PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
SCHOOL OF ENGINEERING

ANALYSIS OF STUDENTS' SELF-REGULATORY STRATEGIES IN MOOCS AND THEIR IMPACT ON ACADEMIC PERFORMANCE

JORGE JAVIER MALDONADO MAHAUAD

Thesis submitted to the Office of Graduate Studies in partial fulfillment of the requirements for the Degree of PhD of Science in Engineering

Advisor:
MAR PÉREZ

Santiago de Chile, June 2019

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
SCHOOL OF ENGINEERING

ANALYSIS OF STUDENTS' SELF-REGULATORY STRATEGIES IN MOOCS AND THEIR IMPACT ON ACADEMIC PERFORMANCE

JORGE JAVIER MALDONADO MAHAUAD

Members of the Committee:

MAR PÉREZ SANAGUSTÍN

MIGUEL NUSSBAUM

JORGE MUÑOZ GAMA

CHRISTIAN BEYLE

CARLOS ALARIO-HOYOS

JORGE VÁSQUEZ

Thesis submitted to the Office of Graduate Studies in partial fulfillment of the requirements for the Degree of PhD of Science in Engineering

Santiago de Chile, June, 2019

Dedicated to my wife Viviana, my daughter María Victoria, and in memory of my grandmother Mélida Ortega.

ACKNOWLEDGEMENTS

"Caminante no hay camino, se hace camino al andar..." (Antonio Machado)

Quiero aprovechar estas páginas de mi tesis para agradecer a todas las personas e instituciones que me acompañaron a recorrer este camino de cuatro años - los que puedo describir como los más increíbles de mi vida - con el objetivo de lograr titularme como Doctorado en la mejor universidad de Latinoamerica, la Pontificia Universidad Católica de Chile, en la Escuela de Ingeniería.

Quiero dar gracias a Dios, por permitirme trabajar en lo que más me gusta, he sentido su presencia en todo el camino recorrido, me ha acompañado en cada desición buena o mala que he tomado, me ha fortalecido en la debilidad y me ha acompañado en el silencio.

Quiero dar gracias a mi tutora Mar Pérez Sanagustín. Como siempre se lo he dicho, me conformaría con llegar a ser la mitad de buen investigador de lo que ella es. No se como agradecer todo lo que me ha enseñado y todo lo que ha hecho por mí. Gracias por inspirarme durante todos estos años, por enseñarme que hacer investigación vá más allá que escribir artículos, sino también, es buscar hacer buenas preguntas de investigación que permitan generar impacto y resolver un problema. Gracias por enseñarme a ser crítico, más aún con mi propio trabajo y por enseñarme a mirar una misma problemática desde distintas perspectivas. Gracias por ayudarme y enseñarme a no conformarme con el trabajo realizado, siempre se puede mejorar, pero eso requiere que se trabaje con dedicación, con voluntad, pero sobre todo con alegría y pasión, disfrutando del camino a recorrer. A la final, la meta es solo un instante, pero lo valioso es todo lo aprendido durante el camino recorrido. Gracias por las largas horas de conversación sobre la investigación y sobre la vida. Solo me queda decir, gracias por todo lo brindado y disculparme si alguna vez he fallado.

Quiero dar gracias a mi esposa Viviana y a mi hija María Victoria por aceptar acompañarme en cada una de mis travesías. Son mis cómplices de vida, sin ustedes nada

de lo que hago tendría sentido. Gracias Viviana por cuidar de mí, por ser más que mi compañera sentimental, has sido mi fortaleza cuando ya no he querido avanzar, has sido los oidos que han escuchado todas mis quejas, has sido el buen humor que ha combatido mi mal humor, has sido mi calma en medio de la tempestad. Puedes ser solamente una persona para el mundo, pero para mí tú eres el mundo. Gracias María Victoria por tus horas de compañía junto a mi escritorio, tu alegría y tus canciones sin duda me llenaron de energía. Espero que todo lo que hago ahora sirva de inspiración para que logres todo lo que quieras en esta vida y sepas que la palabra imposible cuesta un poco más de esfuerzo.

Quiero dar las gracias a mis padres Jorge y Silvia, no solo por darme la vida, sino también por todos los sacrificios que han hecho por mí y por todo el apoyo incondicional que siempre me han brindado. Gracias por haberme hecho entender que más que padres han sido amigos. Hoy soy el reflejo de lo que ustedes me han enseñado durante toda mi vida, aún cuando la distancia nos ha separado, el corazón nos ha mantenido unidos, como los dedos de la mano. Quiero dar las gracias a mi hermanos Silvia y Juan José, que para mí son la sal de la vida. Con ellos he reído y he llorado, y siempre nos hemos apoyado en las buenas y en las malas. La distancia nos ha ayudado a comprender que el tiempo de calidad es más valioso que la cantidad.

Quiero dar gracias a mi comité de tesis quienes me han ayudado durante estos 4 años de distintas maneras. Gracias profesor Nussbaum por las observaciones siempre positivas y por la oportunidad que me dio de poder colaborar en la revisión de los artículos de la revista Computers & Education, sin duda esto me ayudó a aprender a ser crítico con mi propio trabajo y valorar las contribuciones de otros investigadores. Gracias profesor Muñoz por que siempre que me lo encontraba en el pasillo o pasaba por su oficina, siempre tenía una frase de aliento, una sonrisa y me preguntaba “como van esas publicaciones?” también por el tiempo que dedicaba a nuestras reuniones para hablar sobre cómo hacer Process Mining en educación y el apoyo con los recursos necesarios para mis proyectos. Gracias profesor Beyle por aceptar involucrarse en este trabajo de investigación, sus cuestionamientos, puntos de vista y preguntas me ayudaron a ver un mismo tema desde diferentes perspectivas, esto sin duda contribuyó con el resultado final de este trabajo.

Gracias profesor Alario por acompañarme y motivarme durante este trabajo, aprendí mucho con usted durante el tiempo que pudimos interactuar, gracias por abrirme las puertas para colaborar con sus estudiantes de magister, por la invitación a los proyectos europeos en los que aprendí como se ejecuta un proyecto perfecto, y por las charlas sobre la investigación y la vida. Espero que nos podamos seguir viendo en el futuro y como suele ser, en un país distinto. Gracias al profesor Vásquez por llevar cada una de las actividades de seguimiento del doctorado con la formalidad necesaria, y por sus preguntas invaluables que aportaban un punto de vista distinto sobre el tema.

Quiero dar las gracias a mis compañeros del grupo de investigación Technologies for Digital Learning Lab Luis, Ronald, Nicolás, Diego, Isabel y Josefina por compartir la alegrías, las penas y las diferentes situaciones que nos ha deparado la vida durante nuestros estudios. Voy a extrañar los asados y celebraciones con pizza.

También quiero agradecer al equipo de trabajo de la Dirección de Educación en Ingeniería Jorge Baier, Felipe Garrido, Isabel Hilliger y Ximena Hidalgo por el apoyo que recibí durante todo este tiempo para poder desarrollar los cursos MOOCs, así como también a cada uno de los profesores de la Escuela de Ingeniería con quienes interactué durante la creación de estos cursos. Sin duda fue un placer haber trabajo con todos ustedes, y me llevo lo mejor de cada uno.

Quiero agradecer a los siguientes investigadores, quienes me dieron la oportunidad de poder colaborar con ellos: René Kizilcec, Pedro Muñoz Merino, Xavier Ochoa. También a varios de los estudiantes de doctorado con los que pude interactuar y con quienes trabajamos en varios proyectos, algunos de ellos ya graduados como doctores y otros en el camino de lograrlo: María Elena Alonso, Dan Davis, Pedro Moreno Marcos, Wannisa Matcha, Nora'ayu Ahmad Uzir.

Quiero agradecer al Doctor Carlos Delgado Kloos por abrirme las puertas y recibirme en su grupo de investigación GAST en la Universidad Carlos III de Madrid. Al Doctor Nicolae Nistor por su gentil invitación para presentar mi trabajo de investigación en la

Ludwig-Maximilians-Universität München en Alemania, al Doctor Pep Benlloch por su invitación para colaborar en investigación con la Universidad Politécnica de Valencia, al Doctor Gustavo Ramírez por su invitación para hablar de Autorregulación del Aprendizaje en la Universidad del Cauca en Colombia y por supuesto al consorcio de investigadores del proyecto Europeo Erasmus + MOOC-Maker y LALA, ha sido un privilegio poder trabajar y aprender de todos ustedes.

Finalmente y sin ser menos importantes, quiero dar las gracias de manera muy especial a la Universidad de Cuenca, específicamente al Ing. Fabián Carrasco (exrector) y al Ph.D. Pablo Vanegas (rector actual), a la Dirección de Investigación de la Universidad de Cuenca – DIUC, a la Pontificia Universidad Católica de Chile, al Vicerectorado de Investigación VRI y a la Beca Nacional Chilena Conicyt – Conicyt/Doctorado Nacional 2016/21160081, por todas las facilidades y el apoyo económico brindado durante estos 4 años para poder cursar y terminar con éxito el programa de Doctorado en Ciencias de la Ingeniería.

TABLE OF CONTENTS

Pág.

LIST OF FIGURES

LIST OF TABLES

Pág.

ABSTRACT

Massive Open Online Courses (MOOCs) have become a source of digital content anytime and anywhere. MOOCs offer quality content to millions of learners around the world, providing new opportunities for learning. However, only a fraction of those who initiate a MOOC complete it, leaving thousands of committed students without achieving their goals. Recent research suggests that one of the reasons why students find it difficult to complete a MOOC is that they have problems planning, executing, and monitoring their learning process autonomously; that is, they do not effectively self-regulate their learning (SRL). In this thesis, we will explore the possibilities that Learning Analytics (LA) offers to investigate the learning strategies that students use when self-regulate their learning in online environments such as MOOCs. Particularly, the main objective of this research is to develop instruments and methods for measuring students' SRL strategies (cognitive, meta-cognitive and resource management) in MOOCs, and to analyze their relationship with students' learning outcomes. As a methodological approach, this thesis uses mixed methods as a baseline for organizing and planning the research, combining trace-data with self-reported data to better understand SRL in MOOCs. The main contribution of the thesis is threefold. First, it proposes an instrument to measure learners' SRL profiles in MOOCs. This instrument was validated with an exploratory and confirmatory factorial analysis with 4,627 responses collected in three MOOCs. Second, it presents a methodology based on data mining and process mining techniques to extract learners' SRL patterns in MOOCs. The methodology was applied in three self-paced Coursera MOOCs with data from 3,458 learners where six patterns of interaction were identified. Then, this methodology was adapted and applied in an effort of replication for analyzing a synchronous edX MOOC with data from 50,776 learners where twelve patterns of interaction we identified. The third contribution is a set of empirical studies that show the relationship between SRL strategies and academic performance, using data from six self-paced MOOCs in Coursera and two synchronous MOOCs in Open edX. These empirical studies led us to identify self-reported learners' variables (i.e., gender, prior knowledge and occupation) and self-reported SRL strategies (i.e., goal setting, strategic planning) that were identified as the most relevant to predict academic performance. In conclusion, this thesis offers a set of instruments and methods that could be used by other researchers in different contexts to study SRL in MOOCs. The results of this research open up new avenues for personalization and adaptation of MOOC content according to SRL behaviors and set the basis for the study of SRL as a process in other digitally-supported learning environments.

Keywords: Self-regulated Learning, Massive Open Online Courses, Process Mining, Questionnaire, Measures, Learning Outcomes.

RESUMEN

Los Cursos Abiertos Masivos y en Línea (MOOCs – *Massive Open Online Courses*) se han convertido en una fuente de contenido digital que puede ser abordado de forma atemporal y desde cualquier lugar. Los MOOCs ofrecen contenidos de calidad a millones de estudiantes de todo el mundo, brindando nuevas oportunidades de aprendizaje. Sin embargo, sólo una fracción de los que inician un MOOC logran terminarlo, dejando a miles de estudiantes comprometidos sin alcanzar sus metas. Investigaciones recientes sugieren que una de las razones por la que a los estudiantes les resulta difícil de terminar el MOOC es que tienen problemas para planificar, ejecutar y monitorear su proceso de aprendizaje de manera autónoma; es decir, no autorregulan su proceso de aprendizaje de forma efectiva para lograr terminar con éxito un MOOC. En esta tesis, se explorará las posibilidades que ofrece la Analítica del Aprendizaje (LA – *Learning Analytics*) para investigar las estrategias de aprendizaje que los estudiantes utilizan cuando autorregulan su su aprendizaje en entornos en línea como son los MOOCs. El principal objetivo de esta investigación es desarrollar instrumentos y métodos para medir las estrategias de autorregulación del aprendizaje (SRL – *Self Regulated Learning*) de los estudiantes (p. ej. cognitivas, metacognitivas y de gestión de recursos de estudio) en los MOOCs y analizar su relación con los resultados del aprendizaje de los estudiantes. Como enfoque metodológico, esta tesis utiliza métodos mixtos como línea base para la organización y planificación de la investigación, combinando datos de trazas de eventos de los estudiantes con datos de auto-reporte para comprender mejor el SRL en los MOOCs. La principal contribución de la tesis es triple. Primero, propone un instrumento para medir los perfiles de SRL de los estudiantes en los MOOC. Este instrumento se validó mediante un análisis factorial exploratorio y confirmatorio con 4,627 respuestas recopiladas en tres MOOCs. En segundo lugar, presenta una metodología basada en técnicas de minería de datos y minería de procesos para extraer los patrones de SRL de los estudiantes en los MOOC. La metodología se aplicó en tres MOOCs de Coursera (self-paced) con datos de 3,458 estudiantes, en los que se identificaron seis patrones de interacción. Luego, esta metodología se adaptó y aplicó en un esfuerzo de replicación para analizar un MOOC en

edX síncrono con datos de 50,776 estudiantes donde se identificaron doce patrones de interacción. La tercera contribución, es un conjunto de estudios empíricos que muestra la relación entre las estrategias de SRL y el rendimiento académico, utilizando datos de seis MOOCs (self-paced) en Coursera y dos MOOC síncronos en Open edX. Estos estudios empíricos permitieron determinar las variables demográficas de auto-reporte de los estudiantes (p. ej. género, conocimiento previo y ocupación) y estrategias auto-reportadas de SRL (p. ej. establecimiento de objetivos, planificación estratégica) que fueron identificadas como las más relevantes para predecir el rendimiento académico. En conclusión, esta tesis ofrece un conjunto de instrumentos y métodos que podrían ser utilizados por otros investigadores en diferentes contextos para estudiar el SRL en MOOCs. Los resultados de esta investigación abren nuevas vías para la personalización y adaptación del contenido de un MOOC de acuerdo con los comportamientos autorregulados de los estudiantes y establecen las bases para el estudio de la SRL como un proceso en otros entornos de aprendizaje con soporte digital.

Palabras Claves: Autorregulación del aprendizaje, Cursos Masivos Abiertos y en Línea, Minería de procesos, Cuestionario, Métricas, Resultados de Aprendizaje.

Chapter 1

Introduction

Live as if you were to die tomorrow.
Learn as if you were to live forever.

Mahatma Ghandi

This thesis is framed in the domain of Technology-Enhanced Learning (TEL) and, more specifically in the field of Learning Analytics (LA). LA is the measurement, collection and analysis of the records that learners leave behind in their contexts and using those records to improve their learning. LA is grounded around a variety of fields such as Data Mining (DM) and Process Mining (PM). These two fields are closely related research areas, where DM and PM provide the tools to understand how students learn. The primary motivation of this thesis is to explore the possibilities that LA offers to study learners' self-regulated learning strategies in online environments such as Massive Open Online Courses (MOOCs). This chapter introduces the central concepts, terms and definition of Self-Regulated Learning (SRL) in MOOCs that frame the scope of this thesis. The objectives and contributions deriving from this research are also introduced, as well as the impact of the work and a description of the structure of this thesis.

1. INTRODUCTION

1.1 Motivation

Massive Open Online Courses (MOOCs), since its appearance in 2008, have become a source of digital contents for the great majority of students and may be timelessly addressed. MOOCs offer new opportunities to teach millions of people all over the world, delivering quality digital contents in a course format. According to the data collected by Central Class between 2011 and 2018, about 900 higher education institutions have created more than 11,400 courses, offered through several MOOC platforms, such as Coursera, edX, FutureLearn, MiriadaX, etc., which are reaching more than 101 million people all over the world (Shah, 2018).

The pioneer platforms for MOOC, such as Coursera, FutureLearn and edX, started offering free courses based on a series of videos structured in a course format. Nowadays, these platforms are investing in new developments to support new learning experiences (De Waard et al., 2011; Sharples, Kloos, Dimitriadis, Garlatti, & Specht, 2015; Wong & Looi, 2012). For example, Coursera adapted its MOOC platform to be able to offer its courses on demand. MOOCs on demand allows students to register anytime with a lot of flexibility when watching the videos and developing the activities without having to start or end the course on a date determined by the platform. On the one hand, FutureLearn was designed from the beginning to promote discussions and social interactions by means of a series of integrated tools which allow making comments, to answer them and think about them, and which support the use of mobile devices (Sunar, Abbasi, Davis, White, & Aljohani, 2018). On the other hand, edX launched in 2013 its open source code platform named Open edX, in this way offering the opportunity to join the MOOC "wave" to many universities. Also the improvements introduced by the platforms brought with them the development of new learning scenarios based on the MOOC, which are being used in different contexts and following different teaching methodologies (Pérez-Sanagustín, Hilliger, Alario-Hoyos, Delgado-Kloos, & Rayyan, 2016), from those that are only online up to the others which are more combined or hybrid, as for example the flipped classroom (Ho et al., 2015; Kloos, Muñoz-Merino, Alario-Hoyos, Ayres, & Fernández-Panadero, 2015; Soffer & Cohen, 2015).

Nevertheless, and in spite of the attemps to improve platforms and the development of new learning scenarios based on this type of courses, only a small part of the students registered in a MOOC would actually finish it (Jaggars, 2014; Kay, Reimann, Diebold, & Kummerfeld, 2013). The completion rates typically fall below 10% and do not exceed 25% for highly committed learners (Kizilcec & Cohen, 2017). Recent studies show that the students participate in a MOOC selectively regarding the parts of the course content and they decide when and how to engage with course content without any other support than the course content and structure, which can pose a challenge for many learners (Lajoie & Azevedo, 2006). On the other hand, even though of the students claiming to be committed with the course contents can reach their objectives with success (Kizilcec, Pérez-Sanagustín, & Maldonado, 2017), the fact of not getting through the whole course may be attributed in part to the high diversity of learners' backgrounds, motivations, intentions and prior experiences (Kizilcec & Schneider, 2015; Littlejohn, Hood, Milligan, & Mustain, 2016; Zheng, Rosson, Shih, & Carroll, 2015). Nevertheless, research suggests that the main reason students face difficulties when completing a MOOC is due to the **problems when planning, executing and monitoring their learning process, that is, they experience difficulties when self-regulating their learning process** (Kay et al., 2013; Laplante, 2013).

Self-regulated learning (SRL), is defined as the ability that the students have to initiate metacognitive, cognitive, affective, motivational and behavioral processes in order to take actions to achieve their learning goals and persevere until they succeed (Zimmerman, Boekarts, & Pintrich, 2000). SRL is an ability which may be trained so it is developed through the three iterative stages: planning, execution and reflection (Zimmerman, 2015). Also, SRL operationalized through the deployment of the SRL strategies which are necessary for the students to reach the proposed objectives (Pintrich, 2000). SRL, according to the Pintrich model (1999), differentiates three categories of SRL strategies that the students must use to regulate their own learning: **1) Cognitive strategies** related to the strategies which the students use in the acquisition, storage and recovery of the information (i.e., elaboration, rehearsal, organization); **2) Meta-cognitive strategies** related to the activities performed by the students to monitor and consider their learning process with the purpose to achieve a planned objective (i.e., goal setting, strategic planning to comply with the proposed goals and reflection on the reached results); and **3) Resource management strategies** related to the activities that the students perform to manage the learning context and the considered resources (i.e., time management, help seeking, organization of the study context) (Pintrich, 1999). As well as these strategies, the motivation (intrinsic and extrinsic), the value of the task and the expectation to finish the course successfully play an important role in the self-regulation of the students (Kizilcec & Schneider, 2015; Pintrich, 2000; Zheng et al., 2015).

In an online learning context such as a MOOC, where the support of the professor is scarce and often non-existent, SRL plays a key role for the student to effectively manage the study strategies during the learning process and achieve the objectives. For the study of SRL in an online context, there are two different approaches in the bibliography: 1) as an aptitude or, 2) as a process. On the one hand, several instruments have been developed for studying SRL as an aptitude, such as think-aloud protocols, learning diaries and questionnaires, being the last one the most common type of assessment of learners' SRL strategies (Roth, Ogrin, & Schmitz, 2015). On the other hand, the study of the SRL as a process has gained attention of researchers in the past several years. Since SRL can be conceived as a set of events or actions that learners perform (as a process), rather than descriptions of those actions or mental states that these actions generate (Bannert, Reimann, & Sonnenberg, 2014). Nevertheless, there is a lack of empirical knowledge over the students' SRL strategies used in a MOOC and the relationship that exists between the used strategies and the achievements reached in this type of course (Jakešová & Kalenda, 2015). Furthermore, authors such as Panadero & Alonso-Tapia (2015) point out that currently there are many theoretical studies over the psychological models of the SRL, and that there is a lack of empirical studies capable of finding specific causal mechanisms regarding the SRL strategies instead of great theories.

That is, the current approaches to study SRL in online learning environments are not enough to understand this complex process. In fact, many issues remain open and unsolved when studying SRL in a MOOC, either as an aptitude that learners have or as a set of events that account for the self-regulation process as such (Panadero, Klug, & Järvelä, 2016).

This thesis has focused on exploring these challenges through two main research questions: **(1) What instruments and methods are more appropriate to explore learners' self-regulatory strategies used in MOOCs?** and, (**2) What is the relationship between these strategies and academic performance?**

1.2 Challenges

Several specific challenges for studying SRL in a MOOC have been identified in the bibliography. This section reviews how current bibliography have addressed each of these challenges.

1.2.1 Measuring Self-Regulated Learning in MOOCs

Over the last 30 years, SRL has been studied using different proposed approaches to measure and explain the SRL based on the ideas stated in theory (Panadero & Alonso-Tapia, 2015) and models (Boekaerts, 1999) approaches. These proposals follow two different approaches to face the same problem: to consider the ***SRL as an aptitude*** (self-reported theoretical constructions in which the students report how they believe they learn and regulate their learning) or to consider the ***SRL as a process*** (group of events or actions that the students deploy rather than descriptions of the actions or mental status that these actions generate) (Roth et al., 2015; Wirth & Leutner, 2008). This view of SRL as an aptitude and as a process has allowed the development of different evaluation methods allowing to perceive the strategies used by the self-regulated students in an online context (Wirth & Leutner, 2008).

Among the instruments developed by the study of the ***SRL as an aptitude in online contexts,*** the use of the interview, the learning diary, the thinking aloud protocol and the self-reported questionnaire (Boekaerts & Corno, 2005) is emphasized. This last one is the method most used to evaluate the use of the SRL strategies among students (Roth et al., 2015) and for its use, adaptations have been made to the questionnaires developed for traditional learning contexts (i.e., face to face contexts). Nevertheless, the questionnaires developed for traditional contexts may not be useful for online learning context due to the several characteristics of each context (Tallent-Runnels et al., 2006). Even more, smaller changes in the construction of the questionnaires statements may change the meaning, as well as, the validity of the scales used and reliability of the meaning of the information delivered (Karabenick et al., 2007). The aforementioned fact shows the importance of the context in the study of SRL (Hadwin, Winne, Stockley, Nesbit, & Woszczyna, 2001; Hood, Littlejohn, & Milligan, 2015) and the need to construct new instruments specifically adapted to the MOOC context (which is characterized by the heterogeneity of the participants and the possibilities of the learning context) (Barnard, Lan, To, Paton, & Lai, 2009). Out of the bibliography arises then, the first challenge:

[Challenge 1]: *There is a need for new* ***self-reported instruments*** *adapted to the* ***MOOC context*** *which permit measuring the SRL as an aptitude in this type of courses.*

The study of ***SRL as a process is conceived*** as a group of events or actions that the students perform while they are studying (Bannert et al., 2014). In the online contexts, where the students interact through technological platforms, the actions of the students are captured and stored as digital records of events. Through these records part of the cognitive and metacognitive activity of the students can be demonstrated as learning strategies, while they regulate their learning process in the online contexts (Jakešová & Kalenda, 2015).

Several researchers in the self-regulation field have focused on studying how SRL is produced as a process. As a result, study experiences have been reported in online environments about student learning sequences that demonstrate the SRL actions through the use of software tools. These tools have been designed and adapted specifically for the study of SRL as a continuous process of events. For example, the gStudy tool (Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007) was used to explore the learning skills and metacognition of 8 students in one school. Separately, the nStudy tool (Sonnenberg & Bannert, 2015) was used to analyze the use of metacognitive strategies during the learning process of 35 students aged between 18 and 22 at one university. However, the study of learning sequences of students in a MOOC as traces of events that account for the SRL as a process, where there is no tool to support the strategies, is more complex. That is; there is a need for methods that allow the identification of strategies from unstructured events that are not related directly with the SRL. To implement these methods in a MOOC, these must be able to identify strategies in a context to which no direct functionalities associated with the study of the SRL strategies have been added (Veletsianos, Reich, & Pasquini, 2016) and where there is a great variability of behaviors given the flexibility of the context.

But according to a MOOC Research Institute (MRI) report, these type of studies in MOOCs are scarce, and even more in the study of the self-regulation (Gasevic, Kovanovic, Joksimovic, & Siemens, 2014). On the one hand, there are few studies that explore the learning itineraries performed by students as a result of their interaction with the contents of the course (Beheshitha, Gašević, & Hatala, 2015; Siadaty, Gašević, & Hatala, 2016). On the other hand, these studies do not relate self-regulated behavior with academic performance. Therefore, there is a need for studies allowing to explore and understand: (1) what SRL strategies are used by the students in a MOOC, and (2) if there are any differences between groups of students according to the strategies and academic performance allowing to reveal behavior patterns in a MOOC. Out of the bibliography arises, therefore, the second challenge:

[Challenge 2]: *There is a need for new methods to* ***identify and analyze*** *the digital records of the activities of the students in a MOOC and how these activities give account of the* ***self-regulation strategies*** *used in this type of courses.*

1.2.2 Measuring the Relationship between Self-Regulated Learning Strategies and Academic Performance

The students' SRL strategies have been traditionally evaluated in terms of academic achievements (Taub, Azevedo, Bouchet, & Khosravifar, 2014). Until now in a MOOC, the academic achievement has been mainly related with the completion or finishing of the course. On the one hand, the studies performed have focused on analyzing the patterns of students' commitment and attrition regarding the course content (Clow, 2013; Guo & Reinecke, 2014), for example, with video-lectures (i.e., active, passive) (Guo & Reinecke, 2014; Li, Kidzinski, Jermann, & Dillenbourg, 2015), with assessment activities (i.e., honest, cheaters) (Muñoz-Merino, Ruipérez-Valiente, Moreno, & Delgado-Kloos, 2015) and with discussion forums (i.e., participatory or not) (Joksimović et al., 2015). These patterns have served to characterize the students in a MOOC, with the purpose of making informed decisions over the design of the course and concerning future interventions responding to the needs (Ferguson & Clow, 2015; Freitas, Morgan, & Gibson, 2015; Kizilcec & Schneider, 2015). Nevertheless, these patterns do not deliver a clear image of how the students use the SRL strategies to reach a certain status (i.e., active, passive, etc.), nor do they explain why students complete or do not complete a MOOC, and even less, if they complete it because they finally pass it (Zhenghao et al., 2015)

On the other hand, recent studies indicate that the completion of a course is not necessarily the best measure of achievement in a MOOC (Jordan, 2014; Kizilcec & Schneider, 2015). The literature points out that the students have different reasons to register in a MOOC (Zheng et al., 2015) and the initial intentions must be considered when their behavior in the course is analyzed. In a study performed by Kizilcec and Schneider (2015) the authors proposed the Online Learning Enrollment Intentions scale (OLEI scale – a questionnaire to analyze in a systematic manner the students' intention and their relationship with subsequent behaviors in the course). This scale was applied over 14 courses offered in Coursera and Open edX and it was discovered that the students have different intentions when registering in a MOOC, such as: to improve their abilities in english language, as well as a variety of social, vocational, academic and motivational activities according to their interests. The authors point out that as part of the results obtained by the study, it was shown that the new measure of “success” is needed, one that is less related to the top-down traditional models, to understand the students’ needs when taking a course. Therefore, the concept of academic achievement or completion of a MOOC from the perspective of the student (considering the students’ intentions and the objectives) must be redefined and given a new meaning, defining new metrics of completeness/achievement in a MOOC. Out of that need two new challenges arise:

[Challenge 3]: *More analysis is needed to understand the relationship between the objectives and the intentions of the students with the use of the SRL strategies.*
[Challenge 4]: *More analysis is needed to understand the existing relationship between the SRL strategies that the students use in a MOOC and the academic achievements reached in the course.*

1.2.3 Variables Influencing Self-Regulatory Strategies

Recent researches about MOOCs point out that there are **characteristics of the participants, of the MOOC and the context of the course** which may influence how the SRL strategies are used by the students in these courses. **Regarding the characteristics of the participants,** most of the students registered in a MOOC commit in a selective manner with parts of the contents of the course and just a small percentage eventually finishes the course (Anderson, Huttenlocher, Kleinberg, & Leskovec, 2014; Ho et al., 2015; Kizilcec, Piech, & Schneider, 2013; Nesterko et al., 2014). This diversity of objectives may be attributed in part to the remarkable diversity of motivations, intentions (i.e., selectivity with the course content) and the students’ previous knowledge when staring the course (Kizilcec & Schneider, 2015; Littlejohn et al., 2016; Zheng et al., 2015). Authors such as Zheng (2015) identified variations in the students’ behavior according to their use of the course. For example, there are students which use the course only to satisfy the specific learning needs as a set of resources per module; or as “edutainment” (i.e., incorporating didactic resources in learning processes to motivate and make achieving the proposed objective easier) (Zheng et al., 2015). However, there is a lack of studies to explore and understand which characteristics of MOOC participants affect their SRL strategies when they take the course.

Regarding the MOOC, the current bibliography points out the importance of the course design as one of the related variables which may condition the behavior of the students, such as: the type of video-lectures (Guo & Reinecke, 2014), the formative or summative type of evaluation activities (Alario-Hoyos, Muñoz-Merino, Pérez-Sanagustin, Delgado-Kloos, & Parada, 2016; Freitas et al., 2015), if the course offers certification (Hew & Cheung, 2014), the length of the course (duration of four or more weeks), and the nature of the proposed tasks (collaborative or individual) (Margaryan, Bianco, & Littlejohn, 2015). All these variables influence the behavior and the commitment the students establish with the course and if they stay in it until they finish it successfully.

Regarding the course context, it is important to emphasize the *context of use of the MOOC* and the *type of support that the course gives* to the students. ***Regarding the context of use of the MOOC,*** the H-MOOC framework proposed by Pérez-Sanagustín et al. (2016), illustrates how a MOOC may be used: (1) *as a complement to the curriculum* in a formal format of classes but without being formally recognized by the institution (MOOC as a service), (2) *as a replacement for a traditional course* which has a formal recognition by the institution (MOOC as a replacement), (3) *as a guidance* through which the traditional course is organized around the MOOC (MOOC as a driver), and (4) *as an added value* where the institution supplies the necessary help so the students finish a MOOC but without receiving formal credits (MOOC as an added value) (Pérez-Sanagustín, Hilliger, Alario-Hoyos, Delgado-Kloos, & Rayyan, 2016). Depending on the context of use of the MOOC the students establish different SRL strategies allowing them to commit and finish the course. ***Regarding the type of support delivered by the course,*** studies show that the support given to students during learning has an influence on their behavior (Ferguson & Clow, 2015; Joksimović et al., 2015). For example, in a study performed by Alario-Hoyos et al., (2015) it was observed that the instructor's intervention produces an increase in the students' activity in the social tools offered by the course (Alario-Hoyos, Pérez-Sanagustin, Delgado-Kloos, & Munoz-Organero, 2014). In another study performed by Ferguson and Clow (2015) it was observed that the students which participate the most in social type activities (such as discussion forums and social tools) have a higher probability of finishing or completing the course (Ferguson & Clow, 2015). Therefore, the type of support received through the course conditions the interaction of the students with the available resources in the MOOC and at the same time, the SRL strategies deployed during the learning. Out of the previous bibliography arises a new challenge to explore:

[Challenge 5]: *It is required to understand which* ***personal characteristics*** *of the* ***students****, which characteristics related to the* ***MOOC structure*** *and which characteristics associated to the* ***deployment context of the MOOC*** *condition the use of the students' SRL strategies.*

1.3 Research Questions and Objectives

As part of this thesis and based on the general description of the problem (section 1.1) and the challenges found (section 1.2), this research proposal addresses two main research questions (RQ) which cover the five challenges (Ch) through seven specific goals.

1.3.1 Research Questions

RQ1. What ***instruments*** and ***methods*** are more appropriate to explore learners' ***self-regulatory strategies*** used in MOOCs? [Ch1, Ch2]

RQ2. What is the ***relationship*** between ***SRL strategies*** and ***academic performance,*** taking into consideration the characteristics of the participants, the MOOC and the course context that influence the use of these strategies? [Ch3, Ch4, Ch5]

1.3.2 General Objective

The general objective that guides this research in order to address the two research questions is: *To propose instruments and methods for measuring learners' SRL strategies in MOOCs and their relationship with academic performance.*

1.3.3 Specific Objectives

The general objective is divided in specific objectives related with the posed research questions:

Obj1. Implement an instrument that allows us to study the SRL strategies that the students use in a MOOC as an aptitude [RQ1 – Ch1]
Obj2. Implement methods that allow us to explore the SRL strategies that the students use in a MOOC as a process [RQ1 – Ch2]
Obj3. Identify the relationship that exists between the students' self-regulation strategies (self-reported) and their objectives and intentions with the MOOC [RQ2 – Ch3]
Obj4. Identify the relationship that exists between the students' self-regulation strategies (deployed) and their performance in the MOOC [RQ2 – Ch4]
Obj5. Identify the personal characteristics of the students influencing the use of SRL strategies that they deploy in a MOOC [RQ2 – Ch5]
Obj6. Identify the characteristics of the MOOC influencing the use of the SRL strategies that learners use in these courses [RQ2 – Ch5]
Obj7. Identify the characteristics of the course context influencing the use of the SRL strategies that learners deploy in a MOOC [RQ2 – Ch5]

1.4 Methodology

This thesis uses a mixed methods research design as a baseline for the organization and planning of this research work (Creswell & Clark, 2017). This methodology mixes quantitative and qualitative data for extracting conclusions about the research questions. This methodology was selected for three reasons: (1) qualitative and quantitative data together provide a better understanding on how learners self-regulate their learning either as an aptitude or as a process; (2) using only one type of research (qualitative or quantitative) is not enough to address the research questions given the nature of the problem, while exploring data qualitatively and quantitative will help to define an instrument or identity variables to test, obtaining specific information than can be gained from the results of statistical tests; and (3) mixed methods involve the collection and analysis of quantitative and qualitative data, as well as their analysis, integration and joint discussion, in order to make inferences from all the information gathered and gain a better understanding of the phenomenon under study in a real context (Sampieri, 2008). Figure 1-1 introduces the six steps followed to address the two main research questions. Then, an explanation of each step is provided.

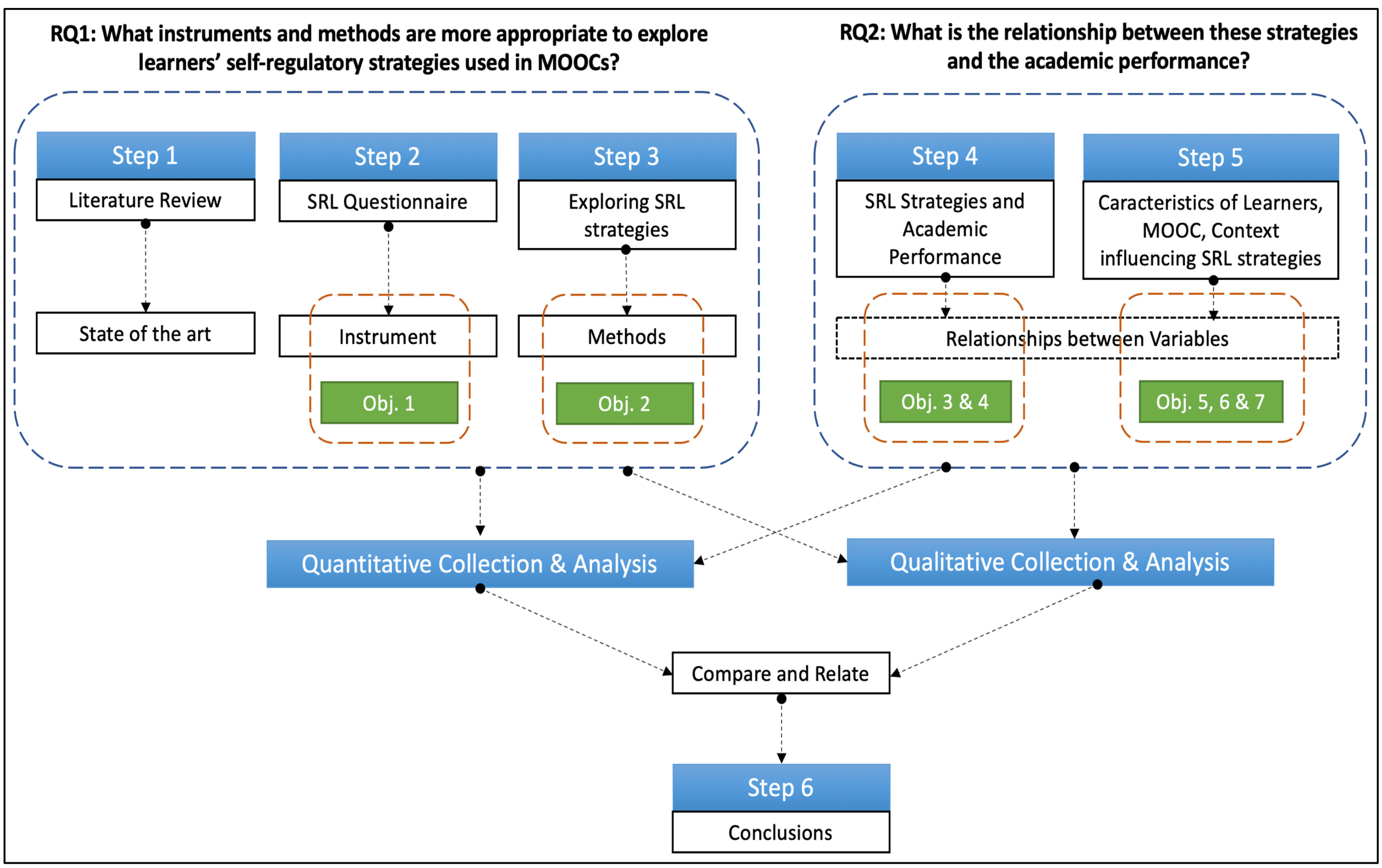


Figure 1-1 Steps defined when applying mixed methods research design

In Step 1, a literature review was conducted. The aim was to make a bibliographic search of SRL models, instruments to measure SRL and experimental studies conducted in online environments and MOOCs that empirically relate SRL strategies with academic performance. To achieve this, a systematic literature review was conducted, following the guidelines defined by Kitchenham (2004), who established three phases for the review: planning, conducting and reporting. From the result of the literature review, an SRL model was adopted and the strategies to be considered in this research thesis were defined.

In Step 2, we developed an instrument to measure SRL in MOOCs. Specifically, a self-reported questionnaire to measure the self-regulation profile of students was created for MOOC context. To achieve this, we took into account the strengths and weaknesses of previous questionnaires and those strategies that showed correlations with student performance according to the bibliography. Then we proposed an instrument evaluated with an exploratory and confirmatory factorial analysis. As a result of this step we obtained a final questionnaire composed by 22 items assessing 5 SRL strategies.

In Step 3, we proposed a methodology to explore SRL strategies in MOOCs. This methodology was proposed as an adaptation of the process mining methodology PM^2 proposed by Van Eck, Lu, Leemans, & Van der Aalst (2015) that combined self-reported data with learners' trace data in the MOOC. The application of this methodology allowed extracting the students' events in the course and relate them with SRL strategies in the literature. Also, we grouped learners according to the strategies they deployed and studied how their behavior relates to academic performance. To validate this methodological approach proposed, we applied it into two different contexts. The first one was with 3,458 learners in 3 MOOCs in Coursera. The second one replicated the methodology applied in the first study but a course in edX with 50,776 learners - this second experimental context provided evidence about the applicability of the methodology in several contexts.

In Step 4, we conducted an empirical analysis to find the relation between SRL strategies with learners' achievements in MOOCs. Specifically, this analysis helped us to better understand the relationship that exists between SRL strategies, obtained from self-reported data and extracted from actual behavior from the MOOC platform, with the academic performance of the learners in a MOOC. To do so, we conducted different steps: (a) an analysis phase where instruments were prepared to collect the data and subsequent analysis of the data; and (b) an evaluation phase where the relationship between achievement and learners' SRL strategies was determined. We applied several statistical techniques to understand what correlation between students' self-regulatory profiles and their achievements exist. As a result of this step, self-reported learners' variables (i.e., gender, age, education), self-reported SRL strategies (i.e., goal setting and strategic planning) and activity sequences patterns (i.e., 'only assessment', 'complete a video-lecture and try an assessment', 'explore the content' and 'try an assessment followed by a video-lecture') were identified as the most relevant to predict academic performance of the learners in MOOCs.

In Step 5 (as a complement of the step 4) we conducted an empirical analysis of variables influencing learner's self-regulation in a MOOC. Specifically, several studies were conducted in order to study the influence of learners' characteristics, MOOC structure and context of the course use on the SRL strategies that learners use. To do so, we conducted the following steps: (a) setting the experimental scenarios, (b) defining and selecting instruments, data gathering techniques and analytic methods, (c) data collection, (d) data analysis. For the analysis, we combined self-reported questionnaire with clickstream data in order to explore SRL as an aptitude and as a process. The results of these analysis helped us to gain insights about how the variables related to learners' characteristics (i.e., gender, age, education and occupation), MOOC structure (i.e., self-paced, synchronous) and context of use of the course (i.e., online, blended) influence students' SRL strategies.

Finally, in Step 6, we conducted a cross-analysis of all the experiments and results obtained to extract conclusions about the main research question addressed. In this case, we followed a convergent parallel design methodology in order to interpret the results, taking advantage of each form of data both quantitative and qualitative. While quantitative data provides for generalizability, qualitative data offers information about the context or setting.

1.5 Contributions

This section summarizes the original contributions of this thesis. Table 1-1 introduces the overall contributions of the entire thesis and includes the following aspects: the research question related (Subsection 1.3.1), its objectives (Subsection 1.3.3); a description of each contribution; the challenges it addresses (Section 1.2); the chapter in which it is presented; and the type of publication in which it was initially proposed (Subsection 1.6.1).

[Contribution 1] Instrument to measure SRL as an aptitude in MOOCs. In this contribution there is a 22-statement questionnaire developed to self-report five SRL strategies (Self-efficacy, Goal setting, Study environment management, Organization and Help-seeking) of the students in a MOOC. This instrument is created based on a bibliography review carried out on how the SRL has been studied in MOOCs. Specifically, the instruments used to measure SRL as an aptitude in face-to-face, online and hybrid learning environments were studied. The results of this contribution have been published in two articles: (1) **[J1] - Journal of Educational Review (literature review)**, and (2) **[J2] – Journal of Research on Technology in Education (the instrument, under review)** and are part of chapter 2 of this thesis

[Contribution 2] A methodology for the discovery of SRL strategies in Coursera MOOCs. In this contribution we present a methodology that combines techniques of data mining and process mining with statistical analysis to extract SRL strategies from students of a MOOC in Coursera. The purpose of the methodology is to answer the following question: What are the most frequent interaction sequences (learning strategies) of learners in MOOCs? The result of this contribution was disseminated through a publication in a journal and a publication in a conference: (1) **[J3] - Journal of Computers in Human Behavior,** and (2) **[C6] – I Learning Analytics Latin America Conference 2018** and are part of the chapter 2 of this thesis.

Table 1-1 Summarizing the contributions of this thesis.

	Specific Objective	Contribution	Challenge	Chapter	Publications											
					Journal						Conference					
					[J1]	[J2]	[J3]	[J4]	[J5]	[J6]	[C1]	[C2]	[C3]	[C4]	[C5]	[C6]
Research Question 1	**What instruments and methods are more appropriate to explore learners' self-regulatory strategies used in MOOCs?**															
	[Obj. 1]	[Cont. 1] Instrument to measure SRL as an aptitude in MOOCs	[Ch. 1]	2	X	X										
	[Obj. 2]	[Cont. 2] A methodology for the discovery of SRL strategies in Coursera MOOCs	[Ch. 2]	2			X									X
		[Cont. 3] Adaptation of the methodology for the discovery of SRL strategies in edX MOOCs	[Ch. 2]	2				X								
Research Question 2	**What is the relationship between SRL strategies and academic performance, taking into consideration the characteristics of the participants, the MOOC and the course context that influence the use of these strategies?**															
	[Obj. 3]	[Cont. 4] Identification of the SRL strategies that are most helpfull to achieve personal goals and intentions in MOOCs	[Ch. 3]	3					X							
	[Obj. 4]	[Cont. 5] Classification of learners based on the relation between their SRL strategies deployed and their achievements in MOOCs	[Ch. 4]	3			X	X			X					
		[Cont. 6] Identification of SRL strategies that predict learners' success in MOOCs	[Ch. 4]	3						X		X				
	[Obj. 5]	[Cont. 7] Identification of the learners' characteristics that predict the use of SRL strategies in MOOCs	[Ch. 5]	3					X				X			
	[Obj. 6 & 7]	[Cont. 8] Identification of SRL strategies employed by students in a MOOC in a Blended context	[Ch. 5]	3										X	X	

[Contribution 3] Adaptation of the methodology for the discovery of SRL strategies in edX MOOCs. This contribution introduces the adaptation of the methodology developed in the [Contribution 2] using the same techniques to extract the SRL strategies of the students of a MOOC in edX. The purpose of this publication was to replicate the study performed in the [Contribution 2] with the purpose to answer to the following question: Which behavioral patterns (learning strategies) can be identified through study sessions in a synchronous MOOC in edX? The result of this contribution is under review in the journal **[J4] – Journal of Computing in Higher Education** and is part of the chapter 2 of this thesis.

[Contribution 4] Identification of the SRL strategies that are most helpful to achieve personal goals and intentions in MOOCs. In this contribution, empirical data over the existing relationship between the self-reported strategies of SRL (i.e., goal setting, strategic planning) with the student personal goals in a MOOC (i.e., earning a course certificate, completing assessments and watching video-lectures) are introduced. Furthermore, the students' self-reported personal intentions are related (i.e., enrolling to earn a certificate, to meet new people, to take the course with others, because of the prestige of the institution or instructor, because the course is relevant to one's research, one's job, or one's school/degree program) which are related with higher SRL skills in a MOOC. For this, an analysis was performed using the previously developed SRL questionnaire which was implemented on six MOOCs in Coursera. With the gathered data (course achievement, intentions, goals and survey responses) and using techniques of statistical regression the following question was answered: Which self-reported SRL strategies are most helpful to achieve personal course goals? The result of this contribution was published in the journal **[J5] – Journal of Computers and Education** and is part of the chapter 3 of this thesis.

[Contribution 5] Classification of learners based on the relation between their SRL strategies deployed and their performance in MOOCs. This contribution presents a classification of the students based on the academic achievements obtained in a MOOC and the SRL strategies used. These groups are: (a) sampling learners, (b) targeting learners, (c) comprehensive low and high learners, (d) low, middle and high self-regulated learners.

For this contribution, the analysis of the data described in the contributions 2 and 3 was used as a basis, using the SRL questionnaire and the clickstream obtained from the courses deployed in Coursera, edX and Open edX platforms to gather data. With the data gathered and using techniques of data mining and technical statistics the following questions were answered: How do the interaction sequences of learners with different academic performance differ? Can we classify learners in different groups according to these behavioral patterns? Is there a difference in terms of academic achievements between the identified groups? The results of this contribution were published in the **[J3] Journal of Computers in Human Behavior**, **[J4] – Journal of Computing in Higher Education (under review)** and in the conference **[C1] - XLII IEEE CLEI 2016 – Informatics Latin America Conference** and are part of the chapter 3 of this thesis

[Contribution 6] Identification of SRL strategies that predict learners' success in MOOCs. In this contribution, the results of a study that analyses the SRL strategies together with other variables of the student profile and their interaction with the course that better predict the good academic performance of the students in a MOOC are presented. These strategies are: (1) self-reported SRL strategies 'goal setting', 'strategic planning', 'elaboration' and 'help seeking'; (2) activity sequences patterns 'only assessment', 'complete a video-lecture and try an assessment', 'explore the content' and 'try an assessment followed by a video-lecture'; and (3) learners' prior experience, together with the self-reported interest in course assessments, and the number of active days and time spent on the platform. For this purpose, the analysis of the data described in the [Contribution 2] was used as a basis, using the SRL questionnaire and the clickstream obtained in a MOOC course deployed in Coursera as instruments to gather data. With the data gathered and using statistical regression techniques, the following question was answered: Which indicators of SRL obtained from self-reported questionnaires and activity sequence extracted from trace data can predict course success in self-paced MOOCs? The result of this contribution was published in the **[C2] - XIII European Conference on Technology Enhanced Learning ECTEL - 2018** and is part of the chapter 3 of this thesis. The results found (identification of the most important indicators of the SRL in

MOOCs) were used for the proposal of design for the NoteMyProgress tool published in the **[J6] - Journal of Universal Computer Science**.

[Contribution 7] Identification of the learners' characteristics that predict the use of SRL strategies in MOOCs. In this contribution an analysis of the personal characteristics of the students that best predict self-reported SRL strategies based on questionnaires before initiating the MOOC is presented. The results indicate that: (1) Older learners reported higher levels of SRL consistently, (2) Women reported higher levels of goal setting, task strategies, and especially help seeking and lower levels of strategic planning, elaboration, and self-evaluation, (3) Learners with a professional or master's degree, and Ph.D. reported higher levels of goal setting, strategic planning, and task strategies, (4) Learners who were employed were more inclined to engage in goal setting, strategic planning, and help seeking, and (5) Learners who had completed more online courses consistently reported higher SRL, especially goal setting. For this purpose, an analysis was performed involving the use of the SRL questionnaire previously developed and applied on six MOOC in Coursera. With the data gathered and using statistical regression techniques an answer was given to the following question: How do self-reported SRL strategies vary by individual learner characteristics? The result of this contribution was published as a contribution in the journal **[J5] – Journal of Computers and Education** and in the **[C3] - III Conference on Learning @ Scale ACM - 2016** and are part of the chapter 3 of this thesis.

[Contribution 8] Identification of SRL strategies employed by students in a MOOC in a Blended context. In this contribution, the SRL strategies used by the students which used a MOOC as part of a flipped classroom (FC) mode are presented. The results indicate that the students with high SRL profiles unlike the students with low SRL profiles show the following behavior in the MOOC: (1) they return to look up or search for specific content before finishing an evaluation activity and then continue, (2) return to the beginning of the module to organize or summarize the learned concepts and (3) try to go through the last module in an autonomous manner. For this, an analysis was performed that involved the use of the SRL questionnaire and the use of PM techniques previously developed for the analysis of data of an MOOC deployed in the open edX platform. It was used in a blended context as part of the FC pedagogical model, with the purpose of answering the following question: How does the behavior of students with different self-regulatory profiles differ when a MOOC is used as part of an FC proposal? The result of this case study was published in the conferences **[C4] – XII IEEE Latin American Conference on Learning Technologies LACLO - 2017** and in the **[C5] - HybridEd Workshop 2018: Successful and Promising Experiences in Blended Learning with MOOCs** and are part of the chapter 3 of this thesis

1.6 Impact

The main results of this thesis have an impact at different levels: (1) at academic level, through scientific publications and collaborations with other institutions; and (2) at national and international level.

1.6.1 Academic Impact

This thesis project has produced a new instrument and a methodology to measure and analyze the SRL strategies that the students use when they studied in MOOC courses and understand the relationship existing between the use of the SRL strategies with its academic performances. Also, we have analyzed the individual characteristics of the students, the structure of the course and the context as factors influencing in the use of these strategies. The results of this thesis have been useful to open the debate in the community of Learning Analytics, on how to advance in the complex analysis of educative data from different perspectives (qualitative as well as quantitative) to show the students' behavior from empirical data and therefore contribute to the current theories of the educational sciences.

The results of the impact are reflected in the number of scientific publications performed, a total of 12 (see Table 1-2) and the number of citations reached since 2016 (358 citations until May 15th, 2019).

Table 1-2 Summarizing the publications of this thesis.

Article	Journal	Conference	Status
1. [J1] Alonso-Mencía, M., Alario-Hoyos, C., **Maldonado-Mahauad, J**., Delgado-Kloos, C., Estevez-Ayres, I., Pérez-Sanagustín, M., (2018). Self-regulated learning in MOOCs: Lessons learned from a literature review, vol 71, pp.1-27.	Journal of Educational Review		Published
2. [J2] Maldonado-Mahauad, J., Pérez-Sanagustín, M., & Beyle, C., (2019). A Questionnaire for Measuring Self-Regulated Learning in Massive Open Online Courses (2019).	Journal of Research on Technology in Education		***Under Review***
3. [J3] Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., & Munoz-Gama, J. (2017). Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses, vol 80, pp. 179-196	Journal of Computers in Human Behavior		Published
4. [J4] Maldonado-Mahauad, J., Davis, D., Alario-Hoyos, C., Delgado-Kloos, C., Pérez-Sanagustín, M., (2019). Adapting a Process Mining Methodology to Analyze Learning Strategies in a Synchronous Massive Open Online Course.	Journal of Computing in Higher Education		***Under Review***
5. [J5] Kizilcec, R., Pérez-Sanagustín, M., **Maldonado J.,** (2017), Self-Regulated Learning Strategies Predict Learner Behavior and Goal Attainment in Massive Open Online Courses, vol 104, pp. 18-33	Journal of Computers & Education		Published
6. [J6] Pérez-Álvarez, R., **Maldonado-Mahauad, J.,** & Pérez-Sanagustín, M. (2018). Design of a Tool to Support Self-Regulated Learning Strategies in MOOCs, vol 24 (8), pp. 1090-1109	Journal of Universal Computer Science		Published
7. [C1] Maldonado, J. J., Palta, R., Vázquez, J., Bermeo, J. L., Pérez-Sanagustín, M., & Munoz-Gama, J. (2016, October). Exploring differences in how learners navigate in MOOCs based on self-regulated learning and learning styles: A process mining approach (pp. 1-12). IEEE.		XLII IEEE CLEI 2016 – Informatics Latin American Conference	Published
8. [C2] Maldonado-Mahauad, J., Pérez-Sanagustín, M., Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., & Delgado-Kloos, C. (2018, September). Predicting Learners' Success in a Self-paced MOOC Through Sequence Patterns of Self-regulated Learning (pp. 355-369). Springer.		XIII European Conference on Technology Enhanced Learning – ECTEL 2018	Published

9. [C3] Kizilcec, R., Pérez-Sanagustín, M., & **Maldonado, J.,** (2016). Recommending self-regulated learning strategies does not improve performance in a MOOC (pp. 101-104). ACM.		III Conference on Learning@ Scale – ACM 2016	Published
10. [C4] Maldonado, J. J., Pérez-Sanagustín, M., Bermeo, J. L., Muñoz, L., Pacheco, G., & Espinoza, I. (2017, October). Flipping the classroom with MOOCs. A pilot study exploring differences between self-regulated learners (pp. 1-8). IEEE.		XII IEEE Latin American Conference in Learning Technologies LACLO - 2017	Published
11. [C5] Maldonado-Mahauad, **J.**, Perez-Sanagustin, M., Pacheco, G., Espinoza, M., Bermeo, J., (2018). Analyzing students' SRL strategies when using a MOOC as a Book (pp. 1-2).		HybridEd Workshop 2018: Successful and Promising Experiences in Blended Learning with MOOCs	Published
12. [C6] Sapunar-Opazo, D., Pérez, R., **Maldonado-Mahauad**, **J.,** Alario-Hoyos, C., Perez-Sanagustin, M., (2018). Analyzing learners' activity beyond the MOOC (pp. 1-8).		I Conference on Learning Analytics in Latinamerica	Published

Academic Reports for International Projects Published

1. Pérez-Sanagustín, M., Hilliger, I., **Maldonado-Mahauad**, **J.,** Pérez, R., Ramírez, L., Muñoz-Merino, P., Tsai, Y., Ortiz, M., Broos, T., Pesantez, P., Sheihing, E., & Whitelock-Wainright, A., (2019). The LALA Framework. LALA project Erasmus + Learning Analytics for Latin America. Link: https://www.lalaproject.org/wp-content/uploads/2019/04/LALA_framework_Spanish.pdf

2. Pérez-Sanagustín, M., **Maldonado, J.,** & Morales, N. (2016). State of the art in the MOOCs adoption in the High Education in Latino America and Europe. MOOC-Maker Construction of Management Capacities of MOOCs in Higher Education. MOOC-Maker.
   Link: http://www.mooc-maker.org/wp-content/files/WPD1.1_ESPAOL.pdf

3. Pérez-Sanagustín, M., **Maldonado, J.**, & Valdenegro, B. (2016). Report on the Technologies and Infrastructure in the Management of the MOOC. MOOC-Maker

Construction of Management Capacities of MOOCs in Higher Education. MOOC-Maker.
Link: http://www.mooc-maker.org/wp-content/files/WPD1.9_ESPAOL.pdf

Research Visits and Initiatives

In addition to the different publications, this thesis has given the author the possibility to perform research visits to specialized institutions and laboratories in which the study of SRL in online learning environments with the help of learning analytics have been the subject of ongoing research. These are:

- Research internship at the GAST Research Group at the Universidad Carlos III de Madrid, Spain. December 2017 – March 2018.
- Research visit to the Research Unit of Empirical Educational Research and Educational Psychology, Faculty of Psychology and Educational Sciences at the Ludwing-Maximilians-Universitat Munchen, Germany. January 2018.
- Research visit to the Computers and System Informatics Department at the Universitat Politécnica de Valencia, Spain. January 2018.
- Development of the Learning Analytics Latin-American Community as a Coordinator under the LALA-project. From October 2018 to present.

1.6.2 National and international impact of the research

The results of this research thesis have had an impact at national and international level:

At national level: The experimental educational scenarios that have been developed in this thesis have been presented as a reference case study in other Chilean institutions for the application of MOOCs in higher education, and which have been performed in the context of the Fondecyt Initiation project called Self-Regulated Learning Strategies in MOOC-based Environments, proposal ID 11150231.

At international level: On the one hand, the MOOC courses developed as part of this thesis have contributed to the training of the Latin American teachers and were performed in the context of the European Erasmus+ MOOC-Maker project (http://www.moocmaker.org/), consisting of 9 partners from Latin America and Europe who have as an objective the development of the training for the design and creation of the MOOCs in the higher education institutions. On the other hand, the data analytic methodologies developed in this thesis and the identification of the SRL strategies influencing the learning results of the students, have been taken as a basis for the design and development of the NoteMyProgress tool. This tool is used today in the context of the European Erasmus+ LALA - Learning Analytics Latin America project (https://www.lalaproject.org/) in which the Pontificia Universidad Católica de Chile participates as a partner with other 3 partners from Latin America and 3 partners from Europe with the objective of developing and adapting Learning Analytics tools in higher education institutions.

1.7 Document Structure

This thesis document is organized following a structure based on chapters that present papers that were sent for review or were published in an ISI journal or a Conference. Four journal and six conference papers had already been accepted and published at the time this thesis was written. This document is structured into four main chapters.

Chapter 1 is an introductory chapter with the aim of giving the reader an overall idea of the research area, presenting the reasons that motivated the work and the challenges identified, the research questions proposed, the objectives, the methodology used, the main contributions and the impact of the results of the thesis.

Chapter 2 presents **"Instruments and methods for measuring SRL strategies in MOOCs"**. This chapter introduces the main contribution regarding the *RQ1* - Instruments and methods to understand SRL strategies in MOOCs. Specifically in this chapter are presented as contributions: 1) an SRL questionnaire for capturing the learners' self-reported SRL strategies (the final instrument was included in the appendix) *[Cont. 1 – Obj. 1]*; 2) a methodology employed to extract SRL strategies from actual students' behavior and its relation with self-reported measures of SRL in Coursera MOOCs *[Cont. 2 – Obj. 2]*, and; 3) an adaptation of the methodology in order to be applied into a different MOOC and discussions on how the methodology proposed can be generalizable for other contexts *[Cont. 3 – Obj 2].*

Chapter 3 introduces the contributions related with the **"Relationship between SRL strategies and academic performance".** This chapter introduces the main contribution regarding the *RQ2 - Relationship between SRL strategies and academic performance.* Specifically in this chapter are presented as contributions: 1) the identification of the SRL strategies that are most helpful in achieving personal goals and intentions in MOOCs *[Cont. 4 – Obj. 3]*; 2) a classification of learners based on the relation between the SRL strategies employed and their achievements in MOOCs *[Cont. 5 – Obj. 4]*; 3) the identification of SRL strategies that predict learners' success in MOOCs *[Cont. 6 – Obj. 4]*; 4) the identification of the learners characteristics that predict the use of SRL strategies in MOOCs *[Cont. 7 – Obj. 5]*; and finally; 5) the identification of SRL strategies employed by learners in a MOOC in a blended learning context *[Cont. 8 – Obj. 6 & Obj. 7]*.

Chapter 4 introduces the **"The main conclusions of this thesis and lessons learned"**. In addition, the aspects to be considered as part of future work in this research area are included.

This thesis also includes two appendices: 1) Appendix A that contains the SRL Questionnaire developed for MOOCs; and 2) Appendix B that contains the citation and the first page of each of the publications presented as part of this thesis and other information as a complete reference about the work done.

Chapter **2**

Instruments and methods for measuring SRL in MOOCs

The best inheritance that a father can leave to his children is his education.

Own authorship

This chapter shows the main contributions concerning the first research question: *"Instruments and methods to measure the SRL strategies in MOOCs"*. This chapter is structured in 5 subsections showing the contributions of four journal articles [Table 1-2; J1, J2, J3, J4] and a conference article [Table 1-2; C6]. Specifically, the subsection 2.1 shows an introduction to the chapter 2 where the contributions are shown. The subsection 2.2 shows the development of a questionnaire as an instrument to capture the SRL strategies self-reported by the students in MOOCs. The subsection 2.3 shows the proposed methodology to extract the students SRL strategies based in process mining. This methodology is applied in a first instance to study the students' behavior in three MOOCs deployed on the Coursera platform. Subsection 2.4 presents the result of applying the same methodology in a MOOC deployed on the edX platform in a replication exercise and discusses how this could be generalized to other contexts of application. Finally, subsection 2.5 presents the main conclusions of the chapter.

2. INSTRUMENTS AND METHODS FOR MEASURING SRL IN MOOCS

2.1 Introduction

This chapter presents the results related to *RQ1: What instruments and methods are more appropriate to explore learners' self-regulatory strategies used in MOOCs?* The main results have been reported in four journal papers. Each journal paper addresses particular sub research questions (Sub-RQ) that arise from the main research question to inform the main conclusions. Table 2-1 summarizes the main Sub-RQ addressed in each paper and the specific objective to which they are related.

Table 2-1 List of sub research questions related to the RQ1.
J[x] and C[x] are the identificators used to refer to journal and conference papers respectively, where "x" indicates the number of the (journal or conference) paper.

<table>
<tr><th>Specific Objective</th><th>Publication</th><th>Sub-research Question</th></tr>
<tr><td colspan="3">RQ1. What instruments and methods are more appropriate to explore learners' self-regulatory strategies used in MOOCs?</td></tr>
<tr><td>[Obj.1] - [Ch1]</td><td>[J1] Self-regulated learning in MOOCs: Lessons learned from a literature review
[J2] A Questionnaire for Measuring Self-Regulated Learning in Massive Open Online Courses</td><td>Sub-RQ 1.1: What SRL models and SRL strategies have been studied in traditional and online contexts?</td></tr>
<tr><td>[Obj.2] - [Ch2]</td><td>[J3] Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses
[J4] Adapting a Process Mining Methodology to Analyze Learning Strategies in a Synchronous Massive Open Online Course
[C6] Analyzing learners' activity beyond the MOOC</td><td>Sub-RQ 1.2: What are the most frequent interactions sequences of learners in MOOC?

Sub-RQ 1.3: To what extend can we replicate (partially or totally) the methodology applied in the previous study [J3] to extract students' learning strategies in a MOOC?

Sub-RQ 1.4: How do students' learning strategies in this new context differ from those from the previous study [J3]?</td></tr>
</table>

Each subsection in this chapter is structured as follows. First, the context to frame the sub research questions addressed in each paper is presented. Second, we present the related work. Third, the analytical methods used to answer the sub research questions are presented. Fourth, the main results are presented. This chapter ends with a conclusion that summarizes the lessons learned of each sub research question in order to inform the RQ1.

2.2 A Questionnaire for measuring Self-regulated Learning in MOOCs

In education, self-regulation of learning is a very important area of study (Boekaerts & Cascallar, 2006), since it is considered one of the most important skills for lifelong learning (Ifenthaler, 2012) and for the XXI century. For the specific case of MOOCs, self-regulated learning strategies are key for students to achieve their objectives. MOOCs require students to have an active and self-directed behavior (Moore, 1986), as they are expected to self-regulate their learning process autonomously. In the last 30 years, for the study of the SRL, formal guidelines have been proposed from the theoretical approach (e.g., operative, socio-cognitive, volitional, constructivist, etc.) (Panadero & Alonso-Tapia, 2015) and models (e.g., Zimmerman, Pintrich, Winne and Hadwin, Efklides, Boekaerts, etc.) (Boekaerts, 1999; Panadero, 2017) that have tried to explain how SRL develops during the learning process. All the authors of the models agree that the SRL is cyclical, consisting of different phases and processes. However, each model conceptualizes the phases and sub-processes in a different way. This difference in the conceptualization of the models affects the type of interventions that can be made and the instruments that are developed to measure it. For example, in relation to the interventions that could be made to promote SRL, in models such as Efklides (2011) the SRL has a top-down approach guided by personal objectives, while in models such as Pintrich (1990) regulation has a data-driven (bottom-up) approach guided and directed by students' actions. On the other hand, the model taken as a reference also affects the instruments that are developed to measure SRL. According to the bibliographic review developed by Roth, Ogrin and Schmitz (2015) on instruments to assess SRL in higher education, the use of interviews, learning journals, think-aloud techniques and questionnaires stands out. The latter are the most common method to evaluate the SRL strategies which students believe to have made effective use of during their learning process (Roth et al., 2015).

The questionnaires are an important part of a research strategy in the SRL study. By means of these it is possible to collect data about what learners are doing, specifically, they are great for measuring opinions as scales (Floden, 1981). In addition, the questionnaires help identify learners' preferences in the use of learning strategies that could influence their learning achievement (Broadbent & Poon, 2015). In a complementary way, Veletsianos et. al. (2016) also highlight the importance of questionnaires when studying students' behavior in online contexts, since they are an essential part of complementing the use of data extracted from platforms and making more complex analysis (Maldonado-Mahauad, Pérez-Sanagustín, Kizilcec, Morales, & Munoz-Gama, 2018).

In the last 3 decades several questionnaires have been developed to measure the SRL strategies of students and have been used in different contexts. On the one hand, for face-to-face traditional contexts, questionnaires have been developed such as the Motivated Strategies for Learning Questionnaire-MSLQ designed to measure learning strategies and motivation (Pintrich, Smith, Garcia, & McKeachie, 1991), the Self-Efficacy for Learning Form-SELF designed to measure learners' perceived self-effectiveness regarding the implementation of specific learning strategies (Zimmerman & Kitsantas, 2007), the Self-Regulated Learning Scale-ASRLS designed to measure self-regulation of college students that is within the context of their learning in higher education (Magno, 2011). However, these questionnaires are not suitable for the use in online contexts to measure SRL. Cho and Summers (2012) have shown that MSLQ could not be validated in an asynchronous online learning environment. The questionnaires designed for the measurement of SRL strategies must take into consideration the learning context to achieve their purpose (Barnard et al., 2009; Hood et al., 2015). For this reason, different authors proposed the development of specific questionnaires for the measurement of SRL in online learning contexts. Examples of these questionnaires are the Learning and Study Strategies Inventor-LASSI designed to collect information about study and learning practices, as well as attitudes (Weinstein & Palmer, 2002), the Online Self-regulated Learning Questionnaire-

OSLQ designed to measure learners' ability to self-regulate their learning in blended and online environments that are wholly or partially web-based (Barnard et al., 2009). However, when looking at the statements contained in these questionnaires, it can be seen that they are based on instruments developed in the decade of the 90's, so that their statements reflect the technological context of the time and could be outdated (e.g., "I prepare my questions before joining the chat room"). In this sense, the questionnaires developed for online environments do not reflect the effect of the context of a MOOC (characterized by the heterogeneity of the participants and the possibilities of the environment for learning). That is why it is important to build instruments that measure what is effective.

In the last two years, efforts have been made to build questionnaires for the MOOC context. Example of these questionnaires are the one proposed by Jansen, van Leeuwen, Janssen, Kester, & Kalz (2016) (SOL-Q), the one proposed by Littlejohn et al. (2016) and the OSLQ for MOOC adapted to the Russian language (Martinez-Lopez, Yot, Tuovila, & Perera-Rodríguez, 2017). However, these questionnaires have their limitations. On the one hand, the first two questionnaires are composed of a high number of items to be answered (between 36 and 38 items) and, as a consequence, students tend not to answer them (Veletsianos et al., 2016), resulting in a low rate of response, which makes a posteriori analysis difficult. On the other hand, the OSLQ questionnaire for MOOC adapted to Russian is not applicable in other cultural contexts where the language is not Russian, such as a cultural context where the language is English or Spanish. The SOL-Q questionnaire was obtained after having performed an exploratory and confirmatory factorial analysis. However, SOL-Q does not consider SRL strategies that according to the bibliography are related to the learners' outcomes (such as goal setting). The questionnaire proposed by Littlejohn et al. (2016) was designed to measure the subprocesses of SRL but for adult learners in informal learning contexts (Fontana, Milligan, Littlejohn, & Margaryan, 2015).

Therefore, *there is the need to create an updated instrument adapted to the context of the MOOCs*, which allows to account for the SRL skills students use and that are related to achievements in MOOCs. Specifically, we present and validate a questionnaire created to measure learners' ability to self-regulate their learning in MOOCs. We call it the MOOC-SRLQ (Self-Regulated Learning Questionnaire). This questionnaire has been built over (1) an analysis related to the actual SRL Questionnaires used in traditional and online learning but suited for the MOOC context, and (2) an analysis related to the SRL strategies that have been related with outcomes in MOOCs. To achieve this purpose, an exploratory and confirmatory factor analysis has been conducted. The final questionnaire is composed of 22 items to assess 5 SRL strategies in MOOCs: Self-efficacy, Goal setting, Study environment management, Organization and Help-seeking.

2.2.1 Related Work

This section contains the bibliographic review related to SRL models and the instruments created from these models in order to answer to the *Sub-RQ 1.1: What SRL models and SRL strategies have been studied in traditional and online contexts?* An analysis of the proposed questionnaires that have been used to measure SRL strategies for both traditional face-to-face contexts and for online learning contexts during the last 25 years is also presented. In a complementary way, we present the results of a bibliographic review on SRL strategies that have been positively related to the achievements of students in online contexts, emphasizing the results of experiments in MOOCs. The results of this bibliographic review are the basis for the construction of the proposed instrument.

2.2.1.1 SRL Models and Instruments

In the last three decades, several theoretical models have been proposed and used to study the self-regulatory skills of students in various contexts. The Panadero (2017) bibliographic review highlighted the most used models that have the highest number of citations per year according to the data collected in google scholar (until 27th April, 2017). These models are: 1) the socio-cognitive model of Zimmerman who developed three models and has 4,169 citations (Zimmerman, 2000), 2) the Boekaerts model focused on the role of emotions and has 1,011 citations (Boekaerts & Corno, 2005), 3) the Winne and

Hadwin model focused on the study of SRL from a metacognition perspective and has 1,037 citations (Winne & Hadwin, 1998), 4) the Pintrich model that emphasizes the role of motivation in the SRL and has 3,416 citations (Pintrich, 2000), 5) the model of Efklides that has a stronger metacognitive background than the other models and has 251 citations (Efklides, 2011) and 6) the model of Hadwin, Järvelä and Miller that positions the study of SRL in the context of collaborative learning and has 196 citations (Hadwin, Järvelä, & Miller, 2011). Almost all the authors of these models agree that the SRL is cyclical and conceptualize it as a process that occurs in three phases (with the exception of Efklides, who coincides only with the first two): a) preparatory phase which includes task analysis and goal setting, b) performance phase in which the task is done while monitoring and controlling the progress of the performance, and c) appraisal phase in which the learner reflects on and regulates for future performance (Puustinen & Pulkkinen, 2001).

These different conceptualizations of SRL have had a strong influence on the development of instruments for its assessment. Most of these instruments are questionnaires that aim to assess the use of SRL skills by defining different SRL strategies (Baumert & Köller, 1996) and administering a number of questionnaire items per class (Wirth & Leutner, 2008). For example, under the Zimmerman model 5 instruments and measures to assess SRL have been developed, of which the Academic Self-Regulation Scale (A-SRL) is the most important. Under the Pintrich model one of the major contributions has been the construction of the MSLQ questionnaire to measure motivation and learning strategies, and according to the bibliographic review by Roth et al. (2015) 94 researchs on SRL have been developed using this questionnaire. Under the Boekaerts model, several reflection articles on how to measure SRL have been written and, as a result, four instruments and methods were developed to evaluate the SRL, with the OMQ questionnaire being the most representative. Under the Efklides model (based on the Pintirich model), the Metacognitive Experiences Questionnaire was created, which explores the cognitive processes of learners. Under the Winne and Hadwin model, although classical instruments have not been built, the model has served as a theoretical framework of reference for studying the traces of SRL. Finally, under the model of Hadwin, Järvelä and Miller, no instruments have been developed to measure SRL. Figure 2-1 shows a summary of the instruments and metrics derived from each SRL model.

Zimmerman	Pintrich	Boekaerts
•Self-Regulated Learning Interview Schedule (SRLIS) •Procedures to assess SRL for writting and dart throwing •Microanalytic measures to assess validity of the cyclical phases model •Measures of self-efficacy to self-regulate •Academic Self-regulation Scale (A-SRL)	•Questionnaire MSLQ that measures cognitive, metacognitive and resource management strategies	•OMQ Questionnaire to measure the sensitivity to learn in concrete situations •Interactive Learning Group System (ILGS) •The confidence and doubt scales: instrument to record student motivation •Development of neural networks for SRL
Efklides	**Winne and Hadwin**	**Hadwin, Järvelä and Miller**
•Questionnaire to measure self-concept for a language task •Metacognitive Experiences Questionnaire (MEQ)	•Tools that measure traces of SRL using the model as nStudy, gStudy computer-supported learning environment	•No measurament instruments for SRL

Figure 2-1 Summary of instruments and metrics derived from each SRL model

Currently, for the study of the SRL, Zimmerman and Pintrich models are the most used. This is based on the number of citations received by both models and the use of instruments that have been built based on these. Additionally, both models include a more complete view of the different types of self-regulation sub-processes compared to the other models (Panadero, 2017). However, the Pintrich model unlike the Zimmerman model combines 4 phases (forethought, monitoring, control and reflection) and 4 areas (cognition, motivation, behavior and context) offering a more comprehensive picture and a greater number of sub-processes that allow us to better understand the SRL process. In addition, Pintrich’s behavior regulation area incorporates the “individuals’ attempts to control their own behavior” (Pintrich, 2000, p. 446) making this model unique in comparison to the others. As a consequence, we follow in the tradition of Pintrich’s model because it focusses on particular strategies, and lends itself more to large-scale quantitative research that can inform targeted interventions to support specific SRL strategies (Kizilcec, Pérez-Sanagustín, & Maldonado, 2017). In another research carried out by Puustinen & Pulkkinen (2001), Pintrich's model is highlighted as one of the most important, since it synthesizes not only the different processes, but also the strategies that contribute to increase the SRL (Montalvo & Torres, 2004; Puustinen & Pulkkinen, 2001). These strategies according to the Pintrich's model can be of the cognitive type (i.e., rehearsal, organization, elaboration, critical thinking), metacognitive (i.e., planning, monitoring, regulation) and resource management (i.e., time management, help-seeking, effort-regulation, study, environment management (Pintrich & De Groot, 1990). For the above reasons, in this work the Pintrich's model will be used as reference for the construction of the theoretical-empirical model of the MOOC-SRLQ questionnaire.

2.2.1.2 Analysis of SRL Questionnaires and Effective SRL Strategies in Online learning

This section summarizes the methodology followed to make a systematic bibliographic review of the questionnaires used to measure SRL, as well as effective SRL strategies that correlate with learning achievements in online learning contexts. The results of the bibliographic review are presented separately and will serve as a starting point to identify the SRL strategies that will be considered as part of the theoretical-empirical model of the MOOC-SRLQ questionnaire.

Methodology of the research. The objectives of this bibliographic review are to: (a) find the existing questionnaires used to measure SRL, the context for which they were developed, and the strategies evaluated, and (b) find the effective SRL strategies that correlate or predict academic learners' achievements in online learning contexts. To achieve these objectives, the bibliographic review was based on the criteria of Kitchenham (2004), which proposes to organize the review in three phases: (1) Plan the review, (2) Conduct the review and (3) Report the results. This was done for both objectives.

Data Collection. For objective (a) - finding the existing questionnaires used to measure SRL, the context for which they were developed, and the strategies evaluated - the initial search focused on scientific databases related to the area of educational technology and psychology: Scopus, ACM Digital Library, Psycarticles, Psycinfo and Web of Science. The Google Scholar search engine was also used. The keywords used to formulate the search queries in the different databases were divided into three categories: (1) Self-regulated learning, Self-regulation, SRL, learning strategy/ies (2) questionnaire, assessment, instrument, and (3) higher education. The search was restricted to articles published between 1991 and 2017 that contained the keywords in the title, summary or in its list of keywords. The formulation of the logical expressions of the consultations can be represented symbolically by: (Self-Regulated Learning, OR SRL OR Self-regulation) AND (questionnaire OR assessment OR instrument) AND (higher education). The year 1991 was taken as a starting point because this is when the MSLQ questionnaire in the bibliography was published.

A total of 523 articles were found that met these criteria, which were subjected to a selection process. To be selected, an article within the set of valid articles should be focused on the proposal of a questionnaire aimed at measuring SRL strategies in higher education. For this selection, the title, summary and list of key words of each article obtained as a result of the search were used. As a result, 201 articles were extracted that fulfilled one of the aforementioned criteria and 322 articles were excluded. Subsequently, duplicate articles (120 articles) were excluded in the different databases. The result was a sample of 81 articles.

For objective (b) - *finding effective SRL strategies that correlate or predict academic learners' achievements in online learning contexts* - the initial search focused on the same databases used for the previous objective. The keywords used to formulate the search queries in the different databases were initially divided into 2 groups. The first group used five categories: (1) Self-regulated learning, Self-regulation, SRL, learning strategy/ies (2) online, MOOC, web based, (3) higher education, learner, course, (4) academic outcome, academic achievement, grade, score, and (5) state of the art, review, systematic review. The search and formulation of the logical expressions were similar to those explained for the previous objective. As a result 1,209 articles were obtained and subjected to a selection process. For this the article should be focused on a review of the literature on SRL strategies and achievement for online contexts for higher education. As a result, only 1 document was found that presented a systematic review of SRL strategies that considered work from 2004 to 2014. For this reason, it was decided to take this article as a reference and expand it by carrying out a second search restricted on articles published between 2004 and 2017. For this a second group with four categories was defined: (1) Self-regulated learning, Self-regulation, SRL, learning strategies (2) online, MOOC, web based, (3) Rehearsal, Elaboration, Organization, Goal Setting, Self-Monitoring, Strategic Planning, Self-Evaluation, Self-Satisfaction, Self-Efficacy, Help Seeking, Time Management, Effort Regulation, Study Environment Management, Metacognition, and (4) academic outcome, academic achievement, grade, score. As a result a sample of 25 articles was obtained.

Three researchers with experience in the research area participated in the selection process. To determine if the article met or not with the selection criteria raised, the title, summary and list of keywords of each article obtained were reviewed. In the event that the three previous sections were not sufficient to determine the validity of the article, we proceeded to read the introduction, as well as the conclusions of the article. To offer greater validity to the results and to ensure that the articles fulfilled the search criteria, all the articles were analyzed by two of the researchers. If the two researchers could not reach a consensus on the inclusion of a particular article, the third researcher participated in the arbitration to achieve consensus.

Classification and Data Analysis. For objective (a) - *find the existing questionnaires used to measure SRL*, the context for which they were developed and the strategies evaluated - the articles were classified into three categories, depending on (1) if the article was a proposal for a new questionnaire to measure SRL strategies, (2) if the article was a study that reused completely or only parts of an existing questionnaire to study the SRL, and (3) if the article proposed a questionnaire (whether it was built based of existing ones or not) that was used in the study to measure SRL. As a result of this process, 69 articles were discarded, since the questionnaires presented in those studies were derived from questionnaires that met all three criteria and were repeated. Finally, as a result of the classification and the analysis, 12 articles were read in their entirety and the results are presented in the following section in Table 2-2.

For objective (b) - *find effective SRL strategies that correlate or predict academic learners' achievements in online learning contexts* - the articles were classified into two categories, depending on (1) if the article talked about one of the strategies of SRL and it correlated it with the learning achievements of the student, (2) if the article talked about one of the SRL strategies and it had been used to predict student's achievement or final grade. As a result of this process, 9 articles that did not meet either of the two criteria were discarded. Finally, as a result of the classification and analysis, 16 articles were obtained that were read in full.

2.2.1.3 Results of SRL Questionnaires

For the 12 articles selected, for each questionnaire we analyzed the following: the scale used to measure each strategy, the number of items that are part of the questionnaire, the context for which it was defined, the year in which the questionnaire was published, the SRL strategies that are included and the authors. Table 2-3 summarizes all this information for all articles analyzed.

Of these, 8 questionnaires were designed to be used in a traditional context (i.e., face-to-face), 3 questionnaires were designed to be used in an online context (2 questionnaires were designed to be used in e-learning courses in general and 1 to be used in MOOC courses) and 1 was designed to be used in a blended and online context. Questionnaires for traditional contexts have between 35 and 120 items to evaluate different SRL strategies and use Likert scales (options from 1 to 4, from 1 to 5 and from 1 to 7) as well as a scale of 100 points, whereas questionnaires for online contexts use between 24 and 36 items to evaluate different SRL strategies using Likert scales (options from 1 to 5 and from 1 to 7) as well as a scale of 100 points. The strategies most evaluated by traditional questionnaires are goal setting, environment structuring, and time management; whereas for online questionnaires the most evaluated strategies are help-seeking and time management.

Of all the questionnaires found and according to a systematic bibliographic review by Roth et al. (2015), the MSLQ questionnaire proposed by Pintrich et al. (1991) has been one of the most used to measure learning strategies and motivations of students in traditional contexts. It is followed by the LASI questionnaire that has been used in 12 studies, while the scale proposed by Zimmerman (Academic Self-Regulated Learning Scale) has been used in only 4 studies according to this review. However, the MSLQ questionnaire, as well as ILS, SESRL, OMQ, SELF, ASRLS, and EFLSRLQ, are questionnaires that were designed to be used in traditional face-to-face contexts. Therefore, if these questionnaires were to be reused in other types of contexts (such as online), their statements would not be adequate, as they would not reflect the effect of the context. This coincides with what the study by Karabenick et al. (2007) mentions, that small changes in the construction of a statement can change its meaning (i.e., "In my science class ..." vs. "My teacher……”). On the other hand, the scales used also provide validity and confidence in the meaning of the information, which can also affect the result (Boekaerts & Corno, 2005; Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007). The recently published questionnaire for a traditional context Scale on Self-regulation in learning (SSRL) (2016), was designed to measure the SRL for university students. However, individual differences such as gender, age, socio-economic level (characteristic variables in a MOOC environment) and the possible influence of the learning environment were not taken into account when validating the questionnaire (Erdogan & Senemoglu, 2016). Therefore, as a result, this questionnaire is not adequate to measure SRL in a MOOC. In addition, in a traditional context (face-to-face) the metacognitive process is usually mediated by the instructor or master teacher (i.e. giving instructions on how to approach the contents), while in a context such as online, mediation is reflected to the extent that the instructional design (or the disposition of the contents) implements it. For this reason, the study of the strategies deployed in different contexts makes it necessary to adapt the statements of the questionnaires to measure them.

The questionnaires designed to measure SRL strategies for an online context, such as LASI and OSLQ, are based on instruments developed in the 90s. As a result, their statements are adapted to the technological context of the time (e.g., "I prepare my questions before joining the chat room", "I have positive attitude about attending my classes"). For a current online context, these statements are outdated, since they do not reflect the effect of the context, and particularly that of a MOOC (characterized by the heterogeneity of the participants, MOOC's massiveness and the possibilities of the environment for learning). In the case of the OSLQ questionnaire according to the authors themselves (Barnard et al., 2009) "*the development and further validation of an instrument like the OSLQ becomes relevant and even necessary given the need to assess courses and learners in emerging online and blended learning environments*" so this questionnaire would have to be adapted for a MOOC context.

The questionnaires for a MOOC context (SOL-Q and SRL Professional) were developed from the combination and adaptation of questionnaire items developed for traditional and online learning contexts. The SOL-Q questionnaire (based on the combination of items from MSLQ, OSLQ and other face to face questionnaires) was obtained after having carried out an exploratory and confirmatory factorial analysis. However, the sample used to obtain it was relatively small, and the strategies proposed by the SOL-Q questionnaire do not consider SRL strategies that according to the bibliography are related to academic performance (such as self-efficacy, goal setting). In addition to the above, this questionnaire has a relatively high number of statements.

The questionnaire proposed by Littlejohn et al. (2016) SRL Professional was a slightly modified version of a published instrument designed to measure SRL of adult learners in informal learning contexts (Fontana et al., 2015). This questionnaire consists of 39 items and takes 15 to 20 minutes to answer, and does not consider the organization, time management and study environment management strategies that, according to the bibliography, are related to academic performance.

After an exhaustive analysis of the selected SRL questionnaires, we identified the following limitations of current instruments to be applied in MOOCs:

1. They are too long in relation to the number of questions and, as a consequence, the response rates are low, limiting the use that can be given to the data posteriori.

2. The statements of the instruments developed for traditional face-to-face contexts are not adequate for measuring SRL strategies in a MOOC context, characterized by the heterogeneity of the participants, MOOC's massiveness and the possibilities of the environment for learning.

3. The statements of the instruments developed for online contexts are based on instruments developed in the 90's and do not reflect the effect of the technological context for MOOC courses.

4. They do not consider strategies that, according to the bibliography, are relevant in this type of course and are related to academic achievements (Table 2-3).

Finally, it is clear that all the questionnaires measure some aspects of the SRL, however none of them measures all aspects of the SRL.

Table 2-2 SRL questionnaires for face to face and online context used between 1991-2016

SRL QUESTIONNAIRES						
Name	**Scale/Measure**	**# Items**	**Context**	**Year**	**SRL Strategies**	**Author**
MSLQ	Likert 1- Not at all true of me 7- Very true of me	81	Face to Face	1991	Intrinsic and extrinsic goal orientation, Task values, Control beliefs, Self-efficacy, Test anxiety, Rehearsal, Elaboration, Organization, Critical thinking, Time and study environment, Metacognitive self-regulation, Effort regulation, Peer learning, Help seeking	(Pintrich et al., 1991)
ILS	Likert Part A 1- I do this seldom or never 5- I do this almost always Part B 1- Disagree entirely 5- Agree entirely	120	Face to Face	1998	Self-regulation, External regulation, Lack of regulation, Construction of knowledge, use of knowledge, Stimulation education, Cooperative learning, personally interested, Certificate oriented, Self-test oriented, Vocation oriented, Ambivalent.	(Vermunt, 1998)
LASI	100 points scales 1- Not at all typical of me 5- Very much tipycal of me	88	Online	2002	Anxiety, Attitude and Interest, Concentration, Information processing, Motivation, Self-testing, Selecting main ideas, Study aids, Time management, Test strategies	(Weinstein & Palmer, 2002)
SESRL	Likert 1- Not well at all 7- Very well or very often	35	Face to Face	2002	General Organization and planning, Environment restructuration, external regulation, Recall ability, Typical study strategies	(Garavalia & Gredler, 2002)
OMQ	Likert 1- Disagree 5- Agree	57	Face to Face	2002	Learning strategies, Organization and planning strategies, Processing ability, External regulations strategies, Typical study strategies	(Boekaerts, 2002)
SELF	100 point scale 0%= definitely cannot do it 100%= definitely can do it	57	Face to Face	2007	Reading item, Study item, Test preparation item, Note taking item, Writing item	(Zimmerman & Kitsantas, 2007)
OSLQ	Likert 1- Strongly disagree 5- Strongly agree	24	Online Blended	2009	Goal setting, Environment structuring, Task strategies, Time management, Help seeking, Self-evaluation	(Barnard et al., 2009)
ASRLS	Likert 1- Strongly disagree 4- Strongly agree	55	Face to Face	2011	Memory strategy, Goal setting, Seek assistance, Self-evaluation, Environmental structuring, Learning responsibility, Organizing	(Magno, 2011)
EFLSRLQ	Likert 1- Not important 4- Essential	40	Face to Face	2015	Intrinsic Motivation, Self-efficacy, Attitude, Organization, memory Strategies, Self-monitoring, Planning & goal setting, Effort Regulation, Regulation of environment, Help seeking	(Salehi & Jafari, 2015)
SRL PROFESSIONAL	Likert 1- Not at all true for me 5- Very true for me	39	Online (MOOC)	2016	Goal setting, Strategic planning, Task interest/value, Self-efficacy, Task strategies, Elaboration, Critical thinking, Help seeking, Interest enhancement, Self-evaluation, Self-satisfaction	(Littlejohn, Hood, Milligan, & Mustain, 2016)
SSRL	Likert 1- Never 5- Always	67	Face to Face	2016	Goal setting and planning, environmental structuring, Organization and transformation, Seeking information, Rehearsing and memorizing, Keeping records and self-monitoring, Seeking peer assistance, Reviewing, Self-evaluation, Self-consequences	(Erdogan & Senemoglu, 2016)
SOL-Q	Likert 1- Not at all true of me 7- Very true of me	36	Online (MOOC)	2016	Metacognitive skills, Time management, Environmental structuring, Persistence, Help seeking	(Jansen et al., 2016)

2.2.1.4 Results of Effective SRL Strategies in Online learning

Recent studies have shown the positive relationship between SRL strategies in online environments and academic achievement (Broadbent & Poon, 2015; Broadbent, 2017; Richardson, Abraham, & Bond, 2012). According to these studies, the use of SRL strategies affects the learning outcomes achieved and its use is typically associated with better academic performance in traditional learning contexts (Beishuizen & Steffens, 2011; Dignath & Büttner, 2008; Pintrich, 2004; Zimmerman, 1989) as well as online learning situations (Broadbent & Poon, 2015; Robbins et al., 2004; Wang, Shannon, & Ross, 2013). A total of 16 investigations were analyzed which studied 35 SRL strategies reported between 2004 and 2013 for online contexts and between 2015 and 2017 for MOOCs. 13 of these investigations were conducted on e-learning courses whose student range was between 26 and 1,395 students, and 3 studies were conducted on MOOC courses whose student range was between 2,439 and 50,000 students.

Of the 35 strategies studied, 15 were found to correlate with academic performance (final grades) of the learners, and 5 strategies were found to predict students' grades. It is important to highlight that for one strategy in particular the studies have applied different approaches calling them differently. For example, the strategy of time management has been considered as a strategy for the management of time in the course, as procrastination time and as students' time for interaction with the environment. Something similar happens with the Help-seeking strategy, since it has been considered as a search for help with additional materials, search for information or search for help by consulting other students in forums.

It has been found that 6 of the strategies studied (metacognition as self-regulation, effort regulation, time management, elaboration, organization and self-efficacy) correlate positively with academic achievements or performance (based on the final grades) in at least 2 studies carried out in traditional e-learning courses, and 1 strategy (help-seeking) that correlates positively with academic achievements in MOOC courses and traditional e-learning courses. On the other hand, it was found that 3 of the strategies studied (time management, information processing, help seeking) are predictors of student grades in the traditional courses of e-learning, whereas in MOOC courses 2 strategies (goal setting and strategic planning) are predictors of student grades. Also 10 strategies correlate positively with academic achievements in at least 1 study (metacognition, time management - as procrastination and interaction time, peer learning - as online interaction and online participation, critical thinking, help seeking, intrinsic goal orientation, study environment management - as environment structuring, reflection and feedback, self-monitoring, verbal ability). Table 2-3 shows a summary of the effective SRL strategies related with academic achievements for traditional online courses and MOOCs.

The following 13 studies were conducted in traditional online courses. In a study conducted by Lynch & Dembo (2004) with 94 students, a significant and positive correlation between self-efficacy and course grades and also between verbal ability and course grades was found. In another study conducted by Van den Boom, Paas, & van Merriënboer (2007) with 47 students, it was found that reflective activities combined with feedback from peers or tutor are beneficial for the development of students' SRL and learning outcomes. Chang (2007) in a study with 99 students found that the self-monitoring strategy had a significant effect on students' academic performance and their motivational beliefs. Puzziferro (2008) in a study with 815 students found that time management, study environment and effort regulation were significantly related to performance. Students who scored higher on these SRL subscales received higher final grades. Valle et al. (2008) in a study with 489 students found that the use of elaboration, organization, time and study environment management strategies explain self-regulation moderately for learning and performance. Wang & Wu (2008) in a study with 76 students found that self-efficacy was not related to student academic performance (which is inconsistent with studies like [Pintrich, 2000]), but the results of the study had also shown that self-efficacy significantly predicted students' use of cognitive strategies and related to students' feedback behavior.

Table 2-3 Effective SRL strategies related with academic achievements

Author	Lynch & Dembo	Van den Boom, et al.	Chang	Puziferro	Valle et al.	Wang & Wu	Johnson, Gueutal and Falbe	Hodges & Kim	Michinov et al.	ChanLin	Klingsieck et al.	Chen & Chau	Cho & Shen	Hood, Littlejohn & Milligan	Kizilcec, Perez-Sanagustin & Maldonado	Corrin, de Barba & Bakharia
Year	2004	2007		2008			2009	2010	2011	2012		2013		2015	2016	2017
n=	94	49	99	815	489	76	914	103	83	118	1,395	26	64	50,000	4,831	2,439
SRL Strategies																
Metacognition							✓				x					
Metacognition (as self-regulation)				x	x			x				✓	✓			
Monitoring											x					
Effort Regulation				✓	x							x	✓			
Effort Regulation (Concentration)										x						
Time Management	x			✓	✓					+		x				
Time Management (Procastination)									x		✓					
Time Management (Interaction time)													✓			
Peer Learning										x						
Peer Learning (as online interaction)				x			✓					x				
Peer Learning (as online participation)									✓							
Information processing										+						
Elaboration				x	✓	x					x	✓			x	
Organization				x	✓						x	✓				
Rehearsal				x		x					x	x				
Critical Thinking				x		x						✓				
Help Seeking				x								✓		x	x	✓
Help Seeking (Support material)										+						
Helps Seeking (Information search)											x					
Goal Setting								x						x	+	
Learning Goals					x											
Intrinsec Goal Orientation	x												✓			
Extrinsic Goal Orientation													x			
Study environments management	x				✓							x				
Environment Structucting				✓												
Reflection and Feedback		✓				x										
Self Satisfaction														x		
Self Evaluation (Testing)										x				x	x	
Self Efficacy	✓			x	x	x		✓					✓	✓		
Self Monitoring			✓					x								
Task Strategies														x	x	
Task Value or interest					x									x		
Control beliefs					x											
Strategic Planning								x			x				+	
Verbal Ability	✓															

✓ Correlation with academic achievements, performance (final grades)
\+ Predicts students' achievements/grades
x No positive correlation was found
n Number of students participating in the study

Johnson, Gueutal, & Falbe (2009) in a study with 914 students, worked on the factors that affect the effectiveness of e-learning and found that age, metacognitive activity, and online interaction are related to course performance. Hodges & Kim (2010) in a study with 103 students found a statistically significant relationship between self-efficacy and achievement. Michinov, Brunot, Le Bohec, Juhel, & Delaval (2011) in a study with 83 students, focused on the specific characteristics of a student, such as management of time - as procrastination - and their role in online learning. They found negative relationship between procrastination and performance (this relationship was measured by the level of students' participation in forums). Apparently, if the level of procrastination is high, online students are less successful compared to those who have a lower level of procrastination. ChanLin (2012) in a study with 118 students examined the relationship between the strategies that students use in an online environment and the outcomes assessed using the LASSI OLL questionnaire. Some of its constructs were significant in predicting students' achievement, including: time management, information processing, use of support material. Klingsieck, Fries, Horz, & Hofer (2012) in a study with 1,395 students found that the procrastination predicted worse grades. Cheng & Chau (2013) in a study with 26 students, used the MSLQ questionnaire to evaluate the students' SRL strategies and their achievements. These were evaluated using the scoring system of an ePortfolio. The results indicate that five learning strategies were significantly positively correlated with the ePortfolio scores of the participants (elaboration, organization, critical thinking, metacognitive self-regulation, and peer learning). Cho & Shen (2013) in a study with 64 students, analyzed the role of goal orientation and academic self-efficacy in students' achievement mediated by effort regulation, metacognitive regulation, and interaction regulation in an online course. They found that intrinsic goal orientation and academic self-efficacy were positively associated with students' achievements and these are mediated by three types of regulation (effort regulation, metacognitive regulation, and interaction regulation) but extrinsic goal orientation did not influence students' achievements.

In the case of MOOCs, 3 studies were found that show the relationship between SRL strategies and the performance of students in this type of course. For example, Hood et al.

(2015) in a study with 50,000 registered students found significant differences in the scores obtained in self-efficacy among the students of a MOOC. Specifically, those differences were between students who worked as data scientist (which was the context of the MOOC) and those who did not work as data scientist. Kizilcec, Pérez-Sanagustín, & Maldonado (2017) in a study made with 4,831 students found that goal setting and strategic planning predicted attainment of personal course goals, while helping-seeking appeared to be counterproductive. Corrin, from Barba and Bakharia (2017), in a study conducted with 2,439 students found five help seeking learner profiles which provide an insight into how learners' help-seeking behavior relates to performance in the course.

In MOOCs, few studies that provide information regarding SRL strategies which impact on students' outcomes have been found. However, it is important that the questionnaire constructed takes into consideration those strategies that have been seen, according to the bibliography, to be related to the performance of students in traditional e-learning courses. Therefore, and according to the data reported in the literature, the strategies of effort regulation, time management, elaboration, organization, help-seeking, goal setting, strategic planning and self-efficacy should be considered when creating a new questionnaire. In the following section the construction of the questionnaire is presented based on the background shown in these previous sections. To develop the SRL questionnaire in MOOCs (MOOC-SRLQ), the theoretical model of Pintrich (Panadero, 2017), the strategies used by the questionnaires proposed in the last 25 years (between 1991 and 2016), as well as effective SRL strategies that are related to students' outcomes in online environments will be taken as reference.

2.2.2 Development of a SRL Questionnaire for MOOCs

This section presents the development of the questionnaire to measure the SRL of students in a MOOC and the analyses carried out to validate the instrument. The construction of the instrument was carried out in 3 different studies. The first study used a sample of 3,665 respondents in which an exploratory factor analysis was carried out to examine the adequacy of the proposed dimensions (13) adapted to the context of MOOCs. As a result of the first study, it was proposed to reformulate the 13 initial dimensions based on a theoretical and semantic analysis of the items used. In this step of the study, it was considered if each of the dimensions correlated with the students' performances or were predictors of students' achievements/grades. From this analysis the dimensions were reduced to five. In the second study, using a sample of 485 respondents, a new exploratory factor analysis was conducted to validate the five dimensions. And finally, in the third study using a sample of 477 respondents, a confirmatory factorial analysis was performed, in which the questionnaire was validated with five proposed dimensions. Below, each of the studies performed is detailed (See Figure 2-2).

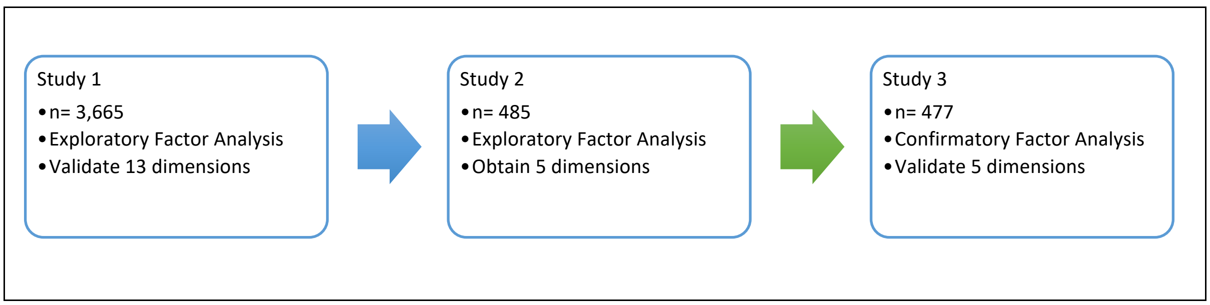


Figure 2-2 Summary of the 3 studies conducted to validate the final questionnaire

2.2.2.1 Study 1: Developing MOOC Self-Regulated Learning Questionnaire

The construction of the questionnaire is based on the Pintrich's model, which classifies SRL strategies as those of cognitive, metacognitive and resource management type. In addition to this, the strategies that have been seen according to the bibliography and are related to the performance of students in traditional e-learning courses will be considered as part of the questionnaire. Those are: effort regulation, time management, elaboration, organization, help-seeking, goal setting, strategic planning and self-efficacy (see Table 2-3). To cover the range of strategies, items from the questionnaires presented in Table 2-2 were extracted and combined. These questionnaires are: SRL Professional, OSLQ, MSLQ, LASSI and EFLSRLQ. The items extracted from these questionnaires had to be adapted for the context of the MOOC. For example, the expression "In a class like this" was changed and contextualized with the expression “In a MOOC” (i.e., “In a class like this, I preferred course material that really challenges me, so I can learn new things” was changed for “In a MOOC, I prefer course material that really challenges me, so I can learn new things”). To answer each of the items a 5-point Likert scale (adapted from the 7-point Likert scale used in the MSLQ, Pintrich et al., 1991) is used, with the ranges (1) - "Not true at all for me" up (5) - "Very true for me", where high scores on this scale indicate better self-regulation in online learning by learners. The resulting preliminary questionnaire consists of 53 items divided into 13 scales to measure different strategies. These are: *self-efficacy, goal setting, strategic planning, rehearsal, organization, elaboration, time management, help-seeking, effort regulation, study environment management, self-monitoring, self-evaluation and self-satisfaction*. Table 2-4 shows a summary of the SRL strategies that constitute the theoretical model of the instrument and the questionnaires from which the items that were adapted during the construction process were extracted.

Table 2-4 Overview of the SRL strategies covered by each of the existing and analyzed SRL questionnaires (Theoretical model)

Strategies	Questionnaires				
	SRL Professional	OSLQ	MSLQ	LASSI	EFLSRLQ
Cognitive					
Rehearsal (4 items)			x		
Elaboration (3 items)			x		
Organization (4 items)			x		
Metacognitive					
Goal Setting (4 items)	x				
Self Monitoring (4 items)				x	
Strategic Planning (4 items)	x				
Self Evaluation (2 items)	x				
Self Satisfaction (2 items)	x				
Self Efficacy (6 items)	x				
Resource Managment					
Help-Seeking (4 items)		x			x
Time Management (6 items)		x		x	
Effort Regulation (4 items)			x		
Study Environment Management (6 items)		x			

Procedure. The data for this study were obtained from three courses offered by Pontificia Universidad Católica de Chile on Coursera. The courses were related to engineering, education and management offered between September 2016 and January 2017. The students of the three courses offered were invited through the e-mail system provided by the platform to complete the questionnaire. This invitation was sent from the first week of the course and each fortnight coinciding with the first work day of the week, in order to reflect on their actual behavior. Informed consent was obtained from all participants before they answered the questionnaire.

All 53 items were presented in random way. It took between 7 to 12 minutes to answer. The participation was totally voluntary. All the procedures for data collection were approved by the ethics committee of the university.

Participants. A total of 3,665 responses were gathered from voluntary participation, of which 33.6% (1,218) reside in Chile, 17.6% (637) in Mexico, and 14.5% (526) in Peru. 45.7% stated that they are between 25 and 35 years old, and 21.8% between 36 and 45 years old. 45.2% stated that they were men, 54.3% stated they were women, and 0.5% (18) preferred not to say. Regarding their education, 89.9% declared having reached at least the tertiary level (higher studies in progress, interrupted, completed and postgraduate). Finally, 30.4% of the sample (1,102) indicated that they had taken at least one MOOC course previously, with an average of 1.58 completed. (median=1, sd=2.922).

Analytic Strategy. Traditionally, factor analysis has distinguished between exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). The EFA is used to try to observe or approach the composition of factors not resolved a priori of the instrument, while the CFA is used when a more complete knowledge about the possible solution has been achieved.

For the case of the present study, although there is no finished knowledge about the final composition of the instrument, the theoretical knowledge is available to propose a tentative structure, in this case, of 13 components (the strategies defined in Table 2-4), so it would be at a point in between. Given the above, the data were analyzed using a 13-component EFA with the FACTOR v.10 software, using the optimal implementation of parallel analysis (PA) technique to determine the appropriate number of dimensions, method for factor extraction by Exploratory Maximum Likelihood (ML), and a Promin rotation. The number of valid observations was 3,269. This selection was based on the fact that parallel analysis is a more effective technique than the traditional Kaiser scree test method to determine the significant components in a factor analysis (i.e., Franklin, Gibson, Robertson, Pohlmann, & Fralish, 1995; Ledesma & Valero-Mora, 2007).

The ML estimation method was adapted to the ordinal scale of the instrument and to the existence of lost data. Finally, an oblique rotation of the Promin type was chosen, which allows the relationship between the factors and presents a solution that authors like Lorenzo-Seva (1999) consider not only effective but also simple. The quality of the instrument was evaluated according to several criteria and indicators suggested by Brown (2014) and Kline (2015), and that began with the extraction of components; followed by a rotated loading matrix with acceptance criteria based on absolute charges greater than .3; and finally, the goodness-of-fit statistics (Chi-square, NNFI, CFI, GFI, RMSEA) and the indicators Simplicity and Construct Replicability were obtained.

Results. The adequacy of the correlation matrix measured by means of the Kaiser-Meyer-Olkin test (KMO) yielded a value of .959, which is considered as very good, while Barlett's sphericity test was χ^2 (1,378) = 87,170, $p < .000$, so the application of a factorial analysis is considered adequate. The results of the extraction of components by means of the Horn's Parallel Analysis method with 500 random samples indicates that an adequate solution should be of 5 empirical dimensions considering the simulation average with random values (Table 2-5), in contrast with the 13 suggested theoretically.

Table 2-5 Parallel Analysis (PA) based on minimum rank factor analysis, 500 random correlation matrices

Variable	Real-data % of variance	Mean of random % of variance	95 percentiles of random % of variance	Variable	Real-data % of variance	Mean of random % of variance	95 percentiles of random % of variance
srl01	29.3**	3.7	4.0	srl28	1.0	1.8	2.0
srl02	5.5**	3.6	3.9	srl29	0.9	1.8	1.9
srl03	5.3**	3.5	3.8	srl30	0.9	1.7	1.8
srl04	4.1**	3.5	3.7	srl31	0.9	1.6	1.8
srl05	3.6*	3.4	3.7	srl32	0.9	1.5	1.7
srl06	3.3	3.3	3.6	srl33	0.8	1.5	1.7
srl07	3.1	3.3	3.5	srl34	0.8	1.4	1.6
srl08	2.7	3.2	3.4	srl35	0.8	1.3	1.5
srl09	2.7	3.1	3.4	srl36	0.8	1.3	1.5
srl10	2.3	3.1	3.3	srl37	0.7	1.2	1.4
srl11	2.1	3.0	3.2	srl38	0.7	1.1	1.3
srl12	1.9	2.9	3.1	srl39	0.7	1.1	1.2
srl13	1.6	2.9	3.1	srl40	0.6	1.0	1.2
srl14	1.5	2.8	3.0	srl41	0.6	0.9	1.1
srl15	1.5	2.7	2.9	srl42	0.6	0.8	1.0
srl16	1.4	2.7	2.8	srl43	0.5	0.8	1.0
srl17	1.3	2.6	2.8	srl44	0.5	0.7	0.9
srl18	1.3	2.5	2.7	srl45	0.5	0.6	0.8
srl19	1.3	2.4	2.6	srl46	0.5	0.6	0.7
srl20	1.2	2.4	2.5	srl47	0.4	0.5	0.7
srl21	1.2	2.3	2.5	srl48	0.4	0.4	0.6
srl22	1.2	2.2	2.4	srl49	0.3	0.3	0.5
srl23	1.1	2.2	2.3	srl50	0.2	0.3	0.4
srl24	1.1	2.1	2.2	srl51	0.1	0.2	0.3
srl25	1.1	2.0	2.2	srl52	0.0	0.1	0.2
srl26	1.0	2.0	2.1	srl53	0.0	0.0	0.0
srl27	1.0	1.9	2.1				

** Advised number of dimensions when 95 percentiles are considered: 4
* Advised number of dimensions when mean is considered: 5

The indicators of goodness-of-fit are generally optimal (NNFI=.990; CFI=.995; GFI=.998, RMSEA=.030) with a Chi-square χ^2 (767) =3,019.38 ($p < .000$). On the other hand, index of factor simplicity were also obtained, which are based on the idea that the commonality of the variables should be related to a reduced number of dimensions, so that the factor load matrix should show zero or close to one (Lorenzo-Seva, 2003). For the case of the instrument presented in this study, Bentler's simplicity index is .0004 and the Loading simplicity index is .5042 (values close to 1 being optimal). Also, the Construct

Replicability was obtained, proposing that a value over .80 (in a range of 0 to 1) suggests that a factor would be stable between studies. From Table 2-6 section "a" it can be seen that the results are heterogeneous, with values from .698 to .944. The sum of these data suggests that while the adjustment indicators – in particular NNFI, CFI, GFI and RMSEA - have a good fit, the results of the Horn's PA, Indices of Factor Simplicity and Construct Replicability show that the 13 dimensions would not be the best solution. This is confirmed by observing Table 2-6 section "b" with the correlations between factors, and Table 2-7 with the factorial loads.

Table 2-6 Construct replicability index and inter-factor correlation matrix

a.- Generalized G-H index

Factor	1	2	3	4	5	6	7	8	9	10	11	12	13
H	.894	.916	.698	.856	.923	.886	.863	.944	.799	.760	.736	.855	.763

b.- Consensus interfactor correlation matrix among multiple imputed datasets.

Factor	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1												
2	.274	1											
3	.190	.016	1										
4	.535	.400	.109	1									
5	.652	.405	.193	.600	1								
6	.520	.430	-.039	.581	.566	1							
7	.608	.257	.209	.439	.484	.435	1						
8	.666	.431	.028	.748	.642	.796	.525	1					
9	.372	.330	-.253	.413	.393	.649	.330	.615	1				
10	.440	.317	-.022	.285	.389	.390	.307	.400	.309	1			
11	-.058	-.236	.152	-.369	-.291	-.213	-.029	-.231	-.166	-.108	1		
12	.487	.349	.023	.683	.430	.579	.500	.714	.481	.393	-.246	1	
13	-.361	-.067	-.315	-.322	-.339	-.192	-.433	-.261	.030	.066	-.042	-.203	1

(1) Self-efficacy (2) Goal setting (3) Strategic planning (4) Study environment management (5) Rehearsal (6) Elaboration (7) Organization (8) Time management (9) Help-seeking (10) Effort regulation (11) Self-monitoring (12) Self-evaluation (13) Self-satisfaction.

Table 2-7 Rotated loading matrix (loadings lower than absolute .300 omitted)

Item	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13
srl01	.787												
srl02	.567												
srl03	.608												
srl04	.818												
srl05	.800												
srl06	.862												
srl07								.350					
srl08				-.330			.311	.433					
srl09							.373	.448					
srl10							.372	.386					
srl11													
srl12												-.312	
srl13													
srl14													
srl15			.739						.691				
srl16			.748						.649				
srl17									.381				
srl18			.391						.824				
srl19									.684				
srl20									.747				
srl21													
srl22								.308					
srl23								.809			-.304		.306
srl24								.850					
srl25					.664			.303				-.309	
srl26				.340	.345							-.333	
srl27					.648							-.352	
srl28							-.489	.660					
srl29							-.316	.606					
srl30							-.435	.479					
srl31							-.402	.577					
srl32						.353							
srl33				.333		.505						.344	
srl34						.390							
srl35							-.447			.377			
srl36													
srl37				-.331			-.349			.351			.317
srl38		.824											
srl39		.899											
srl40		.900											
srl41													
srl42													
srl43						-.510		.333		.317		.354	
srl44													
srl45						-.732		.374				.410	
srl46													
srl47													
srl48													
srl49						.326							
srl50						.396				.392			
srl51						.318							
srl52				-.329	1			-.465				.492	
srl53					.995			-.423				.401	

Discussion. Based on the results obtained, the adequacy of the 13 dimensions was examined and the work was oriented to a simpler proposal more adapted to the MOOC context. Although the results of this first analysis may seem satisfactory in relation to the adjustment index, the size of the sample and the number of variables can produce distortions in them (e.g., Correa, 2010; Hooper, Coughlan, & Mullen, 2008; Iacobucci, 2010; Kenny, 2014; Kenny & McCoach, 2003). The fact that the factorial loads do not coincide with the theoretically expected is added to this, showing loads in more than one dimension or with values below the accepted, which has an impact on poor simplicity and replicability indexes. For this reason, it was decided to perform a theoretical restructuring to approach the solution close to the 5 empirical dimensions suggested by the parallel analysis. A reformulation of the dimensions was then proposed based on a theoretical and semantic analysis of the items used, while integrating the results of the previous study. In this step, it was also considered if each of the dimensions correlates with students' performance or if they are predictors of students' achievement (section 2.2.3). From this analysis emerged 5 dimensions that are: (1) Self-efficacy, (2) Goal setting, (3) Study environment management, (4) Organization and (5) Help-seeking. The dimensions (1), (2), (4) and (3) are correlated with academic achievements, while the dimensions (1), (2) and (5) predict student's achievements/grades.

Dimension (1) Self-efficacy is defined as "individuals' judgments of their abilities to plan and perform the necessary behaviors to achieve specific goals. Learners with high self-efficacy are likely to employ adaptive self-regulatory learning strategies and study skills. When students face a MOOC, they must be able to identify the demands and previous requirements of the course, as well as if the learner is capable to deal with the tasks. Some examples of the statements are "I feel prepared for the demands and requirements of this MOOC" or "I feel that I am capable of studying and learning in a MOOC because I am confident in my skills".

Dimension (2) Goal setting, refers to deciding upon specific outcomes of learning performance, such as completing a major video-lectures in MOOC for deep understanding or attempting to complete most of the assessments in order to get certificated or only attempting to watch some contents. Goals mobilize effort, increase persistence and lead to appropriate use of learning strategies (Schunk, 2000; Garavalia & Gredler, 2002). An example of the statement is “When I study, I set short-term goals (daily or weekly) or long-term goals (for the whole course) for myself”.

Dimension (3) Study environment management is defined as “learners sensitive to their environment, resourceful in creating, altering or changing an environment with less distractions and which facilitates learning can accomplish this (Zimmerman & Martinez-Pons, 1986). Managing adequately the environment shall lead to taking control to change it or leaving it. An example of the statement is “When I study, I choose a place that is conductive to learning and distraction-free”.

Dimension (4) Organization refers to a dimension of reorganizing and elaborating new information in some type of “graphic form” (e.g., creating outlines, mind maps, taking notes), where the forms of graphic organizers have characteristics of active and complex cognitive process (Claire Ellen Weinstein, Acee, & Jung, 2011). In MOOCs, the ability to identify, organize and highlight main points during learning is relevant for the learner (Effeney, Carroll, & Bahr, 2013). An example of the statement is “When I watch the video lectures in a MOOC, I make outlines or summaries of the material to help me organize my ideas”.

Dimension (5) Help-seeking refers to the ability to seek out assistance from peers and teachers when learners encounter a challenge they find too complex to solve (Corrin et al., 2017). It is a key element in SRL, given that in a MOOC environment learners’ autonomy is expected. There are several ways that learners can seek help in a MOOC. The most common is the discussion forum. Here learners can post problems they face with peers and/or teachers looking for a reply. But also, looking for specific information in a video-lecture, or reviewing the answers in assessments are forms of help-seeking. An example of the statement is “I rewind or fast-forward videos in a MOOC to look for specific information on the course topics”.

The dimensions strategic planning, rehearsal, elaboration, time management, effort regulation, self-monitoring, self-evaluation and self-satisfaction were excluded: (1) for theoretical reasons (reformulation of the concept or item given that factor loads do not coincide with what is theoretically expected, evidencing loads in more than one dimension or with values below the accepted), (2) for empirical reasons (the items did not meet the inclusion criteria) and (3) for both theoretical and empirical reasons. The final instrument, consisting of 22 reagents and retaining the response mode using a 5-point Likert scale, was subjected to an EFA using the FACTOR 10 program.

2.2.2.2 Study 2: Reformulation of the SRL dimensions

Procedure. The procedure was the same as described in section 2.2.2.1. The instrument was applied to the participants of 7 courses offered by Pontificia Universidad Católica of Chile on Coursera. The courses were on topics related to engineering, education and management and offered between July 9th until July 31 th of 2017.

Participants. A total of 962 voluntary responses were collected, of which 18.6% (179) came from residents of Chile, 15.6% (150) from Mexico and 12.1% (116) from Colombia. 35% declared being between 25 and 35 years old, and 24.4% between 36 and 45 years old. 57.8% of the sample consisted of men. 91.1% said they had reached tertiary level studies, and 57.1% had previously taken at least one MOOC course, with an average of 2.79 completed (median = 2, sd = 5.5). From the total of 962 responses, we proceeded to randomly divide it into two subsets with approximately 50% of the cases each, being the first of them (n = 485) used for the present EFA of study 2, and the second (n = 477) for CFA for study 3.

Analysis. For this study it was decided to analyze the data through a 5-component EFA with the FACTOR v.10 software, again using the optimal implementation of parallel analysis (PA) technique to determine the appropriate number of dimensions, method for factor extraction by Robust Unweighted Least Squares (RULS), and a Promin rotation. The number of valid observations was 432.

Results. The adequacy of the correlation matrix measured by means of the Kaiser-Meyer-Olkin test (KMO) yields a value of .90718, which is considered to be very good, while Barlett's sphericity test was χ^2 (231) = 4,475, p <.000, which is considered adequate for a factorial analysis. The results of the extraction of components using the Horn's Parallel Analysis method with 500 random samples suggests a solution of 2 empirical dimensions, which differs from the 5 proposed theoretically. The indicators of goodness-of-fit are generally considered optimal and superior to those obtained in the study 1 (NNFI=1; CFI=.1; GFI=.996, RMSEA=.001) y χ^2(131) = 102.321 (p=.969). Table 2-8 section "a" show the factorial loads obtained, which this time coincide with the proposed, an aspect that is also reflected in the increase of the indices of factor simplicity (Bentler's simplicity index = .9967; Loading simplicity index = .7060) and construct replicability (values between .833 and .907). The correlations between factors are also observed more consistent, with correlations of medium magnitude that would account for independent but related dimensions (Table 2-8 section "b").

Table 2-8 Rotated loading and inter-factors correlations

a.- Rotated loading matrix (loadings lower than absolute .300 omitted)					
Item	F1	F2	F3	F4	F5
srl01					.736
srl02					.504
srl03					.609
srl04					.786
srl05					.638
srl06					.667
srl07			.479		
srl08			.770		
srl09			.850		
srl10			.869		
srl11			.711		
srl12		.761			
srl13		.827			
srl14		.785			
srl15	.935				
srl16	.781				
srl17	.752				
srl18	.763				
srl19				.418	
srl20				.434	
srl21				.953	
srl22				.457	

Table 2-8 Rotated loading and inter-factors correlations

b.- Inter-factors correlation matrix

Factor	F1	F2	F3	F4	F5
1	1				
2	.481	1			
3	.565	.448	1		
4	.574	.417	.628	1	
5	.452	.371	.597	.485	1

The results obtained present a 5-dimensional solution well supported by the data. Adjustment indicators improved significantly, the structure of factor loads was simplified, and there seems to be coherence between the theoretical proposal and the empirical results, despite the fact that Horn's PA suggests two dimensions, an aspect difficult to achieve without affecting the very concept of self-regulation of learning as a multifactor behavioral phenomenon. Therefore, it was decided to proceed to the realization of a CFA with an independent sample.

2.2.2.3 Study 3: Confirmatory factor analysis

The objective of study 3 was to test the 5-dimensional factorial structure of study 2, thus the items and the form of response did not vary.

Procedure. The procedure used has been described in section 2.2.2.2

Participants. The sample used has been described in section 2.2.2.2

Analytic Strategy. In this study, a 5-dimensional CFA was performed in the MPLUS v7 software, following the recommendations of Kline (2014) and Brown (2015), the Maximum Likelyhood (ML) estimator and CF-EQUAMAX rotation were used as the most suitable for related latent factors. The quality of the solution will be evaluated through the model fit indicators Chi-square test, RMSEA, SRMR, CFI, TLI; the standardized factorial loads and the R-square estimate for the latent variables will be presented.

Results. Out of a total of 456 valid observations, the model tested has a value for χ^2 (204) = 362.026, p <.000, RMSEA = .041, SRMR = .040, CFI = .962, TLI = .957, which is considered as a good level of adjustment. The standardized factorial loads (Table 2-9) coincide with the proposal and are congruent with the results of the previous study, with values ranging from .489 to .849, all with p <.000. On the other hand, the R-square values for the latent variables range from .336 for Study Environment Management, up to .742 for Help-Seeking.

Table 2-9 CFA loading matrix (standardized model results) and latent variables R^2

Standardized loading matrix					
		Estimate	**S.E.**	**Est./S.E.**	**p-value (2-tailed)**
SE					
	SRL01	.656	.032	20.564	.000
	SRL02	.611	.034	17.782	.000
	SRL03	.597	.035	17.001	.000
	SRL04	.662	.032	20.940	.000
	SRL05	.662	.032	20.965	.000
	SRL06	.728	.028	26.200	.000
GS					
	SRL07	.711	.027	26.762	.000
	SRL08	.805	.020	39.672	.000
	SRL09	.846	.018	47.815	.000
	SRL10	.749	.024	31.106	.000
	SRL11	.700	.027	25.638	.000
SEM					
	SRL12	.799	.028	28.597	.000
	SRL13	.797	.028	28.467	.000
	SRL14	.670	.032	20.691	.000
ORG					
	SRL15	.780	.023	34.137	.000
	SRL16	.831	.020	42.533	.000
	SRL17	.796	.022	36.932	.000
	SRL18	.849	.018	46.404	.000
HS					
	SRL19	.575	.041	14.078	.000
	SRL20	.584	.041	14.399	.000
	SRL21	.652	.037	17.410	.000
	SRL22	.489	.045	10.913	.000
SRL					
	SE	.745	.035	21.435	.000
	GS	.822	.030	27.584	.000
	SEM	.580	.044	13.230	.000
	ORG	.655	.038	17.216	.000
	HS	.862	.038	22.618	.000

Table 2-9 CFA loading matrix (standardized model results) and latent variables R^2

Latent Variables R-Square

Latent Variable	Estimate	S.E.	Est./S.E.	p-value (2-tailed)
SE	0.555	0.052	10.717	0.000
GS	0.675	0.049	13.792	0.000
SEM	0.336	0.051	6.615	0.000
ORG	0.429	0.050	8.608	0.000
HS	0.742	0.066	11.309	0.000

The results of this study are consistent with both the results of EFA of study 2 and the product of the theoretical restructuring that was generated from study 1. The quality of the indicators of adjustment, simplicity and replicability suggest coherence and stability both theoretical and empirical.

2.2.3 Discussion

In appendix A, we present a proposal for a questionnaire to measure SRL in MOOCs. For its design and implementation, a systematic bibliographic review was performed and validated in 3 different studies for an exploratory factorial analysis with 4,627 voluntary responses. The result is a questionnaire with 22 questions that considers the 5 dimensions of SRL: Self-efficacy, Goal setting, Study environment management, Organization and Help-seeking. One of the added values of this article is, in addition to the instrument, an exhaustive and systematic bibliographic review on the work done on instruments to measure SRL and related works in recent years. This bibliographic review serves not only to demonstrate the limitations of existing instruments to analyze self-regulatory profiles in MOOCs, but also to systematically organize the prior work conducted in this line. The conclusions of this bibliographic review indicate that the existing instruments: (1) are too long regarding the number of questions, which affects the response rate in a negative way; (2) have been developed for traditional contexts, and not for that of MOOCs; (3) the instruments designed for an online environment must be adapted because they are based on the technology of the 90's; and (4) do not include as part of their items the strategies that are relevant for this type of context. Therefore, this subsection presents two main contributions to the study area of self-regulation in MOOCs: (1) it contributes to the systematic organization of instruments designed to measure SRL based on the different existing models for online environments and to demonstrate their limitations for MOOCs; and (2) it proposes an instrument that researchers and practitioners can use to measure the profile of self-regulation of MOOC students.

2.3 A Process Mining methodological approach for extracting SRL strategies in Coursera MOOCs

In recent years, masses of fine-grained educational records have become available to researchers and accelerated the nascent field of Learning Analytics (Dietze, Siemens, Taibi, & Drachsler, 2016). Digital learning platforms collect detailed records of each learner's behavior, performance, and other types of interaction. In particular, MOOCs are a major source of data on learner behavior and they enable research to gain a better understanding of how individuals learn in online learning environments (Breslow et al., 2013; Cooper & Sahami, 2013; Daradoumis, Bassi, Xhafa, & Caballe, 2013). Nevertheless, despite the large amount of data that MOOCs are collecting, this information may not be sufficient to build on educational theories and develop new ones. In particular, access to critical information about learners' behavior and learning processes is frequently limited. Data-driven methods can rapidly extract patterns in what learners do throughout a course, but it remains a challenge to interpret the patterns and understand how they relate to theory. One approach to increase the interpretability of large amounts of clickstream data is to triangulate them with other data sources (i.e., taking a mixed-methods approach). For example, clickstream data from MOOCs, which capture learners' actual interactions, can be combined with data from self-reported instruments such as questionnaires or think-aloud sessions (Bannert et al., 2014), or data from external sources like eye-tracking (Trevors, Feyzi-Behnagh, Azevedo, & Bouchet, 2016). To get a better understanding of how learners behave and learn in digital environments there is a need to explore ways to connect educational theory to data-driven methods with behavioral and self-reported data (Lodge & Corrin, 2017).

In this section, we use MOOC data to advance the research of SRL in online context. Recent studies show that for MOOC learners in order to achieve their objectives, they must have the capacity to regulate their own learning (Hew & Cheung, 2014; Kizilcec & Schneider, 2015) or receive active self-regulation support from the platform (Kizilcec & Cohen, 2017). In the absence of the support and guidance that is typically available in brick-and-mortar learning environments (e.g., an instructor setting deadlines and structuring the learning process), the ability to regulate one's learning process is a critical skill to achieve personal learning objectives in a MOOC. Online learners need to determine when and how to engage with course content without any other support than the course content and structure, which can pose a challenge for many learners (Lajoie & Azevedo, 2006). Self-regulated learners are characterized by their ability to initiate cognitive, metacognitive, affective and motivational processes (Boekaerts & Corno, 2005). Moreover, SRL research indicates that successful learning is associated with the active deployment of regulatory activities during the learning process, such as goal-setting, planning or monitoring (Bannert et al., 2014). The ability to develop these learning strategies is an essential skill in order to succeed in an open context such as a MOOC, where the learner should advance independently without support from a tutor or professor. However, how people self-regulate in a MOOC is still an open question.

Over the last 30 years, multiple models have been developed to explain how the process of SRL develops amongst learners (Panadero, 2017). These models serve as a foundation for developing methods to study the use of SRL strategies in the learning process. They can be categorised as either component models or process models (Wirth & Leutner, 2008). Component models describe SRL in terms of different strategies that promote or encourage self-regulation, which are seen as long-lasting characteristics of a person. Process models describe typical requirements that learners have to meet in different phases of the cyclical learning process, but they do not specify the strategies necessary to meet those requirements (Zimmerman et al., 2000). Researchers in the field of SRL have suggested that questions about measuring constructs associated with self-regulation should be seen in terms of aptitudes (for component models; Bannert et al., 2014) and events (for process models; Winne, 2010). Thus, both learner aptitudes and events contribute to a global understanding of how SRL works. On the one hand, aptitudes are essential to researching SRL since they are theoretical constructs underlying observed differences between individual learners in specific contexts such as motivational factors and epistemic beliefs (Snow, 1989). On the other hand, events are the actions that learners perform and provide touch points to map information in order to infer learners' cognitive processes (Winne, 2010).

Prior research studying SRL in MOOCs identified learner characteristics that are predictive of stronger SRL skills based on clickstream behavior data and a survey instrument (Kizilcec et al., 2017). This subsection extends these findings by leveraging process mining methods with the clickstream data collected in three MOOCs. In particular, this subsection focuses on the relationship between the trace-data generated through the interaction of learners with the course content (video-lectures and assessments) in online sessions and learners' self-reported SRL skills. Mukala, Buijs, Leemans and Van Der Aalst (2015) found that learners interact with video-lectures, assessments and other MOOC contents week by week, identifying loopbacks, deviations and bottlenecks. The current investigation additionally incorporates data on learners' assessment submission behavior. In this subsection, formal Process Mining (PM) techniques are used in order to go deeper (looking for broad interaction sequences) and understand the relationship between theoretical self-reported SRL strategies and behavioral patterns on large-scale MOOC platforms. Specifically, an analysis of learners' behavior sequences in a MOOC from a PM perspective could enable us to understand how observed interaction sequence patterns are aligned with SRL strategies. To this end, in this subsection the results of an exploratory sequence analysis to detect patterns in learner's behavior and combining with their SRL profile scores are presented.

2.3.1 Related Work

This subsection contains the bibliographic review regarding how SRL strategies have been studied in online environments. Specifically, this section presents the study of SRL strategies as as set of interaction sequences patterns in online environments and the relation between the SRL profile and academic performance.

2.3.1.1 Self-Regulated Learning in Online Environments: Interaction sequences patterns

Several studies have demonstrated a positive relationship between the use of SRL strategies in online environments and academic achievement (Broadbent & Poon, 2015; Broadbent, 2017; Richardson, Abraham, & Bond, 2012; Robbins et al., 2004; Wang, Shannon, & Ross, 2013). Most research on SRL in online environments adopts an aptitude-based approach where self-reported questionnaire is the most common type of assessment for SRL. Questionnaires assess cognitive, metacognitive and resource management strategies use in order to identify specific learning strategies or tactics. Moreover, self-reports are feasible for large-scale assessment where observational methods are impractical (Roth et al., 2015). In general, these questionnaires can be used to establish aptitude-based SRL profiles for learners: for example, to distinguish between highly self-regulated and less self-regulated learners.

In recent years, there has been a boost in research to understanding SRL in online environments, in particular research that investigates SRL as a process. This is in part due to advances in digital learning environments that can record learner behavior at a fine-grained level (e.g., information collected from a learner's interactions with the course content such as video-lectures or assessments). The aptitude-based approach to studying SRL has relied on questionnaires that reflect a static image of SRL. Yet SRL is a dynamic process sensitive to the specific context where learners perform a task. Thus, the process-based approach offers an opportunity to overcome some of the shortcomings of the aptitude-based approach and self-report instruments (Jovanović et al., 2017).

From this process-based perspective, SRL can be conceived as a set of events or actions that learners perform when they are studying (learning traces), rather than a description of those actions or mental states that these actions generate (Bannert et al., 2014). Recording the context of each trace is possible to obtain a representation of the performed behavior without asking a learner about it (e.g., as with think-aloud methods) (Winne, 2013). In this sense, PM is a suitable approach for studying SRL in online environments from a process perspective. Specifically, PM facilitates the discovery of learning process models, which represent the sequence of learners' interactions with course materials (van Der Aalst, 2011). It also provides robust ways of extracting, analyzing and visualizing learners' interaction traces (Jivet, 2016; Mukala et al., 2015; Romero, Cerezo, Bogarín, & Sánchez-Santillán, 2016). These interaction traces are temporal sequences of events of learners' behavior in the online environment that allow tracing of aptitudes in natural settings (Winne, 2014). For example, Hadwin et al. (2007) examined the performance of eight learners across two study sessions on the gStudy platform. They compared traces of actual study activities to self-reporting on SRL and found that students' self-reports may not align with actual studying activity. More recently, Beheshitha et al., (2015) examined the relationship between 22 undergraduate learners' self-reported SRL aptitudes—such as achievement goal orientation and learning approaches—and the strategies they followed in a learning environment on the nStudy tool. They found differences in transitions between the SRL cognitive strategies performed by both "deep" and "surface" learners. Sonnenberg and Bannert (2015) analyzed sequential patterns in the learning process of 70 undergraduate students in an online environment. They found that using metacognitive prompts to support learners' SRL had an effect on the order in which they participated in learning activities. In a recent experiment in an online environment designed to support SRL at the workplace, Siadaty, Gašević, and Hatala (2016) analyzed trace data to build a transition graph of learning actions of 53 learners, where they show that promoting social awareness strongly influenced with the micro-level processes of SRL of the learners.

This prior work demonstrates the potential of taking a PM approach to study SRL, but there are some notable limitations that need to be addressed. First, the small sample sizes and homogeneity of study participants limits the generalizability of prior findings. Second, participants were unfamiliar with the digital learning tools that were developed to assess SRL and their learning experience with these tools may not have been realistic. It is preferable to study diverse learners' interaction traces and SRL at larger scale and in naturalistic online learning settings. Much research on SRL in online environments has been done on platforms that were either manipulated or adapted to study SRL, by adding functionalities that were associated with a self-regulated strategy (Beheshitha et al., 2015;

Siadaty et al., 2016; Sonnenberg & Bannert, 2015). The use of designated learning platforms to study SRL provides greater experimental control and flexibility in measurement at the expense of external validity. To study learning paths we consider different levels of interaction granularity by which we denote the number of events that occur over time in an interaction sequence (Bannert et al., 2014). The granularity in the interaction sequence can be studied in terms of learning trajectories that learners follow based on the content structure of a MOOC (e.g., a linear trajectory going from one week to the next). Granularity can also be studied in terms of learners' interaction sequences with specific objects in the course, that are part of a learning activity (e.g., learning trajectories between video-lectures, assessments, discussion forums, etc.). Thus, the data gathered can help us gain insights into how learners engage with the course content and provide more information about tactics and strategies that might be useful when studying. Accordingly, we defined the next sub research question as follows: *Sub-RQ1.2 - What are the most frequent interactions sequences of learners in MOOCs?*

2.3.1.2 Self-Regulated Learning Strategies

Self-regulated learning is a very complex process that involves both psychological and behavioral changes. Beside these psychological processes, self-regulated learners must have the ability to initiate behavioral changes in order to take the necessary actions to achieve their learning goals and persevere until they succeed. These behavioral changes manifest as a set of actions or strategies in which learners set goals, attempt to monitor, regulate and control, guided and constrained by their goals and contextual features of the learning environment (Pintrich, 1999). However, observing SRL strategies, even when these manifest as a set of actions and behavioral changes, entails several challenges.

The first challenge is to identify and observe behavioral changes. Even in an online environment, where learners' actions are registered, we are not capturing all the actions involved in learners' learning process. Certain strategies, such as goal setting or help seeking, might be occurring beyond the learning platform. For example, we do not know that MOOC learners complement their learning process with social networks (Chen, Davis, Lin, Hauff, & Houben, 2016; García-Peñalvo, Cruz-Benito, Borrás-Gené, & Blanco, 2015). However, we do not know when this behavior occurs within the learners' learning process and how this relates with SRL strategies. The second challenge is to understand whether an observable behavior relates to a particular SRL strategy or to more than one. For example, is possible to say that when a learner spends a study session watching video-lectures in a MOOC, it could be related to the *Study* strategy as defined by Garavalia and Gredler (2002) ("Study in a particular order"), or as *Rehearsal* as defined by Broadbent (2017) (e.g., "Learner who listens to an online lecture repeatedly"). Moreover, researchers agree that SRL is not a fixed trait, but rather a skill that can be developed through personal experiences and practice applying learning strategies (Azevedo & Cromley, 2004; Schunk, 2005; Zimmerman, 2015). This means that an observable behavior at the beginning of the course may be related to a different strategy when it is observed at the end of the course.

To address these challenges, some researchers have made an effort to associate certain behavioral patterns with learning strategies. For example, Hadwin and Winne (2012) analyzed the learning outcomes of a set of learners when applying certain strategies. They observed that individuals who apply relevant learning strategies would act more strategically and intentionally than the others, such as recalling related prior knowledge and cognitively manipulating new information to connect with their prior knowledge in order to improve retention. Jovanović, Gašević, Dawson, Pardo and Mirriahi (2017) observed that those learners' adopting the learning strategies aligned with teachers' teaching strategy were more successful in online course. This prior work, together with the studies identifying interaction sequences in online environments, shed some light on how to relate observed behavior with learning strategies. However, how MOOC learners' actions and behavior relates with SRL strategies as defined in the theory is still unclear.

2.3.2 Method for extracting SRL strategies in Coursera MOOCs

2.3.2.1 Sample

The final study sample comprised N = 3,458 online learners in three different MOOCs (see 2.4.2.1 Courses). This sample is a subset of 4,871 respondents who answered the initial questionnaire among the 54,935 learners who registered for the MOOCs. We excluded 1,413 responses for one of the following reasons: (1) learners took the survey more than once in the same course (n = 733), (2) empty surveys without answers (n = 133), and (3) survey data could not be linked to platform data (n = 547). The target audiences of the three courses were high school students, college students, and professionals in subject-related industries. Based on the demographic data captured during the registration process on the platform, the average age was 32.0 (SD = 11.07). One quarter of learners were women and 88% held a bachelor's degree or higher (14% a master's or Ph.D.). Data collection occurred between April and December 2015.

2.3.2.2 Courses

This study encompassed three courses offered by Pontificia Universidad Católica de Chile on Coursera. The courses were taught in Spanish on topics related to engineering (n = 2,035 in final study sample), education (n = 497) and management (n = 926). The course materials were organized into different modules, each one composed of several lessons. Each lesson included 9 to 17 video-lectures and assessment activities (only summative). Table 2-10 shows the number of enrolled learners, passing rate, modules, lessons, video-lectures, and assessment activities in each course. The courses followed an on-demand format in which course materials were available all at once without specific predefined deadlines. Figure 2-3 illustrates the structure of each course.

The courses are structured in modules, and each module is composed of lessons. Each lesson includes video-lectures and assessment activities (summative). The '*' represents a video-lecture or assessment activity in each lesson.

Table 2-10 Overview of the MOOCs in our study

	MOOC 1	MOOC 2	MOOC 3
	(n = 497)	**(n = 2,035)**	**(n = 926)**
Enrolled	18,653	25,706	10,576
Passing Rate	1.40%	8.40%	11.40%
Modules	9	4	7
Lessons	9	17	13
Video-lectures	48	83	51
Assessments	7	16	6

MOOC 1: Aula Constructivista	Video-Lecture	Assessm.
Module 1		
Lesson 1	* *	
Module 2		
Lesson 1	* * * * *	
Lesson 2		*
Module 3		
Lesson 1	* * * * * *	
Lesson 2		*
Module 4		
Lesson 1	* * * * * * * *	
Lesson 2		*
Module 5		
Lesson 1	* * * * *	
Lesson 2		*
Module 6		
Lesson 1	* * * * * * *	
Lesson 2		*
Module 7		
Lesson 1	* * * * * * * * *	
Lesson 2		*
Module 8		
Lesson 1	* * * * *	
Lesson 2		*
Module 9		
Lesson 1	*	

MOOC 2: Electrones en Accion	Video-Lecture	Assessm.
Module 1		
Lesson 1	*	
Lesson 2	* * * *	*
Lesson 3	* * *	*
Lesson 4	* *	*
Lesson 5	* *	*
Module 2		
Lesson 1	* * * * * * * * * *	*
Lesson 2	* * * * * * * * *	*
Lesson 3	* * * * * * * *	*
Lesson 4	* * * * * * * *	*
Module 3		
Lesson 1	* * * * *	*
Lesson 2	* * * *	*
Lesson 3	* * * *	*
Lesson 4	* * * *	*
Module 4		
Lesson 1	* * * * * * *	*
Lesson 2	* * *	*
Lesson 3	* * * * *	*
Lesson 4	* * * *	*

MOOC 3: Gestion de Organizaciones	Video-Lecture	Assessm.
Module 1		
Lesson 1	* *	
Module 2		
Lesson 1	* * * * *	
Lesson 2	* *	*
Module 3		
Lesson 1	* * * * *	
Lesson 2	* *	*
Module 4		
Lesson 1	* * * * * *	
Lesson 2	* *	*
Module 5		
Lesson 1	* * * * * * * * * * *	
Lesson 2	* *	*
Module 6		
Lesson 1	* * * * *	
Lesson 2	* *	*
Module 7		
Lesson 1	* * * * *	
Lesson 2	* *	*

Figure 2-3 MOOCs Structure. Each MOOC contains modules and each module contains lessons. Each lesson is composed either a video-lecture or an assessment indicated with "*"

2.3.2.3 Measures

Learners in the three MOOCs completed an optional questionnaire at the beginning of the course. The questionnaire included items related to demographic measures (age, gender, education) and learners' intentions in the course (to watch all lectures or only some of them). In addition, the questionnaire included the Online Learning Enrollment Intentions (OLEI) scale (Kizilcec & Schneider, 2015) translated into Spanish, and a measure of SRL (Kizilcec, Pérez-Sanagustín, & Maldonado, 2016). The SRL measure consisted of 24 statements related to six SRL strategies. Learners rated statements using a 5-point scale (coded from 0 to 4). The six SRL strategies that were assessed are goal-setting strategies (4 statements), strategic planning (4), self-evaluation (3), task strategies (6), elaboration (3) and help-seeking (4). The reliability of the questionnaire was obtained. For each strategy, the individual score was computed by averaging ratings of corresponding statements. The SRL measure exhibited high reliability for all strategy subscales with Cronbach's alpha of at least 0.70, which is generally considered acceptable (Peterson, 1994). The SRL composite, an index of all six subscales, had very high reliability (α = 0.91). Table 2-11 presents descriptive statistics for each SRL strategy and composite, also the Cronbach's α, Pearson's correlation coefficients between strategies.

Table 2-11 Overview of the MOOCs in our study

Strategy	***M* (*SD*)**	**α**	**2.**	**3.**	**4.**	**5.**	**6.**	$\bar{x}$
1. Goal Setting	3.02 (0.75)	.86	.70	.46	.57	.46	.29	.78
2. Strategic Planning	3.11 (0.64)	.73		.60	.65	.58	.31	.84
3. Self-evaluation	3.28 (0.65)	.79			.62	.60	.24	.73
4. Task Strategies	3.10 (0.62)	.78				.72	.34	.87
5. Elaboration	3.31 (0.63)	.76					.32	.77
6. Help Seeking	2.62 (0.78)	.75						.58
$\bar{x}$ SRL Composite	3.06 (0.52)	.91						

2.3.2.4 Procedure

We used the Process Mining PM^2 method (van Eck et al., 2015), which is a simpler and more flexible adaptation of other PM methods such as the L*Life-cycle model (van Der Aalst, 2011). The PM^2 method is structured into four stages (Figure 2-4): (1) extraction - the data is extracted from the Information System data bases (Coursera in our case), (2) event log generation – the table value information is modeled in terms of event logs, defining the concepts of case (execution of a process), activities (steps of the process), and temporal order of the activities, (3) model discovery – process mining discovery algorithms are applied to the event log in order to automatically mine a process model describing the observed behavior of the process, and (4) model analysis – the discovered process models are analyzed in order to understand the observed behavior. This method was selected because it is the one used in disciplines such as healthcare and business to understand users' interactive workflows within a particular system (Chaves & Córdoba, 2014; Rojas, Munoz-Gama, Sepúlveda, & Capurro, 2016). It is also suitable for the analysis of both structured and unstructured processes (van Eck et al., 2015).

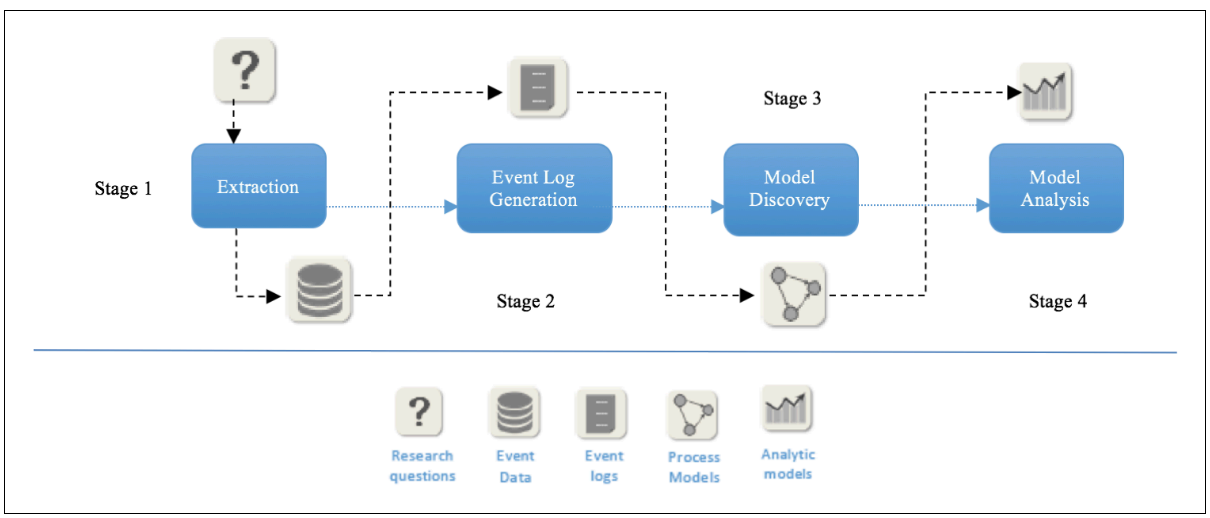


Figure 2-4 Stages for the generation of the process model using PM^2 methodology.

Extraction Stage. In this stage, we extracted the trace data from Coursera database in order to study the interaction sequences of learners in MOOCs. Coursera is a large platform that keeps track of almost all details of student interactions. This raw data is organized into three categories: general data, forums and personal data. It comprises 86 tables of information. For the purpose of this study, we have limited our analysis by selecting only thirteen tables that contain relevant information about students' behavior. The datasets extracted include course information, course content, course progress, assessments, course grades and learner demographics (based on user surveys).

Event Log Generation Stage. In this stage, we defined the event log file we used in the PM algorithm. This event log is a file that stores the information on the learners' interactions within the MOOC, their SRL scores, as well as information necessary to perform the analysis such as the case id, time stamp and other resources. The first step for generating the event log file was to define different concepts to refer to the trace data registered in the Coursera databases. Specifically, we defined the concepts of ***interaction*** and ***session*** as follows:

- An ***interaction*** is an action recorded in the Coursera trace data that registers the interaction of a learner with a MOOC object. We defined six types of interactions depending on the objects that learners interact with: start a video-lecture, complete a video-lecture, review a video-lecture already completed, try an assessment, pass an assessment, and review an assessment already passed. In addition to these interactions, we also included a label to identify the first and last interaction of the learner with the course as *begin session* and *end session*, respectively. All interactions of the learners with the MOOC content extracted from the events log are listed in Table 2-12.
- A ***session*** is a period of time in which the Coursera trace data registers continuous activity of a learner within the course, with intervals of inactivity no greater than 45 minutes. This definition of session was adopted from the prior works by Kovanović et al., (2015) and Liu et al., (2015).

Table 2-12 Definitions of six interaction types with course materials to characterize consecutive learner behavior

Interaction	Definition
(1) Video-Lecture begin	Begin watching a video-lecture without completing it. The video-lecture was not previously completed.
(2) Video-Lecture complete	Watch a video-lecture in its entirety on the first attempt.
(3) Video-Lecture review	Go back to a video-lecture that the learner had previously watched in its entirety (not necessarily on the first attempt).
(4) Assessment try	Unsuccessful attempt to solve an assessment.
(5) Assessment pass	Successful attempt to solve an assessment for the first time.
(6) Assessment review	Go back to an assessment that was previously completed successfully (not necessarily on the first attempt).

In addition to the interactions, the event log file included the learners' SRL scores that we obtained from the SRL self-reported questionnaire. Finally, the event log also included whether the learner completed the course or not: a) True (finished the course), or b) False (did not finish the course). All this information is included in the event log for each session and learner. Therefore, the result of this stage is a log of events documenting the learners' interactions with the course content within a session, their SRL scores, completion of the course, and other complementary data to identify the session ID, the event ID and the timestamp in which each registered event was produced. Table 2-13 shows an example of the event log generated.

Table 2-13 Example of the event log generated for the process analysis

Case ID	Time Stamp	Interaction	SRL Scores	Course completion	Session
acc92cf40b27	1451023929	Begin session	3.162	False	1
acc92cf40b27	1448567431	Video-Lecture.begin	3.162	False	1
acc92cf40b27	1448567737	Video-Lecture.complete	3.162	False	2
acc92cf40b27	1448568139	Assessment.try	3.162	False	2
acc92cf40b27	1449103918	Video-Lecture.review	3.162	False	1
011ff41dfa72	1449104348	Assessment.pass	3.433	True	1
011ff41dfa72	1449104694	Assessment.review	3.433	True	2
011ff41dfa72	1449105157	End session	3.433	True	1

Discovery of the model. We processed the event log with a discovery algorithm to obtain a process model representing the behavior of the learners within the MOOC. In the PM literature, there is a wide range of discovery algorithms that can be used to identify interaction patterns (van Der Aalst, 2011). Given our situation, we selected the Disco algorithm (Günther & Rozinat, 2012) and Celonis algorithm and their implementations in the Disco and Celonis commercial tools. With some differences, both algorithms are based on the Fuzzy algorithm concept (Günther & Aalst, 2007) combined with some characteristics from the Heuristic algorithm family (van Der Aalst, 2011).

Both algorithms were specially designed to handle complex processes, such as learner interactions in a MOOC, and they result in process-map models that can be operated and understood by domain experts with no previous experience in PM. Finally, both commercial tools integrate a set of metrics and filtering options to adapt the event log to the specific questions and to analyse the process interactively. We used Disco and Celonis to generate initial process models for analysis.

Model analysis. Once the process model was generated, we analyzed and identified learners' most frequent ***interaction sequences***. An ***interaction sequence*** is defined as a set of concatenated interactions (from one interaction to another) of the same learner within a session. That is, the path that a learner follows through the MOOC content within a session. The interaction sequences were first used for an exploratory analysis and then for clustering. As a result of applying the algorithms, we obtained a *spaghetti process* model (Figure 2-5). The *spaghetti process model* is a term used in the PM field to refer to a model with so many arcs and crossings that it is difficult to understand or observe patterns. This process model is composed of a start-point and an end-point represented with a white hexagon with a play image and a stop image inside, respectively. The interactions in Table 2-12 are represented with a coloured filled hexagon.

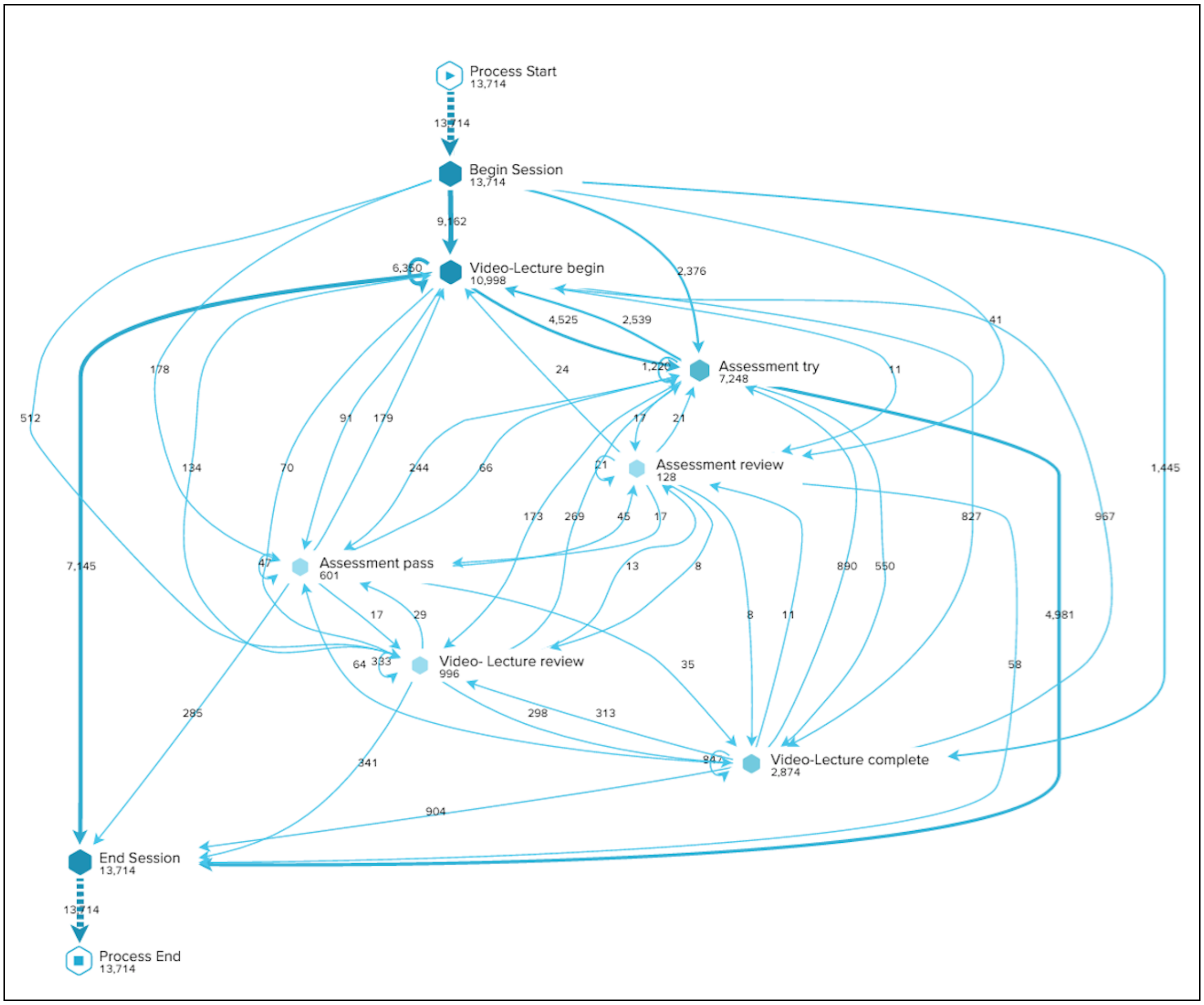


Figure 2-5 Spaghetti process model containing all interaction sequences of 3 MOOCs by sessions

The arcs and arrows connect two or more interactions into what we call *interaction sequences* that were repeated by different learners. For example, an interaction sequence would be from *Begin session* to *(→) Video-lecture-begin* to *(→) End session*, which indicates that a learner began a session, then watched a video-lecture and then ended a session; or from *Begin session* to *(→) Video-lecture-begin* to *(→) Assessment-try* to *(→) End session*, which indicates that a learner began a session, then began a video-lecture, then attempted an assessment and then ended a session. Figure 2-6 shows a subset of interaction sequences extracted from the main process model to provide a better explanation about its semantics. The process model also contains numbers next to each hexagon. These numbers mean the amount of times the interaction inside the hexagon was repeated across all sessions in the dataset. For example, Figure 2-6 shows that the event log contains 13,714 *Begin session* interactions; that is, there were 13,714 sessions registered in the dataset. The numbers over the arcs with arrows means the amount of interaction sequences from the two interconnected interactions that have been identified within a session, and the arrows indicate the direction.

Figure 2-6 shows that the interaction sequence from *Begin-session* to *(→) Video-lecture-begin* was performed 9,162 times. This means that from the 13,714 sessions that were initiated, only 9,162 interaction sequences were performed toward *Video-lecture-begin*. The process model contains the six possible interactions that learners can perform with the course content like video-lecture begin, video-lecture complete, video-lecture review, assessment try, assessment pass, assessment review. Also, the process model specifies the number of sessions that start (begin session) and end (end session).

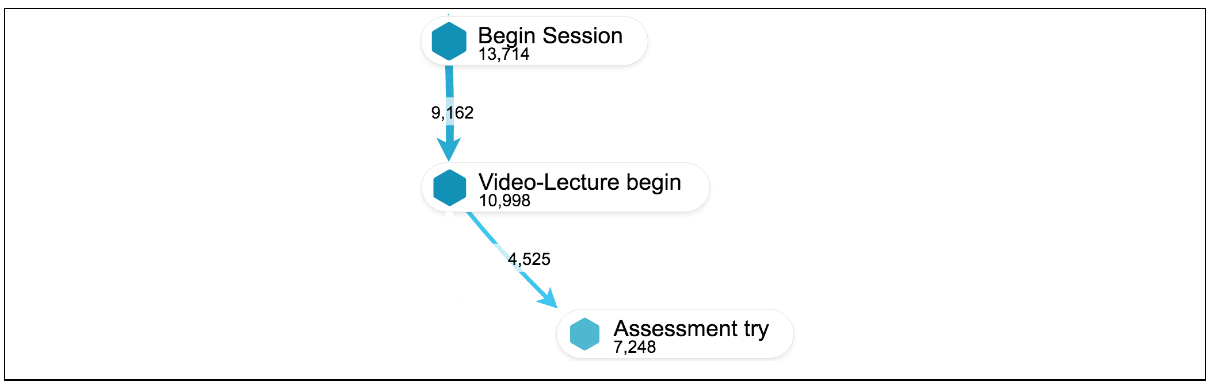


Figure 2-6 Representation of interaction sequences extracted from the spaghetti full process model

This extract of the process model shows that the interaction sequence from Begin-session to (→) Video-lecture-begin was performed 9,162 times and the interaction sequence from Video-lecture begin to (→) Assessment try was performed 4,525 times. Also, the numbers under the interaction caption next to each coloured hexagon indicates the number of times the interaction caption was repeated. For this case 10,998 times for Video-lecture begin interaction and 7,248 times for Assessment try interaction. Once the process model was generated, we applied filters to the event log in order to obtain more specific process models and extract information about the learning strategies deployed by learners.

2.3.3 Results

We analyzed the process models in the model analysis stage to identify the most frequent interaction sequence patterns. First, we analyzed the models, considering all the data from the three courses. Second, we analyzed the data from each course separately. After having identified the most common interaction sequence patterns among MOOC learners in a session, we analyzed how these patterns vary according to whether or not learners completed the course. To achieve this, we filtered the log file by completer (n = 258) and non-completer (n = 3,200) status. This allowed us to observe differences between the various interaction sequence patterns. We also generated process models for completers and non-completers. Then we use an agglomerative hierarchical clustering technique for grouping learners (N = 3,458) based on the identified interaction sequence patterns (e.g., learning strategies). That is, we cluster learners based on their distinct use of learning strategies. We use the scores obtained through the self-reported SRL questionnaire in order to observe how learners are distributed across the different clusters.

Sub-RQ 1.2 - What are the most frequent interactions sequences of learners in MOOCs?

We generated the process model shown in Figure 2-5 based on 13,714 sessions. There were 1,956 different types of sessions, each containing a set of interaction sequences that characterized the session. Figure 2-7 shows a screenshot of the Disco software, which provides a list of the 1,956 types and an overview of its related interaction sequences. In Figure 2-7, the type 21 (variant) shows 4 interactions (events) with 3 interaction sequences and the time associated with the duration of the session. The types of sessions were ordered from the most common to the least common.

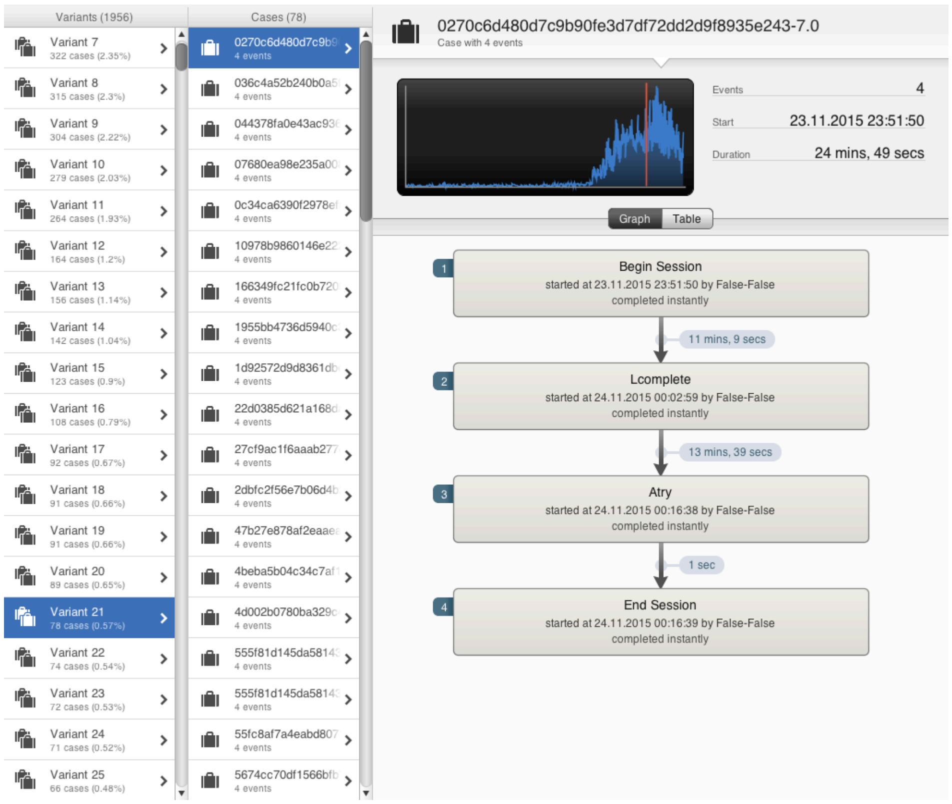


Figure 2-7 List of the 1,956 types of the sessions obtained using Disco software

The most common we assigned to a category that describes the interaction sequence pattern. For example, we analyze the first most common types of sessions and we observed that these consists in video-lecture begin interaction sequences. So, a pattern of Only video-lecture is defined. Then, we filtered the log file marking these types of sessions. After that, the procedure is repeated, identifying the rest of the sessions types that remains without mark in the log file. It was done through a python script developed. As a result, we obtained the following seven interaction sequences patterns:

1. Only Video-lecture: 2,539 repetitions of the type of session
2. Only Assessment: 604 repetitions of the type of session
3. Explore: 583 repetitions of the type of session
4. Assessment try to Video-lecture: 304 repetitions of the type of session
5. Video-lecture complete to Assessment try: 78 repetitions of the type of session
6. Video-lecture to Assessment complete: 15 repetitions of the type of session

7. Others: 3 repetitions of the type of session

Those types of sessions that fit into multiple interaction sequence patterns (given that they are long and disperse) or they do not fit into any interaction sequence pattern, were classified as "Others". The description of each interaction sequence pattern is based on whether a session only contains certain type of interaction (defined in Table 2-12) or whether the session contains certain type of interaction sequences between interactions that are important in the learning process (for example, pass from try an assessment to a video-lecture which represents how the learner looks for missing information after not passing the assessment). Once the most common interaction patterns were extracted from the main process model (Figure 2-5), we defined for each pattern a process model (Figures 2-8, 2-9, 2-10, 2-11, 2-12, 2-13 and 2-14), in order to observe the learner behavior as a result of the interaction with the MOOC content in a session. We described the seven distinct interaction sequence patterns extracted by PM as follows:

(1) *Only Video-lecture*: interaction sequence pattern dedicated only to watching video-lectures, in which the most common interaction sequences are *Begin session* to *video-lecture-begin* or *video-lecture-complete* or *video-lecture-review* and combinations of them before *End session* (Figure 2-8).

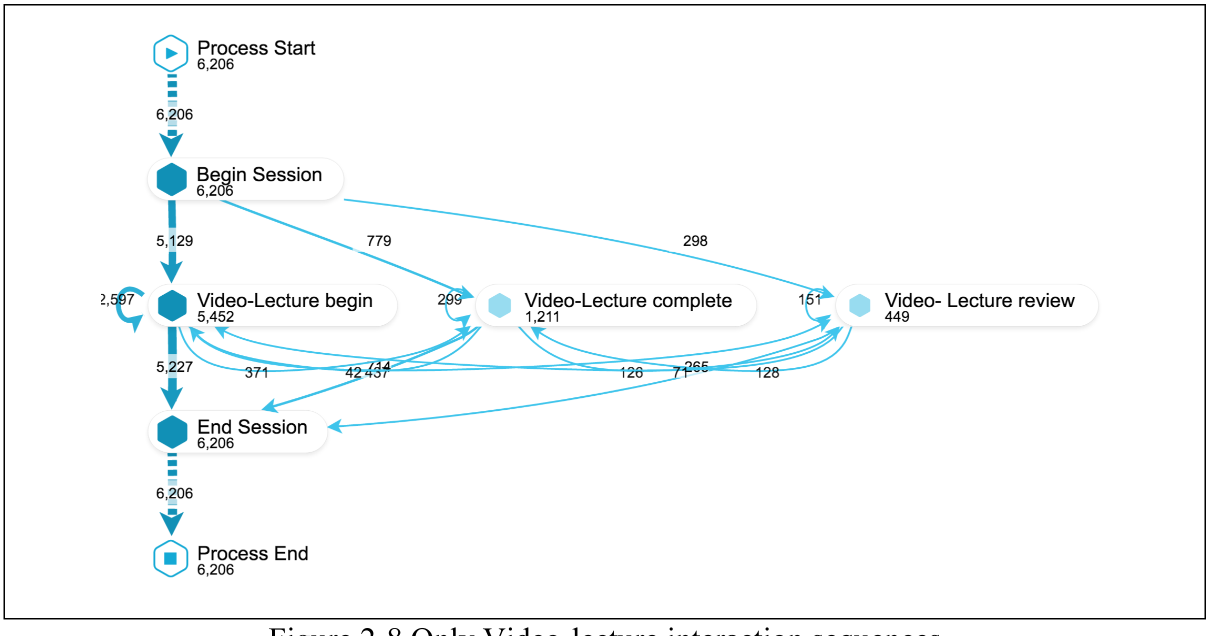


Figure 2-8 Only Video-lecture interaction sequences

Figure 2-8 presents a process model generated using Celonis software, containing interaction sequences by sessions performed with only video-lectures contents (Video-lecture begin, Video-lecture complete. Video-lecture review) being the interaction sequence Begin-session to (→) Video-lecture-begin to (→) End session the most common interaction sequence pattern.

(2) *Only Assessment:* interaction sequence pattern dedicated to working only with assessments in which the most common interaction sequences are *Begin session* to *assessment-try* or *assessment-pass* or *assessment-review* and combinations of them before *End session* (Figure 2-9).

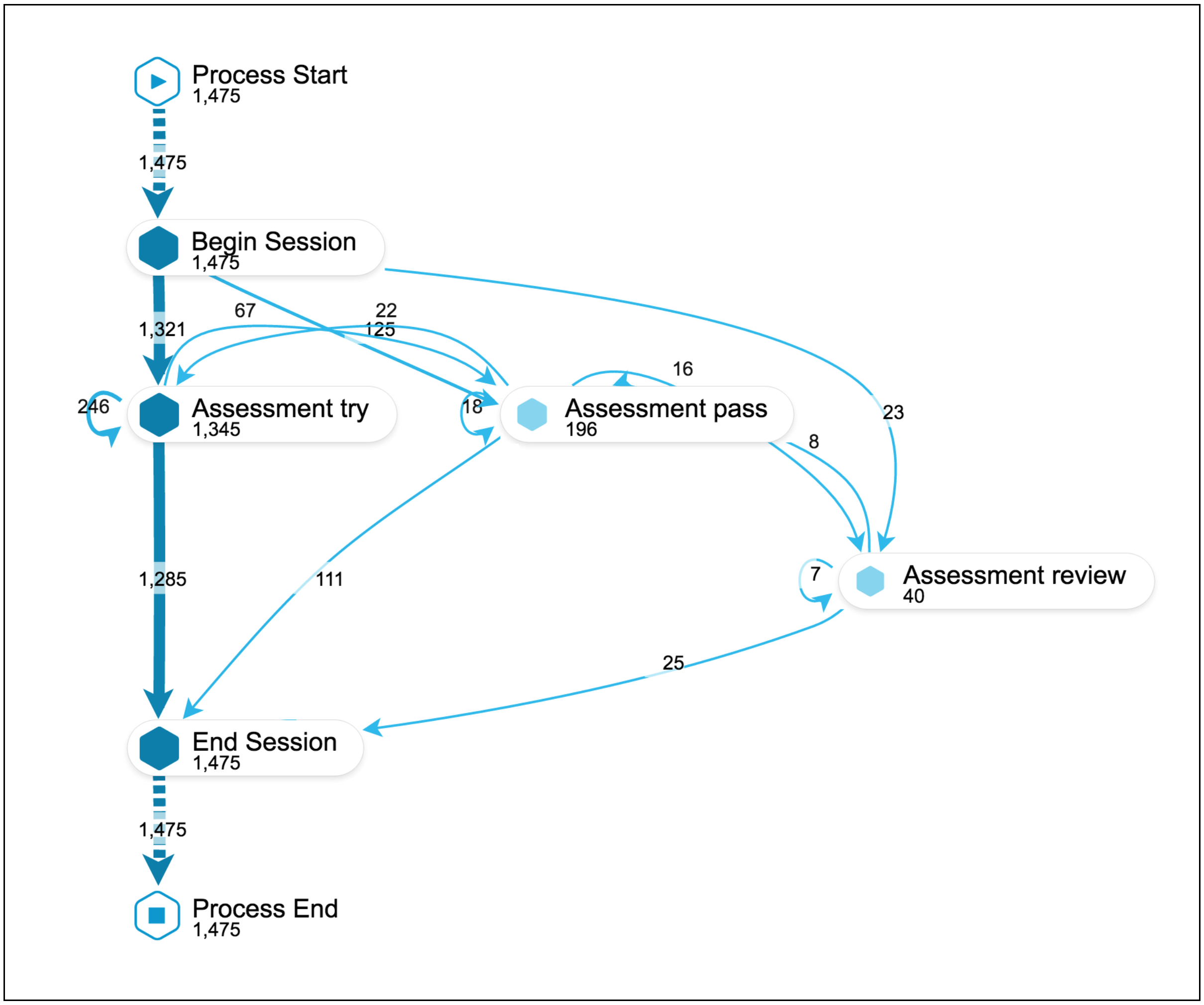


Figure 2-9 Only Assessment interaction sequences

(3) *Assessment-try to Video-lecture*: interaction sequence pattern where the most common interaction sequences observed are (a) *Begin session* to *Assessment-try* (with the intention of trying to solve an assessment) then to *Video-lecture-begin* (looking for information in a new video-lecture) then to *Assessment-try* and *End session*, (b) *Begin session* to *Assessment-try* then to *Video-lecture-complete* (consuming the video-lecture information) then to *Assessment-try* and *End session,* and (c) *Begin session* to *Assessment-try* then to *Video-lecture-review* (looking for specific information) then to *Assessment-try* and *End session* (Figure 2-10).

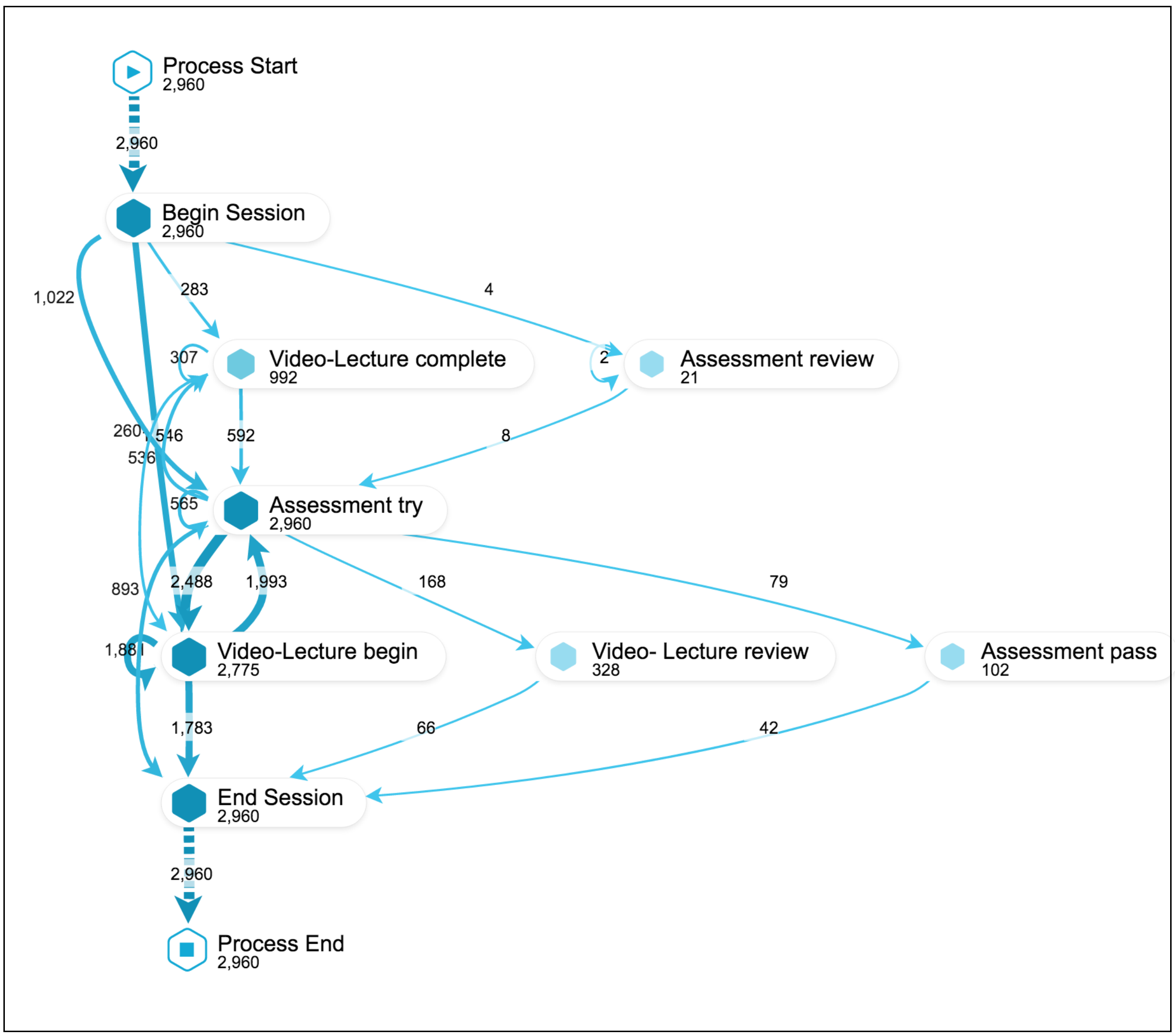


Figure 2-10 Assessment try to video-lecture interaction sequences

(4) *Explore:* interaction sequence pattern composed of an *assessment-try* and a *video-lecture-begin*, where learners only superficially inspect the contents without any intention to complete them (Figure 2-11).

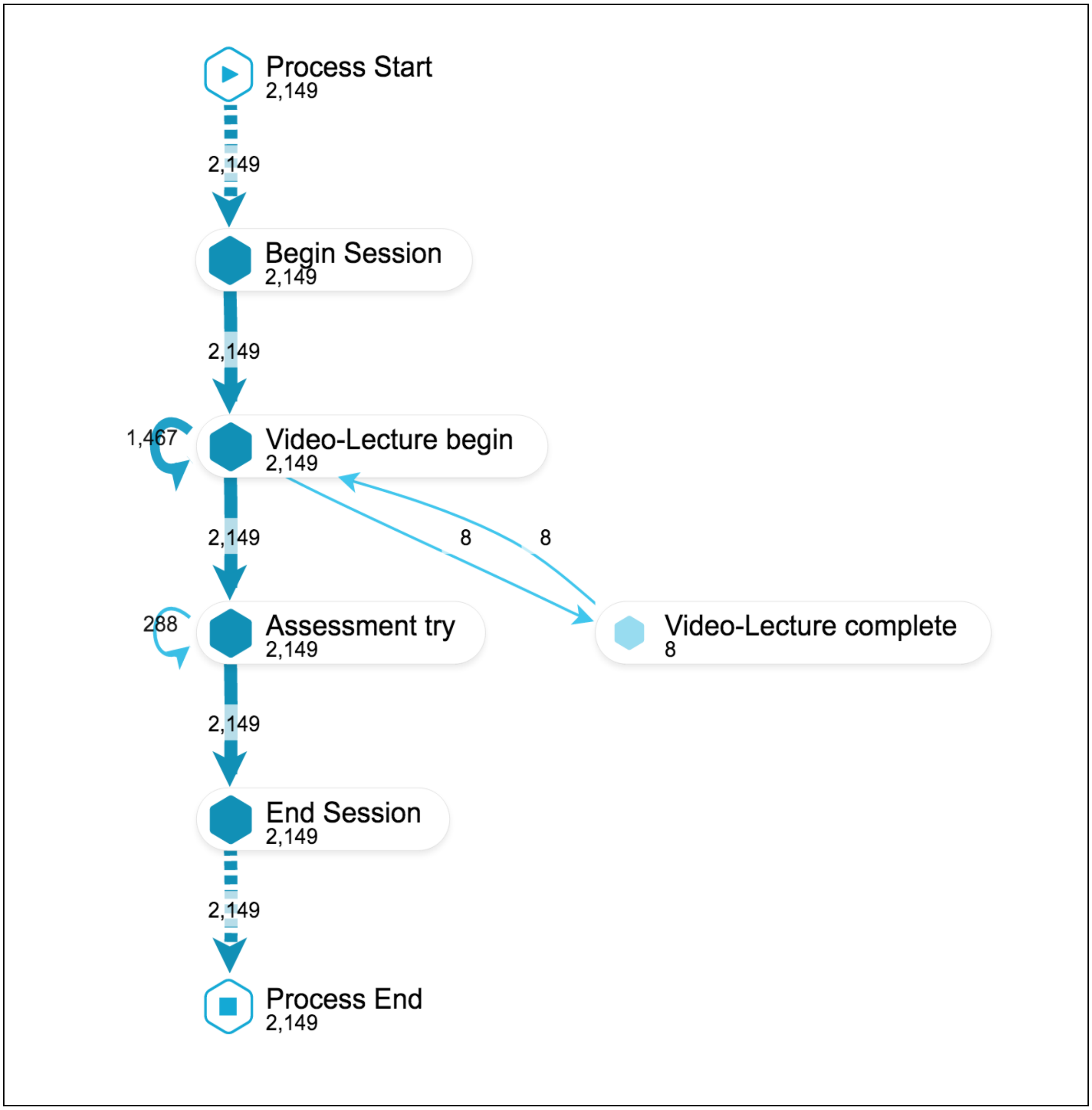


Figure 2-11 Explore interaction sequences

(5) *Video-lecture-complete to Assessment-try*: interaction sequence pattern where the most common interaction sequences observed are (a) *Begin session* to *Video-lecture-complete* then to *Assessment-try* (without achieving it and with no more attempts to complete it) and then *End session* (Figure 2-12).

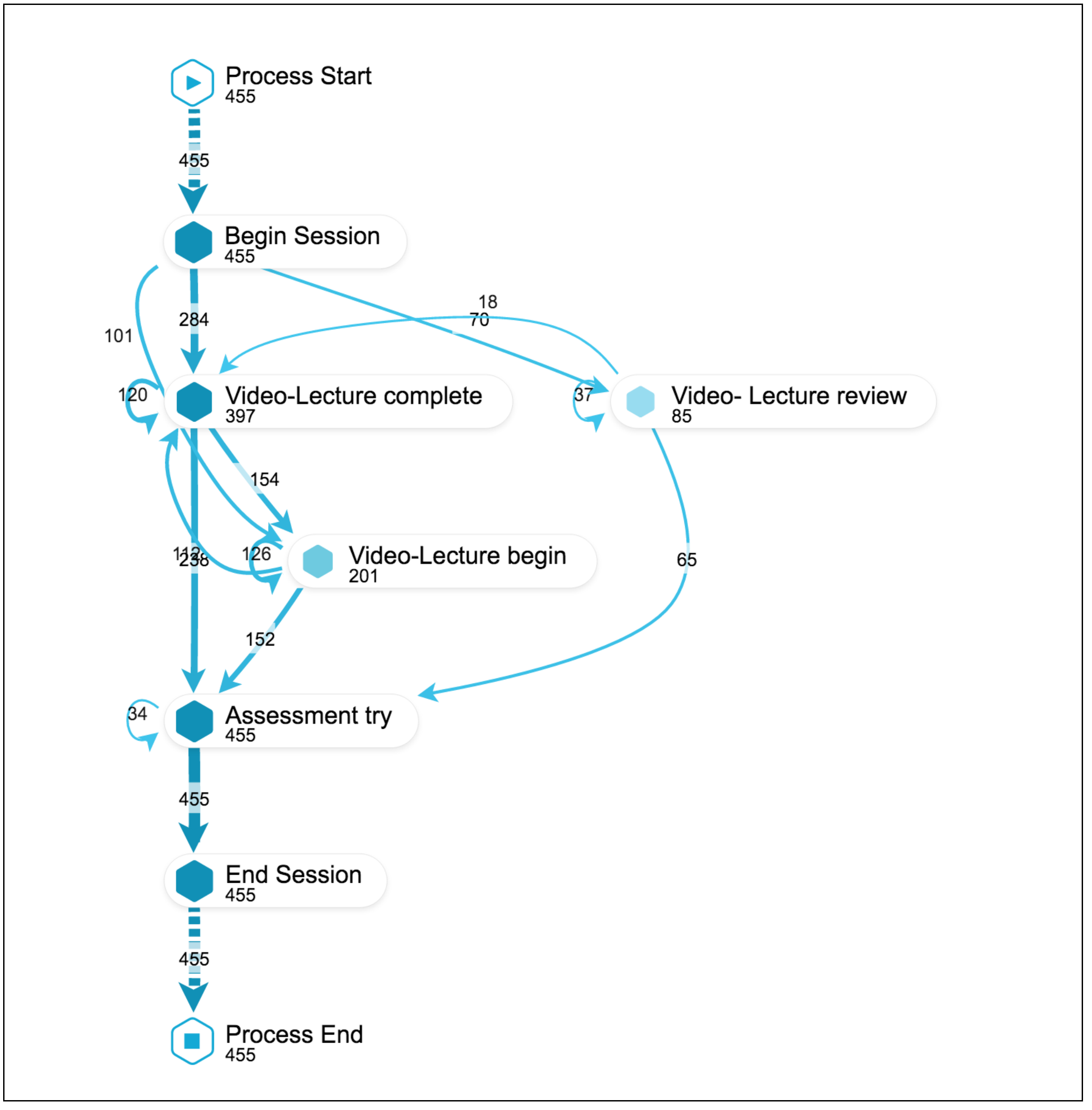


Figure 2-12 Video-lecture complete to assessment try interaction sequences

(6) *Video-lecture to Assessment-pass*: interaction sequence pattern where the most common interaction sequences observed are (a) *Begin session* to *Video-lecture-begin* then to *Assessment-pass* and then *End session,* (b) *Begin session* to *Video-lecture-complete* then to *Assessment-pass* and then *End session,* (c) *Begin session* to *Video-lecture-review* then to *Assessment-pass* and then *End session,* and (d) *Begin session* to *Video-lecture-begin* then to *Assessment-try* then to *Assessment-pass* and then *End session* (Figure 2-13).

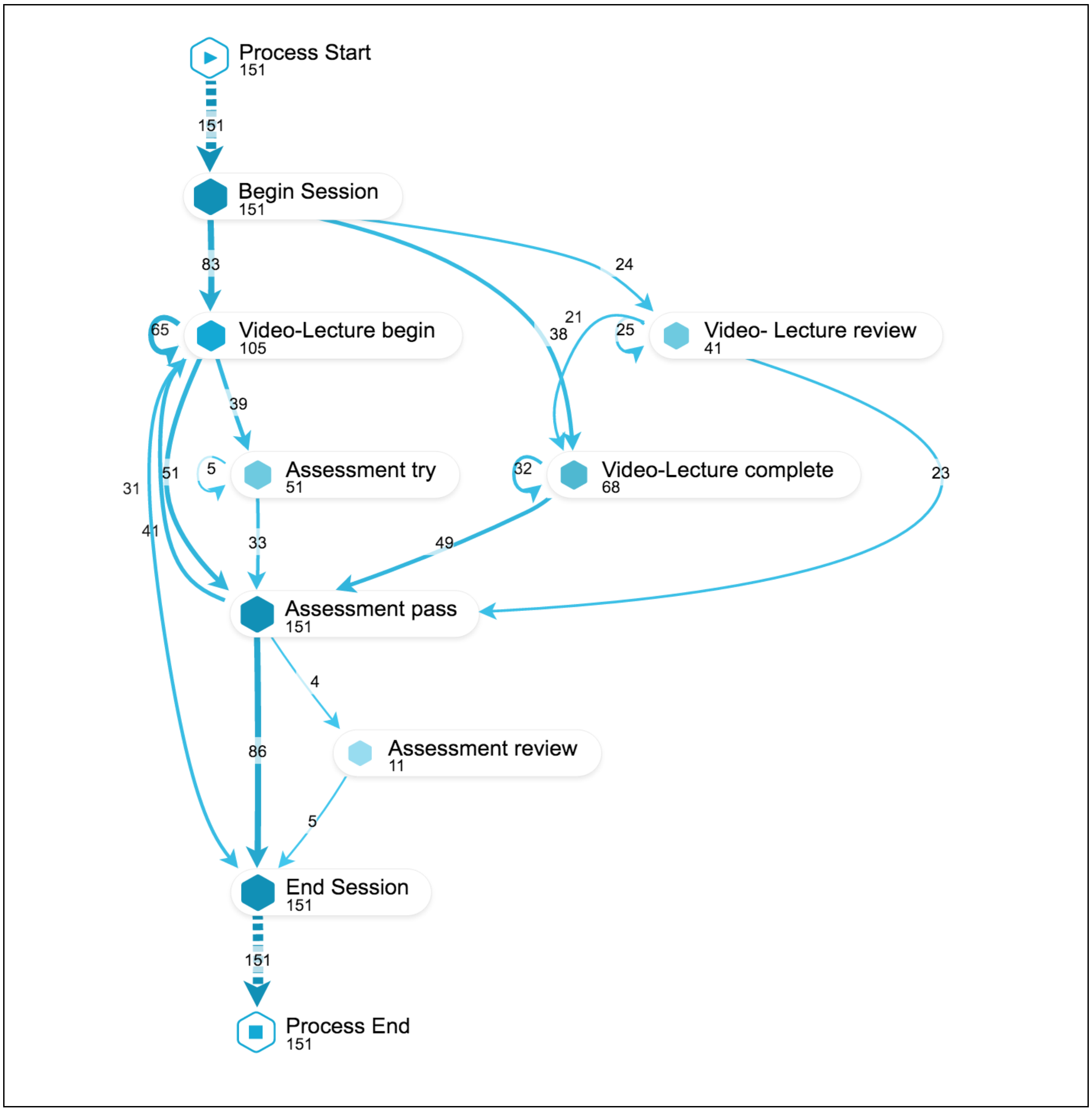


Figure 2-13 Video-lecture to assessment pass interaction sequences

(7) *Others:* interaction sequence patterns that are long and disperse and they do not fit into any interaction sequence pattern mentioned before. The most common interaction sequences observed are (a) *Begin session* to various *Video-lecture-begin*s then to *Assessment-try* and then *End session* (Figure 2-14).

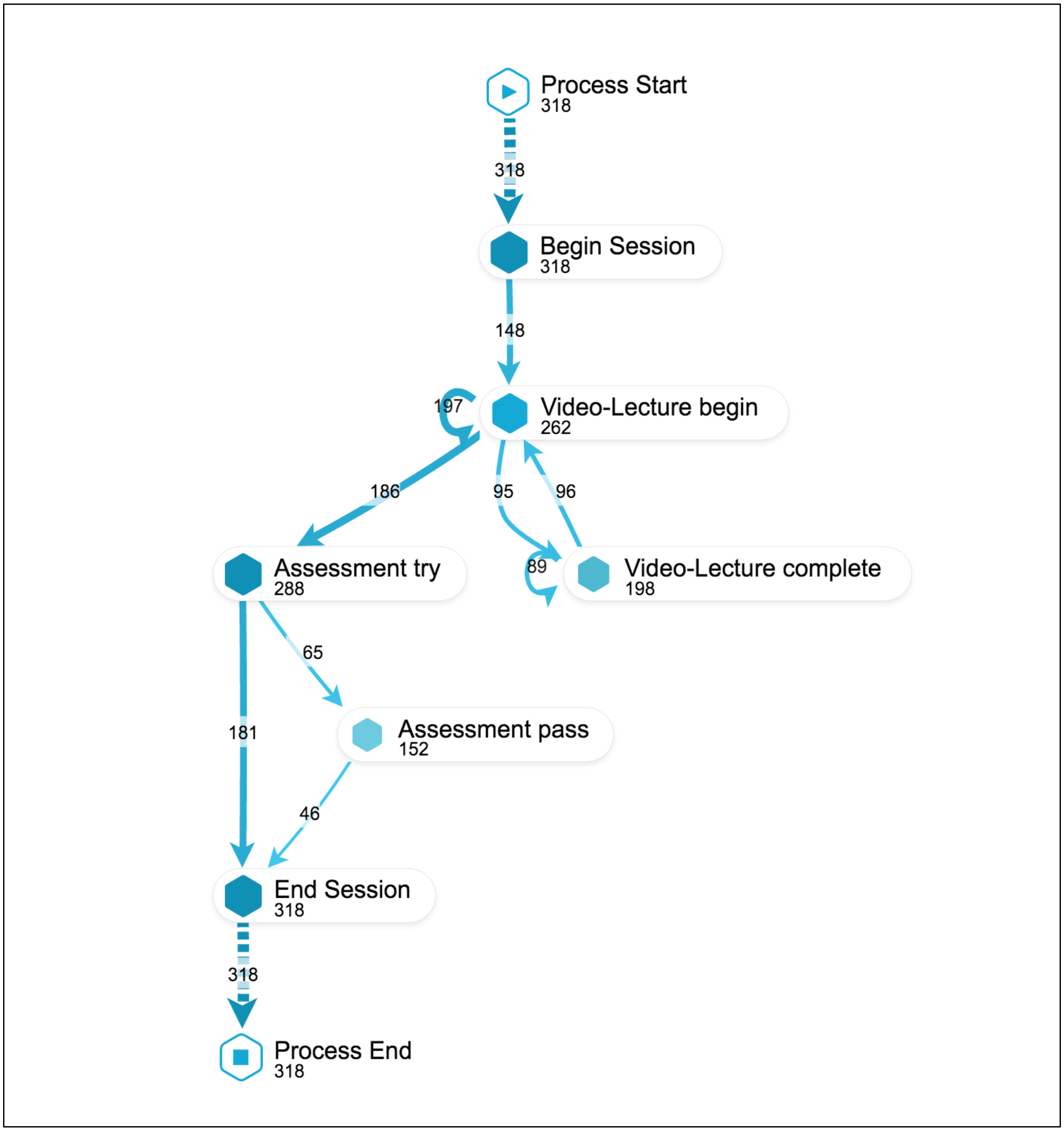


Figure 2-14 Other interaction sequences

The four most common patterns of interaction sequences among MOOC learners (93.26% of the sessions registered) are as follows, in order of frequency (Table 2-14): (1) ***Only Video-lecture*** (45.25% of the sessions follow this type of pattern). The most common interaction sequence in this type of interaction pattern is *Begin session,* then *Video-lecture-begin*, then *End session* without completing the video-lecture; **(2)** ***Assessment try → Video-lecture*****:** 21.58% of the sessions follow this type of pattern, with the most common interaction sequence of this interaction pattern being a loop between *Begin session → Assessment-try → Video-lecture-begin → Assessment-try → Video-lecture-complete → Assessment-try → End session;* **(3)** ***Explore*****:** 15.67% of the sessions follow this type of pattern, in which the most common behavior of the learners is to follow a disorganized interaction sequence in which they go from one type of content (assessments or video-lectures) to another without completing them; **(4)** ***Only Assessment*****:** 10.76% of the sessions follow this type of pattern, in which the most common interaction sequence is *Begin session → Assessment-try → End-session* without completing the assessment; finally, ***Video-lecture complete → Assessment-try*** (3.32%), ***Video-lecture → Assessment-pass*** (1.10%); and ***Others*** (2.32%) interaction sequence patterns are the least common. These patterns help us to understand how learners behave in a session, whether they complete the course or not.

The findigs mentioned above are supported by Table 2-14, that presents the proportions of the interaction sequence patterns based on the number of sessions (N_sessions = 13,714) performed by learners in 3 MOOCs and derived from the MOOC process models. Also, we have analyzed how distinct types of learners (based on academic performance and SRL scores) perform these interaction patterns (excluding *Others*) that provide insights about what strategies they used throughout the course.

Table 2-14 Proportions of the interaction sequence patterns based on the number of sessions (N = 13,714)

Interaction sequence patterns	**ALL 3 MOOCS**		
	N_sessions	**%**	**Learners**
Only Video-lecture	6,206	45.25	2,495
Assessment try → Video-lecture	2,960	21.58	1,271
Explore	2,149	15.67	1,195
Only Assessment	1,475	10.76	865
Video-lecture complete → Assessment try	455	3.32	358
Video-lecture → Assessment pass	151	1.10	132
Others	318	2.32	258
Total	**13,714**	**100%**	**-**

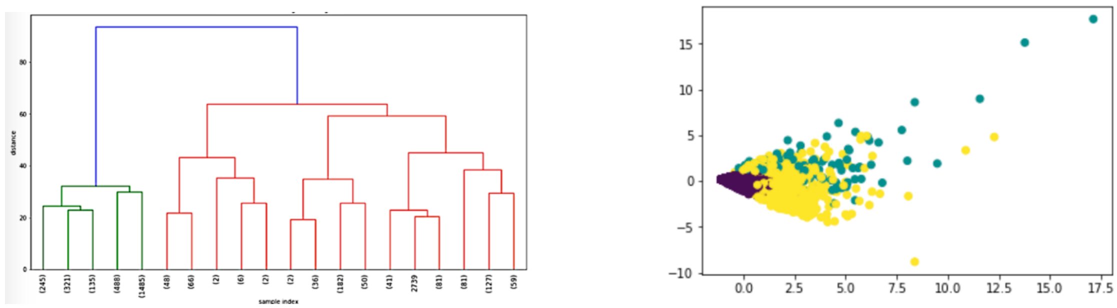
After having identified the most common interaction sequence patterns among MOOC learners in a session, we analyzed how these patterns vary according to whether or not the group of learners complete the course. Specifically, we looked for differences in interaction sequence patterns that completers perform, which should help reveal how their behavior impacts their learning and how it relates with SRL strategies. We analyzed interaction sequence patterns per session. **We found that for completers were more common to perform sessions that contain more assessments than non-completers.** Completers' sessions mainly consist of: (a) taking one assessment after another (called *Only Assessment*) or (b) trying an assessment and then watching a video-lecture (called *Assessment try → Video-lecture*) or (c) watching video-lectures and trying an assessment without completing either (called *Explore*). By contrast, non-completers' sessions consist of watching one video-lecture after another (called *Only Video-lecture*). We found statistical differences between the percentage of sessions of each type performed by these two types of learners (Table 2-15).

Table 2-15 Proportions of the interaction sequence patterns based on the number of sessions (N = 13,714) performed in 3 MOOCs derived from the process models for Completers and Non-Completers

Interaction sequence patterns	**Completers**		**Non-Completers**		χ^2	***p***	***r***
	N_sessions	**%**	***N_sessions***	**%**			
Only Video-lecture	1,253	**36.29**	4,953	**48.27**	149.26	**<0.001*****	0.1043
Assessment try → Video-lecture	922	**26.70**	2,038	**19.86**	71.42	**<0.001*****	0.0722
Explore	610	**17.67**	1,539	**15.00**	13.94	**<0.001*****	0.0319
Only Assessment	417	**12.08**	1,058	**10.31**	8.43	**<0.01*****	0.0248
Video-lecture complete → Assessment try	111	**3.21**	344	**3.35**	0.16	0.690	0.0034
Video-lecture → Assessment pass	44	**1.27**	107	**1.04**	1.26	0.262	0.0096
Others	96	**2.78**	222	**2.16**	4.34	**0.036****	0.0178
Total	**3,453**	**100%**	**10,261**	**100%**	**-**	**-**	

Note. *p < .1; **p < .05; ***p < .001

Then we started grouping learners (N = 3,458) based on the identified interaction sequence patterns in order to detect differences between learners with distinct SRL profiles. We used agglomerative hierarchical clustering based on Ward's method. This clustering technique is advisable for detecting learner groups in online contexts (Kovanović et al., 2015). To select the optimal number of clusters we inspected the resulting dendrogram and check for different ways of cutting the tree structure, in order to obtain a minimal number of interpretable clusters explaining user behavior (Jovanović et al., 2017). Also, we use other clustering techniques as Gaussian mixture and K-means to define the appropriate number of clusters based on the silhouette score. This led to selecting the solution with 3 clusters as the best one (Figure 2-15).



(a) Dendrogram (b) Scatter Plot

Figure 2-15 (a) Dendrogram obtained using agglomerative hierarchical clustering; (b) Scatter Plot with silhouette score = 0.5320

As a result, Table 2-16 describes the resulting clusters in terms of (1) the six identified interaction sequence patterns (we discarded "*Others*" interaction sequence pattern as variable) used for clustering; (2) the SRL score obtained from the self-reported questionnaire; and (3) the course completion.

Table 2-16 Summary statistics for the three learner clusters (sampling, comprehensive and targeting learners): median and standard deviation

Session patterns	**Cluster 1 – Sampling learners**	**Cluster 2 – Comprehensive learners**	**Cluster 3 – Targeting learners**
Only Video-lecture	4.67 (5.41)	22.57 (33.79)	15.72 (13.13)
Assessment try → Video-lecture	3.39 (7.09)	19.85 (18.60)	19.52 (21.42)
Explore	1.84 (3.61)	8.61 (9.40)	10.18 (11.37)
Only Assessment	0.65 (1.62)	4.18 (5.39)	4.39 (6.04)
Video-lecture complete → Assessment try	0.00 (0.00)	1.75 (3.70)	3.84 (4.95)
Video-lecture → Assessment pass	0.00 (0.00)	8.70 (6.05)	0.09 (0.80)
SRL score	3.06 (0.51)	3.12 (0.49)	3.11 (0.52)
Learners	2,674 (77.32%)	124 (3.59%)	660 (19.09%)
Completers	22 (0.8%)	36 (29.03%)	200 (30.30%)
Non-Completers	2,652 (99.2%)	88 (70.97%)	460 (69.70%)

* For completers and non-completers learners the number of them and its percentage are presented under each correspondent cluster.

We have analyzed similarity in the SRL profiles between each group of clusters. As a result, we did not observe statistically significant differences between Cluster 2 and 3, while we observed statistically significant differences when comparing with Cluster 1. Table 2-17 shows the differences between each cluster based on the SRL profile score.

Table 2-17 Differences between each cluster based on the SRL profile score

Cluster #	**Cluster #**	***t***	***p***
2	3	0.1030	0.9179
1	2-3	-2.7333	0.0063***

Note. *p < .1; **p < .05; ***p < .001

The resulting clusters indicate different kinds of learning strategies that learners have adopted while they are facing the MOOC. If we look for specifically particular differences between the different interaction sequence patterns performed by each cluster, we can describe them as follows:

- **Cluster 1 – Sampling learners:** this cluster is composed by learners with least SRL scores compared with their counterparts. Learners in this cluster in average per session perform low number of video-lectures and in average per session perform few attempts to try to solve assessments. These learners have a low activity in the course (generally learners in this group watch just a single video-lecture or start “sample” at the beginning of the course exploring materials with the course already started).

- **Cluster 2 – Comprehensive learners:** this cluster is composed by learners with SRL scores higher than the learners in cluster 1, so they can be considered as more self-regulated (see Table 2-17). Learners in this cluster have developed a variety of learning strategies per session. They watched more video-lectures on average per session than learners in the other clusters. Based on the observed interaction sequences, learners in this cluster tend to follow the path that is provided by the course structure. They also invest more time watching video-lectures and therefore exhibit a higher level of engagement than learners in cluster 3. Thus, learners in cluster 2 focus on performing interaction sequence patterns in a specific order which sets them up for deeply learning the course content.

- **Cluster 3 – Targeting learners:** this cluster is composed of learners with similar SRL scores to those in cluster 2, which suggests that the difference in observed behavior is not due to differences in their SRL profiles. Learners in clusters 2 and 3 also complete the course at similar rates (29% and 30% respectively). However, learners in cluster 3 watch fewer video-lectures and complete more assessments on average per session. They also tend to explore the course contents more than learners in clusters 1 and 2. These differences lead us to describe this group of learners as more strategic or goal oriented. According to Biggs (1999), strategic learners tend to focus their efforts on assessments to achieve performance-oriented objectives and exhibit less engagement overall. This interpretation is consistent with the observation that the level of engagement in cluster 3 is lower than in cluster 2.

Table 2-18 presents the differences found between clusters 2 and 3 in relation to interaction sequence patterns. We found statistically significant differences with significance level of .05 for the Only Video-lecture, Video-lecture complete → Assessment try and Video-lecture → Assessment pass patterns; and statistically significant differences with significance level of .1 for Explore pattern, with effect sizes (r) ranging from small (Only Video-lecture, Explore); medium (Video-lecture complete → Assessment try) and big (Video-lecture → Assessment pass).

Table 2-18 Comparisons respect interaction sequence patterns performed between Comprehensive and Targeting learners

	Comprehensive learners	**Targeting learners**	***t***	***p***	***r***
Only Video-lecture	22.57	15.72	2.2276	0.0276**	0.1917
Assessment try → Video-lecture	19.85	19.52	0.1788	0.8583	0.0129
Explore	8.61	10.18	1.6393	0.100*	0.1159
Only Assessment	4.18	4.39	0.3880	0.6984	0.0284
Video-lecture complete → Assessment try	1.75	3.84	5.4396	**<0.001***	0.3476
Video-lecture → Assessment pass	8.70	0.09	15.8244	**<0.001***	0.6859

Note. $*p < .1$; $**p < .05$; $***p < .001$

2.3.4 Discussion

We identified the following interaction sequence patterns as the most frequently repeated by learners in a MOOC: (1) watching one video-lecture after another; (2) taking one assessment after another; (3) trying an assessment and then watching a video-lecture; (4) watching a video-lecture and then passing an assessment; (5) completing a video-lecture and then trying an assessment; and (6) watching video-lectures and trying an assessment without completing either. The extracted patterns can be interpreted as manifestations of specific learning strategies (Winne, 2013) and thus it is possible to link behavioral patterns to learning strategies. However, these patterns are only a first step towards understanding how learners self-regulate in a MOOC. More research is needed to refine and extend the identified patterns, for instance by incorporating more information such as the amount of time spent in each interaction sequence. This type of information would shed more light on how much effort learners invest in applying a particular strategy.

We found that learners who completed the course exhibited different interaction patterns than those who did not complete it. Unsurprisingly, completers were more engaged with assessments than non-completers. Going deeper, we were able to identify three types of learners in terms of their behavioral and SRL characteristics: (1) Comprehensive learners, who have a high SRL profile, tend to follow the sequential structure of the course materials in the MOOC (i.e., guided by instructional design), and engage in more organized sessions that allow them to gain a deeper understanding of the content; (2) Targeting learners, who also have a high SRL profile but who strategically seek out specific information to pass the course assessments; and (3) Sampling learners, who have a low SRL profile, tend to behave in irregular ways, and are the least likely to complete the course. This clustering is consistent with findings in prior research. Kizilcec, Piech, and Schneider (2013) originally identified four clusters of prototypical MOOC learners: Completing, Disengaging, Auditing, and Sampling learners. In comparison, Sampling learners explore parts of the course, while Comprehensive and Targeting learners appear to be two types of Completing learners who may pursue different goals: deep learning and certification, respectively. Beheshitha et al. (2015) examined learners' cognitive SRL strategies while using the nStudy tool and found differences between Deep and Surface learners that partly map onto the current distinction between Comprehensive and Targeting learnings. Relatedly, Kovanovic et al. (2015) identified three profiles and interpreted them in terms of deep versus surface approaches to learning and performance versus mastery achievement goal orientations.

We attempt to reconcile the identified behavioral patterns with SRL strategies that are established in the literature. Table 2-19 summarizes the relationship between these observed patterns and SRL theory. We were able to associate each interaction sequence pattern to one or more theory-based SRL strategies.

First, the *Only Video-lecture* interaction pattern was associated with three SRL strategies in the literature: **studying** (Garavalia & Gredler, 2002), **rehearsing** (Broadbent, 2017), and **repeating** (Sonnenberg & Bannert, 2015). All three are cognitive SRL strategies in which learners invest time to better understand a particular idea or knowledge component in the course. Interpretation of this interaction pattern could be enriched with additional information from external resources (e.g., capturing trace data outside the platform). This would provide more insight into whether learners use organizational SRL strategies, such as note taking, creating concept maps, or using other means to make sense of the content. As Veletsianos et al., (2016) state, "*automatically collected data by learning platforms does not necessarily offer a comprehensive and complete representation of learners' behavior*."

Second, the *Only Assessment* interaction pattern was associated with two cognitive SRL strategies: **elaboration** (Weinstein, Acee, & Jung, 2011) and **evaluation** (Sonnenberg & Bannert, 2015). This interaction pattern was most frequently observed among the strategic Targeting learners who are likely to complete the course (cf. Tables 2-15 and Table 2-18). Information about this interaction pattern could be complemented with additional information about the action's learners perform to connect the new information to their prior knowledge, and to gain more insight into whether they process information in a deep or superficial way.

Third, the *Assessment try→Video-lecture* interaction pattern, which was most common among completers (cf. Table 2-15), was associated with **help-seeking** (Karabenick et al., 2007; Lodge & Corrin, 2017). Help seeking in online environments can mean that a learner looks for human help through forums, chats, or other online communication mechanisms (Broadbent & Poon, 2015). However, help can also be sought from course-internal resources (e.g., video-lectures, forums, assessments) or external resources (digital or physical material outside the platform). Thus, to better understand applications of this strategy, there is a need to collect qualitative data from interviews or focus groups asking learners about their help-seeking behavior in MOOCs.

Fourth, the *Video-lecture→Assessment pass* interaction pattern, which was most common among Comprehensive learners (cf. Table 2-18), was associated with the **reviewing record strategy** (Zimmerman & Pons, 1986). This interaction pattern may reflect MOOC teachers' and instructional designers' intentions for how learners should proceed in the course: first watch a video-lecture and then pass an assessment.

Fifth, the *Video-lecture-complete→Assessment* interaction pattern, which was most common among Targeting learners (cf. Table 2-18), was associated with **self-evaluation** (Zimmerman & Pons, 1986). This is a metacognitive SRL strategy that has learners tracking themselves and checking their progress in the course. With the appropriate feedback, it would be possible to develop a mechanism of self-monitoring that could help learners regulate how they approach the learning process.

Finally, the *Explore* interaction pattern was associated with **task exploration** (Van Der Linden, Sonnentag, Frese, & Van Dyck, 2001). This pattern was mainly performed by Targeting learners (cf. Table 2-18) and it appeared to be a strategic behavior, for instance, switching between video-lectures and assessments without completing them to investigate how the topics and the materials are organized.

Based on this preliminary pattern-strategy mapping, we found that Comprehensive learners tended to use rehearsal, repeating, studying, reviewing record, and self-evaluation SRL strategies. Moreover, these learners tended to go back and forth over the course content to review video-lectures before and after completing an assessment, a behavior that could be a form of *cognitive retrieval practice* (Davis et al., 2016; Johnson & Mayer, 2009; Roediger III & Butler, 2011). Conversely, Targeting learners tended to use evaluation, elaboration, and task-exploration SRL strategies. These learners acted strategically, since they sought out specific information that would help them pass course assessments. Both Comprehensive and Targeting learners tended to use a form of help-seeking SRL strategy.

Table 2-19 Connecting Theory-based SRL strategies to patterns from observed learning behavior

Interaction Pattern	Description	SRL Strategy
Only Video-lecture	Interaction pattern dedicated to working only with video-lectures (2 or more consecutively). The interaction sequence patterns consist of: *Begin session* to *video-lecture-begin* or *video-lecture-complete* or *video-lecture-review* and combinations of them before *End session.*	The interaction sequences referring to *video-lecture begin* and *video-lecture complete* could be related to the ***Study*** SRL strategy described by Garavalia and Gredler (2002) (e.g., "Study in a particular order"). *Video-lecture review* in isolation is related to the ***Rehearsal*** SRL strategy described by Broadbent (2017) (e.g., "Learner who listens to an online lecture repeatedly") or by Weinstein et al. (2011) (e.g., "Go over information"). This pattern could also be related to ***Repeating,*** an SRL strategy defined by Sonnenberg and Bannert (2015) as "Watching (part of) a lecture that was completed in the past."
Only Assessment	Interaction pattern dedicated to working only with assessments (2 or more consecutively). The interaction sequences patterns consist of: *Begin session* to *assessment-try* or *assessment-pass* or *assessment-review* and combinations of them before *End session*.	The interaction sequences referring to *assessment-try* and *assessment-pass* could be related with the ***Elaboration*** SRL strategy described by Weinstein et al. (2011) (e.g., "Answering possible test questions"). When assessment review occurs, it could also be associated with the ***Evaluation*** SRL strategy described by Sonnenberg and Bannert (2015) (e.g., "Look up an assessment that was completed in the past").
Assessment try →Video-lecture	Interaction pattern where the learner tries an assessment and then performs a video-lecture interaction. The interaction sequence patterns consist of: (a) *Begin session* to *Assessment-try* (with the intention of trying to solve an assessment) then to *Video-lecture-begin* (looking for information in a new video-lecture) then to *Assessment-try* and *End session*. (b) *Begin session* to *Assessment-try* then to *Video-lecture-complete* (consuming the video-lecture information) then to *Assessment-try* and *End session*. (c) *Begin session* to *Assessment-try* then to *Video-lecture-review* (looking for specific information) then to *Assessment-try* and *End session.*	These interaction sequences (a), (b) and (c) could be associated with the ***Help-seeking*** SRL strategy (Karabenick & Dembo, 2011; Corrin, de Barba, & Bakharia, 2017). This help-seeking could be classified as internal if the learner looks for information inside the MOOC environment, or as external if they look for information outside the MOOC platform, using resources such as web pages, digital books, learning objects, etc.
Video-lecture→Assessment pass	Interaction pattern where the learner passes an assessment after performing many video-lecture interactions. The interaction sequence patterns consist of: (a) *Begin session* to *Video-lecture-begin* then to *Assessment-pass* and then *End session*. (b) *Begin session* to *Video-lecture-complete* then to *Assessment-pass* and then *End session*. (c) *Begin session* to *Video-lecture-review* then to *Assessment-pass* and then *End session*. (d) *Begin session* to *Video-lecture-begin* then to *Assessment-try* then to *Assessment-pass* and then *End session.*	The interaction sequences performed in (b) correspond to those proposed in the MOOC instructional design in the MOOC platform (*Video-lecture-complete→Assessment pass*). Interaction sequences (a), (b), (c) and (d) could be associated with the ***Reviewing record*** SRL strategy described by Zimmerman and Pons (1986) (e.g., "Learner initiated efforts to try, complete or review test, notes, or textbooks to prepare for a test").
Video-lecture-complete→Assessment try	Interaction pattern where the learner attempts to solve an assessment after completing a video-lecture. This interaction sequence pattern consists of: *Begin session* to *Video-lecture-complete* then to *Assessment-try* (without achieving it and with no more intentions made to complete it) and then *End session.*	This interaction pattern could be associated with the ***Self-evaluation*** SRL strategy described by Zimmerman and Pons (1986) (e.g., "Student initiated evaluations of the progress of their work").
Explore	Interaction pattern performed by lurker learners, who only superficially inspect the video-lectures and assessments (*video-lecture-begin* and *assessment-try*) without any intention to complete them.	This interaction pattern could be associated with the ***Task exploration*** SRL strategy described by Van Der Linden et *al*. (2010) (e.g., "The task exploration strategies performed in order to obtain more information and plan for learning a new computer program").

2.4 An adaptation of a Process Mining methodological approach for extracting SRL strategies in edX MOOCs

In the past years, and due to the massive amount of data collected from MOOC platforms, several researchers in the Learning Analytics (LA) community have focused on the analysis of learners' trace data to unveil their learning strategies and propose new classifications accordingly (Fincham, Gasevic, Jovanovic, & Pardo, 2018; Jovanović et al., 2017; Maldonado-Mahauad et al., 2018). Several methods and techniques have been applied to analyze these trace data, such as unsupervised machine learning techniques, sequence mining algorithms, transition graphs or hidden Markov models (Fincham et al., 2018; Jovanović et al., 2017). All these methods are event-based approaches; where an event is defined as an action of the learner with the course content, tools or learning platform functionalities. However, recently, researchers from the Process Mining (PM) field, who are experts in the analysis of data processes, proposed novel methods to unveil learning strategies from big data looking for other representations to understand how self-regulated learning processes occurs (Maldonado-Mahauad et al., 2018; Van den Beemt, Buijs, & Van der Aalst, 2018).

Process Mining techniques can be used to discover models that describe and represent sequences of interactions between learners and course materials (Van den Beemt et al., 2018). In these recent studies, PM techniques have shown to be very robust to understand users' interactive workflows within a particular system in both structured and unstructured processes. Moreover, compared with other techniques such as sequence mining, transition graphs or hidden Markov models, whose outputs are difficult to relate with natural learning processes and to draw meaningful insights about them. In this sense, PM provides encouraging results for understanding learning processes (van Eck, Lu, Leemans, & van der Aalst, 2015) and is a suitable approach for studying learning strategies, as a dynamic regulatory activity carried out during a learning task (Sonnenberg & Bannert, 2015), facilitating the discovery of end-to-end learning process models using the recorded events.
But, despite the encouraging results obtained using PM techniques, results from one study do not necessarily apply to other contexts. So, there has been an increasing interest in LA

research in replicating studies across contexts (Ferguson et al., 2015; Gardner, Brooks, Andres, & Baker, 2018; Kizilcec & Brooks, 2016), although studies of this nature are still scarce in part due to the variation of the instructional conditions (Gašević, Dawson, Rogers, & Gasevic, 2016). Therefore, new analyses with different data should be done to understand the validity of PM methods in other learning environments and contribute providing more evidence about the impact of the learning context on learners' behavior and study strategies.

To continue this trend of reproducible science, this subsection builds upon the analytical methodology proposed in subsection 2.3 for unveiling students' learning strategies in self-paced MOOCs in Coursera (subsection 2.3). In that section, seven different learning strategies were identified, and learners were classified into three groups: sampling, comprehensive, and targeting learners. In this subsection, we adapt this particular PM methodology and analyze its application in a MOOC deployed over the edX platform, delivered in a synchronous mode, where the digital resources were developed in English language and consisted in video-lectures, graded and non-graded assessments and other resources. The aim of this adaptation effort is two-fold: 1) to understand whether we could replicate (partially or totally) the analysis conducted in subsection 2.3 and what methodological decisions we had to change for this purpose and; 2), to extend the current knowledge about students' learning strategies in MOOCs and the influence of the learning context.

2.4.1 Related Work

2.4.1.1 Analysis of learning strategies in MOOCs: methods and techniques across contexts

Observing learning strategies in MOOCs, even when these manifest as a set of events or actions, involves several challenges, such as: 1) how to transform traces of fine/coarse-grained data into interpretable behavior (learning strategies); 2) how to identify and observe behavioral changes; and 3) how to understand whether an observable behavior relates to a particular learning strategy or to more than one (Pashler & Wagenmakers, 2012). Recent advances in the evolving disciplines of LA and PM have contributed to overcome these challenges. LA focuses on the human interpretation of data and could provide insights into learning strategies (Boekaerts, 1997), while PM focuses on the application of computational techniques on event-based learning activities to discover sequence of learning behavior (Van den Beemt et al., 2018). Examples of these advances are the work done by Mukala, Buijs, & Van Der Aalst (2015), who applied PM techniques in a MOOC in Coursera with 43,218 learners to understand their learning processes analyzing how they performed watching video-lectures and taking assessments. In Maldonado-Mahauad et al. (2018) the authors used the fuzzy miner algorithm to extract seven types of learning strategies from learners enrolled in four MOOCs in Coursera. Authors in Juhanák, Zounek, & Rohliková (2017) used PM to explore learners' quiz-taking behavior and interaction patterns in a learning management system. Finally, authors in Van den Beemt et al. (2018) also used PM and clustering techniques to describe the learning behavior of 4 groups of learners. These prior works set the basis to start considering PM as a suitable technique for analyzing sequences of learning behavior. However, more examples and replication studies are needed, since both the methodological decisions involved in the use of PM and the context in which the data is gathered may strongly condition the final results.

One of the most important concerns in today's scientific community is that of reproducibility. A key domain in which reproducibility has been identified as a particularly important problem is that of Psychology (Pashler & Wagenmakers, 2012). Psychology researchers have observed a systematic trend wherein results from studies carried out in one (original) context do not reliably transfer or generalise to other contexts (Pashler & Wagenmakers, 2012; Stanley & Spence, 2014). Examples of contextual factors and changes include everything from demographic variables of participants to the physical or virtual environment in which the study is carried out. This trend has highlighted that fact that results from scientific experiments should always be: 1) sufficiently contextualised and reported on accordingly and; 2) replicated across different contexts.

Research in education has found that, just as is the case in Psychology research, the outcomes regarding the impact on learning are also highly dependent on context. A number of studies have found that learning outcomes and learner engagement are highly dependent on the context in which the learning occurs (Meyer & Muller, 1990; Trigwell & Prosser, 1991). This issue has recently begun to be explored in the LA literature by examining the effect of a course structure/design on passing rates (Davis, Seaton, Hauff, & Houben, 2018). By leveraging the literature on learning design (the science of structuring and sequencing instructional activities) Laurillard (2013) found that certain course designs (context) lead to significantly different passing rates than others (Davis, Seaton, et al., 2018). Ferguson and colleagues (2015) also demonstrated in a replication study that classifications of learners according to their behavior varies from a MOOC deployed in Coursera or in FutureLearn, a platform created for promoting a socio-constructivist learning approach (Ferguson et al., 2015).

2.4.1.2 Research questions

Two sub research questions drive this subsection with the aim of understanding how the methodology for detecting learning strategies proposed in subsection 2.3 adapts to other learning contexts:

Sub-RQ 1.3: To what extend can we replicate (partially or totally) the methodology applied in the previous subsection 2.3 to extract students' learning strategies in a MOOC? The objective of this research question is to analyze and discuss what the methodological decisions are needed for applying the same methodology in a different context and see the implications on the final analysis.

Sub-RQ 1.4: How do students' learning strategies in this new context differ from those from the previous subsection 2.3? Learning is highly dependent on context, and the structure and characteristics of a course can have a direct effect on learners' behavior. In order to understand whether the learning strategies found in subsection 2.3 vary in this new context, we will analyze two aspects: a) the learners' behavioral patterns in a synchronous MOOC in edX; and b) how learners can be classified according to their behavior and learning outcomes.

2.4.2 Adapting the methodological approach for extracting SRL strategies in edX MOOC

Some decisions were taken during the process to adapt the methodology developed in subsection 2.3 to the new learning context. We bold the text indicating **[Decision-X]**, where "X" corresponds to the number of the methodological decision taken and described the decision in italics.

2.4.2.1 Context: MOOC and Sample

We used data from one MOOC on Programming in Java offered by Universidad Carlos III of Madrid in edX. The course was taught in English and the materials were organized into five modules. This MOOC included video-lectures and numerous interactive activities as formative and summative assessments. Figure 2-16 presents the course structure. This MOOC followed a synchronous approach and the contents were released weekly. The course was open from April 28th, 2015 until June 30th of the same year. The estimated learners' workload was between 5 to 7 hours per week. To pass the course the learners needed to obtain 60% of the final grade. Summative assessments (exams) had a weight of 75% of the final grade. The rest, 25% of the grade, was assigned to programming activities that consisted of two peer assessments. The final study sample comprised N = 50,776 online learners that at least completed one video-lecture in the MOOC. *The sample selection differs from the subsection 2.3, where the subjects were selected based on if they had answered or not a self-reported SRL survey* **[Decision-1].**

Week 0	
General Topic	VL VL VL VL VL
Week 1	
Topic 1	VL AF AF AF VL AF AF AF VL AF AF
Topic 2	VL AF AF AF AF VL VL AF AF AF AF AF AF
Topic 3	VL AF AF VL AF AF AF AF
Topic 4	VL VL AF VL AF AF
LAB 1	VL AF AF AF AF AF AF AF AF AF AF
RECAP	VL VL VL VL
EXAM 1	AS
SELF-EVALUATION	AF AF AF AF AF AF AF
SELF-EVALUATION	VL
Week 2	
Topic 1	VL AF AF VL AF AF AF AF AF
Topic 2	VL AF AF VL AF
Topic 3	VL VL AF AF AF
Topic 4	VL AF VL AF VL AF VL AF
LAB 2	VL AF AF AF AF AF AF AF AF AF
RECAP	VL VL VL
EXAM 2	AS
SELF-EVALUATION	AF AF AF AF AF AF AF AF AF
SELF-EVALUATION	VL
Week 3	
Topic 1	VL AF VL AF VL AF AF AF
Topic 2	VL VL AF AF AF AF
Topic 3	VL AF VL AF VL AF AF VL AF AF
Topic 4	VL VL AF VL AF VL AF AF
LAB 3	VL AF AF AF
RECAP	VL VL VL VL VL VL VL
EXAM 3	AS
PEER ASSESSMENT 1	AS
SELF-ASSESSMENT 1	AF
SELF-EVALUATION	AF AF AF AF AF AF AF AF
SELF-EVALUATION	VL
Week 4	
Topic 1	VL AF VL AF AF VL AF
Topic 2	VL AF VL AF AF AF VL VL
Topic 3	VL AF AF VL AF VL AF
Topic 4	VL AF AF AF VL AF AF VL AF VL AF VL VL
LAB 4	VL AF AF AF AF AF AF AF AF
RECAP	VL VL
EXAM 4	AS
SELF-EVALUATION	AF AF AF AF AF AF AF AF AF
SELF-EVALUATION	VL
Week 5	
Topic 1	VL AF AF VL AF VL AF AF
Topic 2	VL AF VL VL AF VL
Topic 3	VL AF VL AF VL AF AF VL AF VL AF
Topic 4	VL VL VL
LAB 5	VL AF AF AF AF AF
RECAP	VL VL
EXAM 5	AS
PEER ASSESSMENT 2	AS
SELF-EVALUATION	VL

Figure 2-16 Structure of the course presenting the contents of each week. VL=video-lecture, AF=formative-assessment, AS=summative-assessment

2.4.2.2 Procedure

To extract students' learning strategies, we followed the stages proposed in subsection 2.3. Specifically, we applied the PM^2 methodology (van Eck et al., 2015), and defined 4 phases to obtain the process model from learners' behavior in interaction with the course content: 1) extraction stage, 2) event log generation, 3) model discovery and 4) model analysis.

Extraction stage. The data used in this subsection were related to learners' commitment with the MOOC contents. These contents were presented in the course as a sequence of different digital resources such as video-lectures, and formative/summative activities. In subsection 2.3 we only considered interactions with video-lectures and summative activities. *In this subsection, we extended the data employed to characterize the learners' interaction with the course content by considering the following resources: LTI activities* *(integrating an external development environment called Codeboard), graded activities, navigation between modules, tabs and clicks on the home page in edX* **[Decision-2]**. Each time a learner interacted with a digital resource in edX, a log with a learning event was generated and stored. This raw data was organized in different files classified in general data, forums, and personal data containing information about learners' behavior.

Event log generation stage. For creating the event log in this stage, we built upon the two conceptual assumptions defined in subsection 2.3: 1) to adopt the same definition of study session as a period of time in which the MOOC platform registered continuous activity of a learner within the course, with intervals of inactivity no greater than 45 minutes and; 2) to adopt the same definition of an interaction as an event triggered by a learner when this interacts with resources from the MOOC. In comparison with subsection 2.3 where only six possible interactions were defined*, we defined ten types of possible interactions (Table 2-20) depending on the MOOC structure and the digital resource the learner interacted with* **[Decision-3]**. This extension on the number of interactions was a necessary step in order to consider the content provided in the course. Table 2-20 presents the ten types of interactions defined, which are related to video-lectures, assessments, home view page, and navigation between modules and tabs.

As a result, we defined an event log that contained: 1) the user identification, 2) a time stamp, 3) the interaction performed, and 4) the number of the session in which the event was triggered when learners engaged with MOOC contents. Table 2-21 presents part of the event log used as an example. We also defined success in a synchronous MOOC based on the grades that learners achieved during the course (at least 60% of the grade in the course), as also we did in subsection 2.3. On the contrary, *we did not include the SRL profile as part of the event log* **[Decision-4]**.

Table 2-20 Types of interactions defined based on course resources

Course resource	Interaction	Description
Video-lecture	Begin	Begin but not complete watching a video-lecture that was not previously completed.
	Complete	Complete watching more than the 75% of the video-lecture for the first time.
	Review	Watch (part of) a video-lecture that was completely watched in the past.
LTI activity	Assessment Formative	Attempt to solve a non-graded activity at the first time.
	Assessment Formative Review	Go back to a non- graded assessment that was previously visited.
Graded activity	Assessment Summative Try	Attempt to solve a graded activity without achieve it.
	Assessment Summative Complete	Successful attempt to solve a graded assessment for the first time.
	Assessment Summative Review	Go back to a graded assessment that was previously completed successfully.
Home Page	Home View	Go to the home page of the course
Modules, Tabs	Navigation	Go through modules (vertically) or tabs (horizontally) looking for specific content.

Table 2-21 Example of the minimal columns of the event log generated

UserId	Time stamp	Interaction	# Session
28	1434522567	Assessment-Formative	1
28	1434522567	Video-Lecture-Complete	1
161	1430520885	Assessment-Formative	1
161	1430520885	Navigation	1
161	1430520885	Navigation	1

Model discovery stage. Given the exploratory context of this study in which it was necessary to handle complex processes, we selected the same Disco algorithm and their implementation in the Disco commercial tool (Günther & Rozinat, 2012) as we did subsection 2.3. The resulting process model was confirmed using the implementation of the Celonis algorithm. Both implementations use a variation in the fuzzy miner algorithm that produced interesting synopses of the learning process in comparison with other techniques (Saint, Gašević, & Pardo, 2018).

Model Analysis stage. As a result of the previous stages, we generated a process model that contained learners' behavior (see *Figure 2-17*). Then, we analyzed the observed behavior in order to unveil learning strategies. For this stage, we identified the most frequent interaction sequences performed by learners that characterized each session, that is the learner' s path followed in the MOOC within a session.

We ordered the different variants of the sessions from the most common to the least common (as well as in subsection 2.3). The most common ones were assigned to a category that described a session pattern. For example, we analyzed the first variants of these sessions and observed that comprised interactions consisting in beginning a video-lecture, then completing or reviewing a video-lecture and then ending the session. Therefore, a pattern of "*Only video-lecture"* was defined (i.e., learners working in sessions only with video-lectures).

As in subsection 2.3 we recommend repeating this procedure several times for analyzing the rest of the variants in the sessions. This was done using the same Python script developed ad hoc to do this classification task. As a result, we obtained twelve types of sessions (interaction patterns) that learners made.

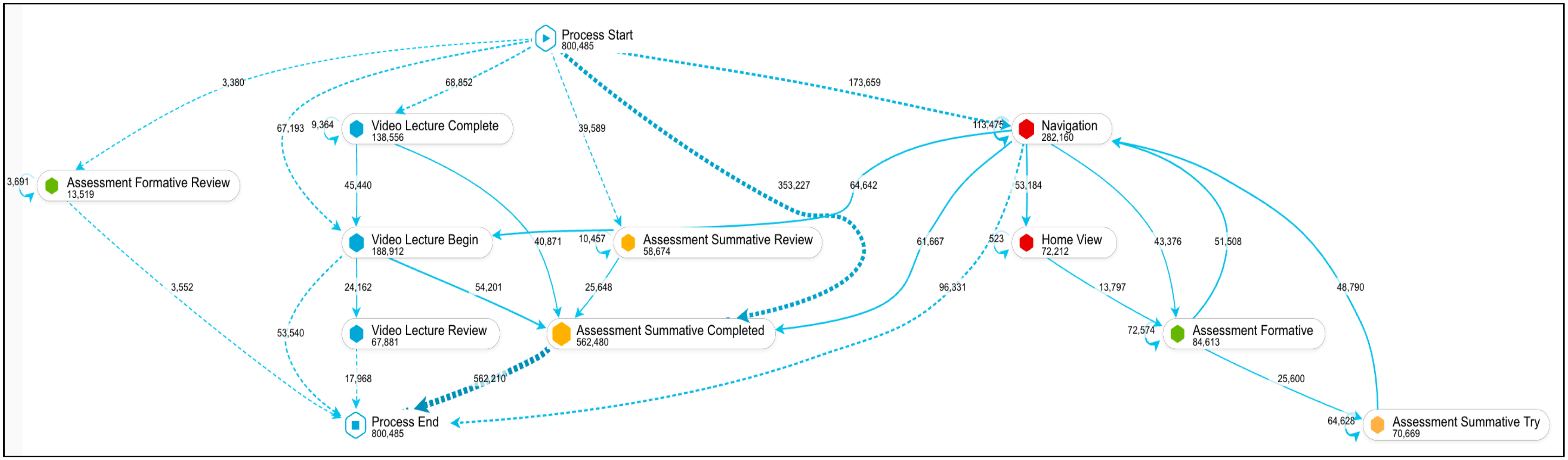


Figure 2-17 Full process model obtained using Celonis software, containing all the interactions by sessions. The process model shows ten possible interactions that learner can perform with the course content. Thick dotted line represents the most common path followed by learners.

2.4.3 Results

Sub-RQ 1.3: To what extend can we replicate (partially or totally) the methodology applied in the previous subsection 2.3 to extract students' learning strategies in a MOOC?

Most of the process developed in subsection 2.3 could be applied to the new MOOC. However, some methodological decisions were made to adapt to the structure and data collected in the edX platform, especially in the data-set extraction and log-data construction. These decisions were:

[Decision 1] Study Sample. The study sample of the synchronous MOOC deployed in edX was composed of online learners that at least completed one video-lecture, unlike in the case of the previous subsection 2.3 in which the sample was composed of learners who completed an SRL survey. This decision was made because two other previous studies (Kizilcec et al., 2017; Maldonado-Mahauad et al., 2018) observed that learners' behavior in the platform was not related with the self-regulatory profile reported in that questionnaire, which is also related to the discussion about the validity of self-reported data in psychological studies (Veletsianos, Reich, & Pasquini, 2016).

[Decisions 2 and 3] Mapping the nature of interactions with course resources. The MOOC structure of the edX course contained more digital resources compared with the ones in Coursera due to the course design characteristics (video-lectures, formative activities, graded activities, navigation between modules, tabs and clicks on the home page). Accordingly, we mapped the course resources with the possible interactions of the learners and defined ten types of interactions instead of the six defined in the previous study (asynchronous MOOC in Coursera).

[Decision 4] Self-reported information. This study did not include a self-reported SRL profile of the students (as it was done in subsection 2.3) as part of the event log. This variable was found to not have an influence in the process of exploring the patterns of the behavior found. However, knowing the self-reported profile of the learners helps to have a better understanding of the characteristics of the students and relate their profile to their actions. To sum up, these four decisions lead us to adapt the methodology developed in subsection 2.3 in a new data set context from a different MOOC platform.

Sub-RQ 1.4: How do students' learning strategies in this new context differ from those from the previous subsection 2.3?

To answer this sub research question, two analyses were conducted. Next, we present the results of these analyses.

a) Analysis of learners' behavioral patterns in a synchronous MOOC in edX.

We obtained twelve types of interaction sequence patterns that learners made when they engaged with the MOOC (see Table 2-22).

Table 2-22 Percentage of sessions patterns based on the number of sessions (N = 800,485) and performed by learners

Session patterns	# sessions (%)
(1) Only assessment-summative-complete	353,090 (44.11%)
(2) Only video-lecture → assessment-summative-complete	107,623 (13.44%)
(3) Only video-lecture	86,306 (10.78%)
(4) Only assessment-summative	80,310 (10.03%)
(5) Only assessment-formative	76,791 (9.59%)
(6) Combined	33,253 (4.15%)
(7) Only assessment	18,205 (2.27%)
(8) Only-video-lecture → assessment-formative	18,000 (2.24%)
(9) Explore	10,095 (1.26%)
(10) Assessment-summative-try → only-video-lecture	9,463 (1.18%)
(11) Others	6,644 (0.83%)
(12) Video-lecture-complete → assessment-summative-try	705 (0.08%)

The description of each interaction sequence pattern was grounded upon whether a session only contained a certain type of interaction (e.g., sessions consisting of only video-lectures without any assessment activity) or whether the session contained certain type of interaction sequences between interactions that are considered important for the learning process (e.g., sessions where learners went from trying a summative-assessment to a video-lecture activity). Once the most common sessions patterns were extracted from the main process model (see Figure 2-17), we obtained a specific process model for each pattern (see an example in Figure 2-18). Twelve distinct types (patterns) of sessions were extracted:

(1) Only assessment-summative-complete: Session pattern in which learners worked only passing graded assessments. This is the most common type of session: 44.11% of the total number of sessions corresponded to this type.

(2) Only video-lecture to assessment-summative-complete: Session pattern in which learners began working with video-lectures (either beginning, completing) and then successfully solved a graded assessment (summative) for the first time (see Figure 2-18): 13.44% of the sessions corresponded to this type.

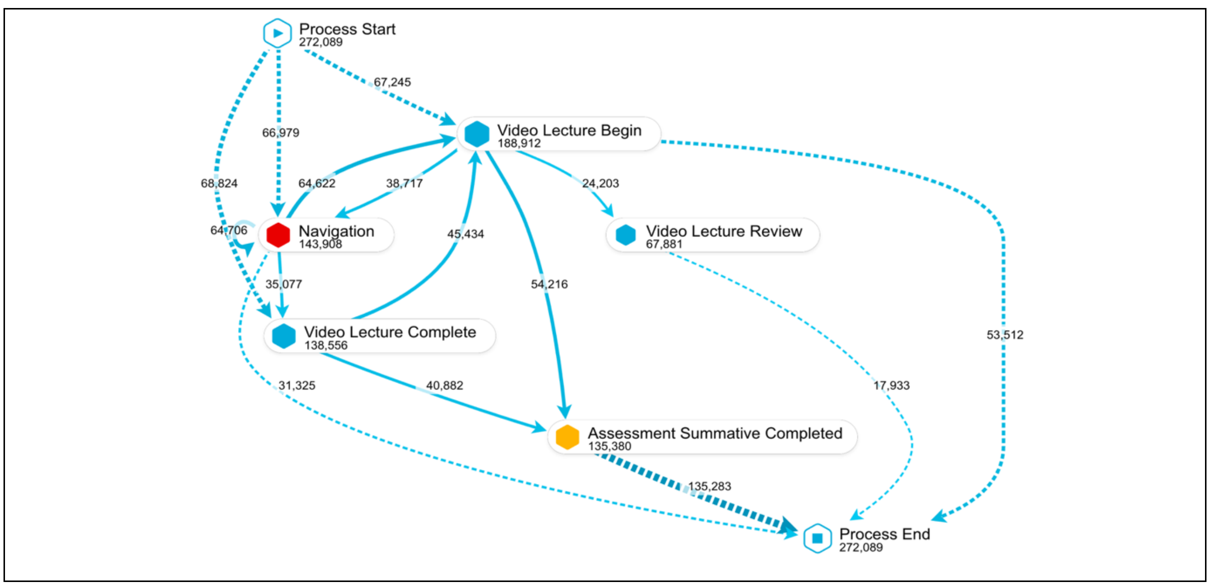


Figure 2-18 Only video-lecture to assessment summative complete session pattern performed by learners in a MOOC.

(3) Only video-lecture: Session pattern in which learners worked only with video-lectures. Learners performed sessions that consisted of watching at least one video-lecture and did not contain assessment activities. Learners could begin, complete, review video-lectures or perform combinations of them (i.e. begin and then complete, begin and then review, complete and then review): 10.78% of the sessions corresponded to this type.

(4) Only assessment-summative: Session pattern in which learners worked only with summative assessments. Learners performed sessions that consisted in trying at least one summative assessment and did not watch any video-lecture. Learners could try, complete, review summative assessments or performed combinations of them (i.e., try and then complete, try and then review, complete and then review) while they were interacting with the course: 10.03% of the sessions corresponded to this type.

(5) Only assessment-formative: Session pattern in which learners worked only with formative assessments. Learners performed sessions that consisted of attempting at least one formative assessment and did not watch any video-lecture. Learners could attempt or review formative assessments or perform combinations of them (i.e., attempt an assessment and then end the session, attempt and then review, review and then end the session): 9.59% of the sessions corresponded to this type.

(6) Combined: Session pattern in which learners combined from two up to four sessions patterns mentioned in this section: when the combination is up to two, all types of sessions were considered as part of this combined session pattern; when the combination is up to three, sessions consisting in work only with video-lectures and only with assessments were not considered as part of this combined session pattern; when the combination is up to four, sessions consisting in working only with video-lectures, only with assessments and explore were not considered as part of this combined session pattern: 4.15% of the sessions corresponded to this type.

(7) Only-assessment: Session pattern in which learners worked between formative and summative assessments in the same session. Learners could attempt to solve or review a non-graded assessment activity (formative) and try to complete (pass) a graded assessment activity (summative) while they were interacting with the course: 2.27% of the sessions corresponded to this type.

(8) Only video-lecture to assessment-formative: Session pattern in which learners began working with video-lectures (either beginning, completing or reviewing) and then attempted to solve a non-graded activity at the first time: 2.24% of the sessions corresponded to this type.

(9) Explore: Session pattern in which learners worked only beginning video-lectures (without completing) or attempting some non-graded formative assessments.

(10) Assessment-summative-try to Only-video-lecture: Session pattern in which learners attempted to solve a graded activity incorrectly and then worked with video-lectures (begin, complete, review video-lectures or combinations of them).

(11) Video-lecture-complete to assessment-summative-try: Session pattern in which learners completed a video-lecture and then attempted to solve a graded activity without managing to do it.

(12) Others: We have classified as other to those sessions that were long and disperse, as they do not fit into any of the above-mentioned session patterns.

b) Learners' classification according to their behavior and learning outcomes.

To answer this question learners (N = 50,776) were grouped based on the identified sessions patterns. We use the agglomerative hierarchical clustering as subsection 2.3 recommended. This led to selecting the solution with 4 clusters (see Figure 2-19). Table 2-23 describes the resulting clusters in terms of: (1) the ten session patterns used for grouping the learners (we discarded *video-lecture-complete to assessment-summative-try* and *others* given that both types are less than 1% of all sessions), (2) the mean in terms of session performed, (3) the number of learners, (4) the number of learners that passed/failed the course.

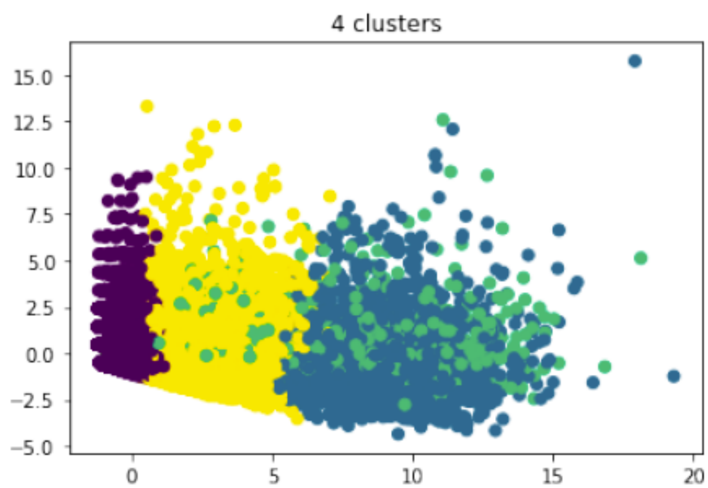


Figure 2-19 Scatter Plot with silhouette score 0.571

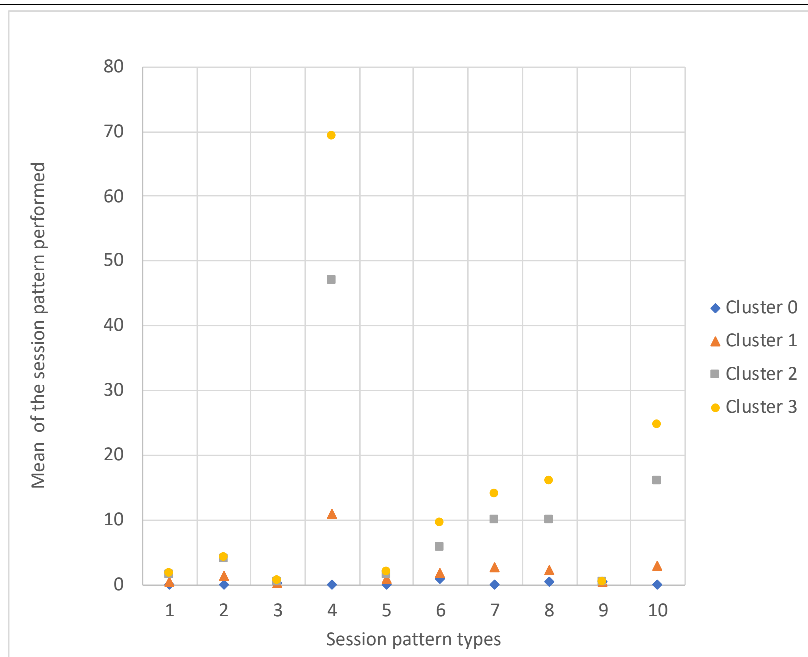


Figure 2-20 Mean of every session pattern by cluster. The numbers in “x” axis represents the patterns found (Table 2-22) and the numbers in “y” axis represents the mean of the patterns performed in a session

Table 2-23 Means of session patterns per cluster (N = 800,485) performed by learners

Session patterns	**Cluster 0**	**Cluster 1**	**Cluster 2**	**Cluster 3**
	mean (sd)	**mean (sd)**	**mean (sd)**	**mean (sd)**
1 - Assessment-summative-try → only-video-lecture	0.006 (0.079)	0.294 (0.625)	1.492 (1.635)	1.627 (1.685)
2 - Combined	0.012 (0.111)	1.269 (1.339)	3.914 (3.403)	4.086 (3.362)
3 - Explore	0.149 (0.388)	0.232 (0.568)	0.395 (0.667)	0.626 (0.851)
4 - Only-assessment-summative-complete	0 (0)	10.791 (13.664)	47.031 (32.673)	69.157 (27.796)
5 - Only-assessment	0.007 (0.088)	0.741 (1.101)	1.431 (1.313)	2.043 (1.352)
6 - Only-assessment-formative	0.765 (0.0956)	1.787 (2.316)	5.708 (4.906)	9.541 (5.429)
7 - Only-assessment-summative	0.004 (0.075)	2.653 (3.888)	9.961 (6.938)	14.031 (6.198)
8 - Only-video-lecture	0.336 (0.686)	2.227 (3.182)	9.943 (8.635)	15.909 (9.635)
9 - Only-video-lecture → assessment-formative	0.371 (0.617)	0.323 (0.731)	0.407 (0.724)	0.376 (0.641)
10 - Only-video-lecture → assessment-summative-complete	0 (0)	2.861 (4.303)	15.926 (11.815)	24.591 (10.642)
N_sessions on average per cluster	1.697 (1.379)	23.371 (25.593)	97.774 (61.687)	142.761 (45.771)
N_learners	30,415	17,829	651	1,881
Fail_course	30,415	17,786	492	1,005
Pass_course	0	43	159	876

The resulting clusters indicate different types of learning strategies deployed by learners while they were facing the MOOC. If we compare the resulting clusters obtained in Table 2-23 (4 clusters) with those obtained in Table 2-16 in subsection 2.3 (where only 3 clusters were obtained) we can see that one new cluster emerged. This is given the type of activities introduced in the edX MOOC. Specifically, edX MOOC contain formative and summative assessments unlike Coursera MOOC, where only summative assessments were considered as evaluation activities. This bring us new insights about how the design of the course influenced over the behavior deployed by learners in the course (i.e., deploying different learning strategies). If we look for specific differences between the different clusters, we can describe them as follows (see Table 2-23; Table 2-24; Table 2-25 and Figure 2-20):

Cluster 0 – Sampling learners: this cluster was composed of learners that on average visited only once or twice the course exploring the course content. Specifically, they visited the video-lectures and follow through the proposed path by the course to visit formative assessments but without attempting or ending any activity proposed, just exploring the content to see the big headlines (as the same cluster 0 in subsection 2.3). This cluster is composed of the largest number of learners (n = 30,415), but they fail passing the course.

Cluster 1 – Targeting learners: this cluster was composed of learners that on average performed a low number of sessions. Although they were active learners, they had low activity in the course in comparison with the next groups (clusters 2 and 3, see Table 2-23, Table 2-24 and Table 2-25). They worked superficially with the course materials. These learners after watching video-lectures attempted to pass summative assessments leaving formative assessment aside (sessions were mainly oriented to passing the summative assessments). This behavior shows that learners in this cluster focused on passing the course more than on achieving a deep understanding of the contents and self-evaluating their progress (as the same cluster 1 in subsection 2.3). This cluster is composed of a great number of learners (n = 17,829), but only a few of them passed the course (n = 43, compared with clusters 2, 3).

Cluster 2 – Low Comprehensive learners: this cluster was composed of learners that on average performed a large number of sessions in comparison with the previous two groups (clusters 0, 1). They worked intensively with the course materials. These learners watched the video-lectures, attempted formative and then summative assessments (which is the path designed by the instructors in the course). They focused on summative more than formative assessments (see Table 2-23, Table 2-24 and Figure 2-20). Also, after watching video-lectures they intended to pass summative assessments and worked less with formative assessments (in comparison with cluster 3). However, learners in this cluster performed more sessions working with summative assessments than with formative ones. In this cluster, a large number of learners passed the course (n = 159, in comparison with cluster 1).

Cluster 3 – Highly Comprehensive learners: this cluster was composed of learners that on average performed a large number of sessions and worked with more intensity with the course contents than learners in the rest of the clusters (see Table 2-23; Table 2-24 and Figure 2-20). Learners in cluster 3 performed more sessions that consisted in working with video-lectures before they passed a summative assessment. Also, they performed more sessions either with formative or summative assessments in comparison with learners in cluster 2. This behavior showed the intention of learners to achieve a deep understanding of the contents and self-evaluate their progress. Learners in this cluster also performed sessions in which they worked intensively only with video-lectures in comparison with the rest of the learners in the different clusters.

Table 2-24 Differences in session patterns between cluster 2-3

Session patterns	**Cluster 2**	**Cluster 3**	***t***	***p***	***r***
	mean	**mean**			
Assessment-summative-try → only-video-lecture	1.492	1.627	-0.953	0.342	0.064
Combined	3.914	4.086	-1.441	-0.499	0.072
Explore	0.395	0.626	-6.943	**<.001****	0.395
Only-assessment-summative-complete	**47.031**	**69.157**	-8.028	**<.0001****	0.493
Only-assessment	1.431	2.043	-5.382	**<.0001****	0.339
Only assessment-formative	5.708	**9.541**	-8.911	**<.0001****	0.504
Only assessment-summative	**9.961**	**14.031**	-6.913	**<.0001****	0.434
Only video-lecture	**9.943**	**15.909**	-7.868	**<.0001****	0.457
Only-video-lecture → assessment-formative	0.407	0.376	0.505	0.614	0.035
Only video-lecture → assessment-summative complete	**15.926**	**24.591**	-8.633	**<.0001****	0.515
N_sessions on average per cluster	97.774	142.761	-17.05	**<.0001****	0.493
N_learners	**651**	**1,881**			

Note. ** *p < .05;* Marks statistically significant differences

Finally, Table 2-25 presents comparisons between the four clusters based on the distributions of the session patterns. Between clusters 2 and 3 there are no statistically significant differences, while pair comparisons between clusters 0-1, 1-2 and 1-3 showed statistically significant differences.

Table 2-25 Comparisons between clusters of learners based on the session patterns

Cluster #	**Cluster #**	$\chi 2$	p
0	1	281.3519	**0.000***
1	2	194.9919	**0.000***
1	3	529.9969	**0.000***
2	3	15.1820	0.231

Note. *Marks statistically significant

Learners in clusters 2 and 3, classified as low and highly comprehensive learners respectively, behaved differently in terms of passing the course. Although learners in these clusters worked on average the same number of sessions in the course (no statistical differences observed), their study strategies differ (Table 2-26).

Table 2-26 Differences in sessions patterns performed on average by learners in clusters 2-3 that passed the course

Session patterns	**Cluster 2 (pass)** mean	**Cluster 3 (pass)** mean	t	p	r
Assessment-summative-try → only-video-lecture	2.252	1.495	4.896	**<.001****	0.326
Combined	5.899	3.825	6.244	**<.001****	0.409
Explore	0.346	0.509	-3.073	**.002****	0.179
Only-assessment	1.906	2.128	-2.0413	**.045****	0.135
Only-assessment-formative	10.943	11.857	-2.085	**.038****	0.140
Only-video-lecture → assessment-summative complete	30.623	31.859	-2.205	**.028****	0.139
N_sessions on average per cluster	88.811	88.814	-0.0029	0.998	0.000
N_learners	**159**	**876**			

Note. ** $p < .05$; Marks statistically significant differences

Highly comprehensive learners (cluster 3): 1) worked more in sessions that consisted in watching video-lectures and then passing summative assessments, 2) worked more with formative assessments and worked in combination with summative and formative assessments, and 3) on average explored more the course contents. In contrast, low comprehensive learners (cluster 2): 1) worked more in sessions in which they tried to pass a summative assessment (but failed) and then went back to work with video-lectures (begin, complete or review), and 2) worked more with combinations of the different session patterns in comparison with highly comprehensive learners. In addition, low comprehensive learners tried to pass summative assessments but when failing, they work in video-lectures, probably trying to find information in the video-lectures that helped them to pass the summative assessments. In contrast, highly comprehensive learners worked first with video-lectures and then passed summative assessments. This behavior suggests that this type of learner is trying to achieve a deep understanding of the contents and self-evaluate their progress working more with formative assessments.

2.4.4 Discussion

Even if conducting the same study across different context is complicated by variations in instructional conditions (Gašević et al., 2016). In this subsection, we made an effort of replicability of the PM methodology developed in subsection 2.3 and applied to a data set of a synchronous MOOC in the edX platform. Two main results were obtained. Firstly, the PM methodological approach can be replicated, but it requires taking 3 key decisions that are dependent to the context of application: (1) the sample size, which will vary from experiment to experiment; (2) mapping the nature of the interactions based on the structure of the MOOC under analysis, but keeping the metric of session and interaction; and (3) eliminating students' SRL profile obtained from a SRL questionnaire as a control measure. Secondly, the adaptation of this methodological approach extends the findings in subsection 2.3 by identifying new learning strategies (one new cluster) that are highly dependent on the course structure. In contrast to the six self-regulatory patterns and three groups of learners identified in the prior work, we identified twelve patterns and four groups: 1) Sampling learners, 2) Targeting learners, 3) Low Comprehensive learners, and

4) Highly Comprehensive learners. The present findings have implications both for (1) the methods used in the LA community for analyzing trace data, and (2) for theory and practice of SRL.

Regarding the implications in LA methods, the work developed in subsections 2.3 and 2.4 sheds some light on the aspects to be considered when doing replication studies using students' trace data. Replicating an analytical method requires taking decisions about how raw data is processed. In order to evaluate the reproducibility of the results, these decisions should be carefully reported, especially when they require some level of pre-processing or abstraction. When applying PM approaches, the data pre-processing and data abstraction is key. For example, how students' work session is defined or how student's interactions with the course content are mapped into a logfile may have an impact on how learners' strategic patterns are observed. This study shows that, when replicating methodological approaches based on PM, the granularity of the data when defining students' interaction should be maintained from one study to another. That is, if student's interaction with the course content is defined by interaction with a particular resource, this should be the level of granularity for the analysis, and no combinations of interactions should be considered for the analysis. In current literature, most of studies take as a reference the interactions with the course content as a basis (Jovanović et al., 2017; Saint et al., 2018); however, this could vary when changing platform, since the nature of the data collected may vary. The results of this subsection emphasize the importance of including the decision-making process on data preprocessing as part of any analysis in order to be able to compare the results from one study to another. Moreover, this pre-processing should consider simplifying the raw data by keeping only those types of interaction that could be translated from one platform to another, even if this means losing some data in the process. Of course, simplifying the data may mean also simplifying the results, but more studies of this type should be reported so that the community arrives to agreements such as a standard of a minimum logfile to facilitate replication studies.

Regarding the implications for SRL theory and practice. The adaptation of this methodology extends the findings in subsection 2.3 by identifying new learning strategies that are highly dependent on the course structure. Twelve sessions patterns and four groups of learners were found. Learners classified as Sampling and Targeting in this subsection are similar to those found in subsection 2.3. However, in contrast to subsection 2.3, Comprehensive learners can be classified into highly and low comprehensive. Highly comprehensive learners seemed to be deeper learners following the designed path of the course, trying to achieve a deep understanding of the contents and self-evaluating their progress through the intensive work with formative activities. In contrast, low comprehensive learners seemed to be more strategic, following a pattern that consisted in passing summative activities and working less with formative ones.

While in subsection 2.3 analyzed a MOOC with only summative assessment activities, the MOOC in this subsection included more than 160 formative activities. These results suggest that the strategies adopted by the learners are highly dependent on the context, and in particular, on the course content and structure. Moreover, these results align with prior work that show how course structure and design conditions students' behavior (Alario-Hoyos et al., 2014; Ferguson et al., 2015; Laurillard, 2013).

However, more studies, and particular A/B experimental experiments, should be conducted in order to provide robust evidences on how context affects learners' behavior. Moreover, and beyond replication efforts, we believe that the identified behavioral patterns can inform the design of learning environments by either supporting the implementation of precise learner modelling or by providing enough scaffolding to at-risk learners who remain working actively in the MOOC.

2.5 Conclusions

This chapter has presented the work performed in order to answer the **RQ1: *What instruments and methods are more appropriate to explore learners' self-regulatory strategies used in MOOCs?*** To address the RQ1, four sub research questions (Sub-RQ) have been proposed:

- Sub-RQ 1.1: What SRL models and SRL strategies have been studied in traditional and online contexts?
- Sub-RQ 1.2: What are the most frequent interactions sequences of learners in MOOC?
- Sub-RQ 1.3: To what extend can we replicate (partially or totally) the methodology applied in the previous study (subsection 2.2) to extract students' learning strategies in a MOOC?
- Sub-RQ 1.4: How do students' learning strategies in this new context differ from those from the previous study (subsection 2.2)?

The work developed to address these four Sub-RQ has allowed to achieve three main contributions in this thesis that are presented below:

The first contribution is a questionnaire adapted to the context of the MOOC that allows measuring the SRL as an aptitude. The questionnaire consists of 22 questions that consider 5 SRL strategies: Self-efficacy, Goal setting, Study environment management, Organization and Help-seeking. One of the added values of this questionnaire is that it was built upon an exhaustive and systematic bibliographic review on the work done on questionnaires to measure SRL and related works in recent years. This bibliographic review serves not only to demonstrate the limitations of existing questionnaires when analyzing self-regulatory profiles in a MOOC context, but also to systematically organize the work conducted in this line.

The second contribution is a methodology based on process mining for studying SRL in MOOCs as a process. This methodology consists of four stages (i.e., extraction, event log generation, model discovery, model analysis) and helps to identify and analyze the digital records of the learners' activities in a MOOC that account for the SRL strategies. One of the added values of this methodology is that it combines an aptitude-based approach with a process-based approach to investigate SRL strategies in MOOCs across contexts by relying on both a self-report instrument and PM of behavioral learner data. This methodology allows us to: (1) identify the most frequent interaction sequence patterns that learners exhibit in a MOOC; (2) to differentiate interaction sequence patterns between learners with different characteristics; (3) to identify learner profiles based on their observed interaction sequence patterns and; (4) to associate observed interaction sequence patterns with SRL strategies established in SRL theory. This novel way of measuring SRL "on the fly" characterizes the SRL as a process and not as a trait and allows to study the self-regulatory processes of the students at a specific moment during their activity.

The third contribution is the adaptation of this methodology across contexts. This methodology was proposed using a dateset from Coursera platform, and was validated using a dataset from edX platform. This adaptation revels us that the instructional design proposed in the course, influenced in the behavior of the learners and the set of learning strategies that they deploy in a MOOC. This contribution will help in proposing future development of tools to scaffold specific strategies at the same time that serves for capturing students' "records" about their current behavior.

Chapter 3

Relationship between SRL strategies and academic performance

If you can't explain it simply, you don't understand it well enough.

Albert Einstein

This chapter shows the main contributions related to the second research question: *"Relationship between the SRL strategies and the academic performance of the students in a MOOC"*. Specifically, this chapter shows the main conclusions about this research question considering the influencing factors in the SRL strategies such as the characteristics of the students, the MOOC characteristics and the context in which the MOOC is deployed. This chapter is structured in 6 subsections collecting the results reported in four journal articles [Table 1-2; J3, J4, J5, J6] and five Conference articles [Table 1-2; C1; C2; C3; C4; C5]. Specifically, the subsection 3.1 is an introduction where sub research questions (Sub-RQ) from the main RQ2 are shown and the contributions associated to each one of them. The subsection 3.2 shows SRL strategies that are most helpful to achieve personal goals and is related to the learners' intentions; the subsection 3.3 shows the classification of learners based on the relationship between SRL strategies deployed and their achievements in MOOCs; the subsection 3.4 shows SRL strategies that predict learners' success in MOOCs; the subsection 3.5 shows SRL strategies employed by learners in a MOOC deployed in a blended learning scenario. Finally, the subsection 3.6 shows the main conclusions of the chapter.

3. RELATIONSHIP BETWEEN SRL STRATEGIES AND ACADEMIC PERFORMANCE

3.1 Introduction

This chapter presents the results related to *RQ2: What is the relationship between SRL strategies and academic performance, taking into consideration the characteristics of the participants, the MOOC and the course context that influence the use of these strategies?* All the results have been reported in 4 journal papers and 5 conference papers. Each journal and conference paper address a particular sub research question derived from the main research question RQ2. Table 3-1 summarizes the main sub research questions addressed in each paper and the specific objective to which they are related.

Table 3-1 List of sub research questions related to the RQ1. J[x] and C[x] are the identificators used to refer to journal and conference papers respectively, where “x” indicates the number of the paper.

Specific Objective	Publication	Sub-research Question
RQ2. What is the relationship between students’ SRL strategies and their academic performance in a MOOC, taking into consideration the characteristics of the participants, the MOOC and the course context that influence the use of these strategies?		
[Obj.3]-[Ch.3]	**[J5]** Self-Regulated Learning Strategies Predict Learner Behavior and Goal Attainment in Massive Open Online Courses	Sub-RQ 2.1: Which self-reported SRL strategies are most helpful to achieve personal course goals?
[Obj.5]-[Ch.5]	**[C3]** Recommending self-regulated learning strategies does not improve performance in a MOOC	Sub-RQ 2.2: How do self-reported SRL strategies vary by individual learner characteristics?

	[C1] Exploring differences in how learners navigate in MOOCs based on self-regulated learning and learning styles: A process mining approach.	Sub-RQ 2.3: How do the different SRL and Learning Style profiles manifest themselves in a MOOC in terms of learning sequences?
[Obj.4]-[Ch.4]	**[C2]** Predicting Learners' Success in a Self-paced MOOC Through Sequence Patterns of Self-regulated Learning. **[J6]** Design of a tool to support self-regulated learning strategies in MOOCs	Sub-RQ 2.4: Which indicators of SRL obtained from self-reported questionnaires and activity sequence extracted from trace data can predict course success in self-paced MOOCs?
[Obj.6] – [Ch5] & [Obj.7] – [Ch5]	**[C4]** Flipping the classroom with MOOCs. A pilot study exploring differences between self-regulated learners **[C5]** Analyzing students' SRL strategies when using a MOOC as a Book	Sub-RQ 2.5: How does the behavior of the learners with different SRL profiles differ when a MOOC is used as part of a Flipped Classroom proposal?

Each subsection in this chapter is structured as follows. First, the context to frame the sub research questions addressed in each paper is presented. Second, we present related work. Third, the analytical methods used to answer the sub research questions are presented. Fourth, the main results are presented. This chapter ends with a conclusion that summarizes the lessons learned of each sub research question in order to inform the RQ2.

3.2 Individual differences in SRL strategies of the learners that are most helpful to achieve their personal goals and intentions in MOOCs

A primary goal of MOOCs has been to provide more people with an opportunity to learn and grow. Most learners who enroll in MOOCs selectively engage with parts of the course content and a small proportion eventually completes the course (Anderson et al., 2014; Breslow et al., 2013; Evans, Baker, & Dee, 2016; Ho et al., 2015; Kizilcec et al., 2013; Perna et al., 2014; Seaton, Bergner, Chuang, Mitros, & Pritchard, 2014). This variation in behavior can be partly attributed to the remarkable diversity of leaners' backgrounds, motivations, intentions, and prior experiences (de Barba, Kennedy, & Ainley, 2016; Kizilcec & Schneider, 2015; Zheng et al., 2015). In fact, only half of the survey respondents in a typical MOOC report that they intend to complete the course to receive a certificate (Kizilcec & Schneider, 2015; Littlejohn & Milligan, 2015; Reich, 2014). However, even among learners who hold ambitious goals for the course and express a commitment to achieve them, a majority of learners remains unsuccessful. The primary reasons why learners drop out of MOOCs are related to their time management and course difficulty, according to quantitative (Kizilcec & Halawa, 2015; Nawrot & Doucet, 2014) and qualitative (Zheng et al., 2015) accounts. This raises the question of how to support learners in achieving their goals in learning environments like MOOCs.

To address this question, we investigated self-regulation strategies in MOOCs. Our work builds on SRL theory, which describes ways for learners to take control of their learning process (cf. Chapter 2). We examined which self-regulation strategies predict attainment of personal course goals, and how strategies vary by individual characteristics. The goal of this subsection is to provide a foundation for future research and interventions that support SRL in MOOCs and comparable environments. We leveraged MOOCs as an environment in which to investigate authentic learner behavior over time—a research paradigm that holds promise for advancing educational science and practice (Reich, 2015; Winne & Nesbit, 2010)—and used methods from educational data mining and learning analytics in our analysis (Roll & Winne, 2015; Winne & Baker, 2013). We surveyed 4,831 online learners across six distinct MOOCs about their SRL strategies and individual characteristics, including demographics, motivations, and intentions for completing course material. Their responses were combined with detailed records of their interactions with course content and their overall course achievement, yielding a longitudinal account of in vivo SRL in context.

This subsection makes two contributions to the literature on SRL. First, we provide new insights into SRL in MOOCs for a heterogeneous adult learner population. Second, leveraging the heterogeneity of the present sample, we demonstrate multiple individual differences in SRL that can inform targeted interventions, such as adaptive scaffolding.

3.2.1 Related Work

3.2.1.1 Self-regulated learning strategies for achieving personal goals

Online learners are expected to actively and autonomously engage in the learning process (Wang, Shannon, & Ross, 2013). This demands a high level of confidence in learners' own abilities and the ability to manage their own learning process (Liang & Tsai, 2008; Sun & Rueda, 2012; Tsai, Chuang, Liang, & Tsai, 2011). Learners who struggle to regulate their learning process effectively tend to experience frustration and become less engaged in the course (Sun & Rueda, 2012), and consequently, they are less successful (Lee, Shen, & Tsai, 2008; Samruayruen, Enriquez, Natakuatoong, & Samruayruen, 2013; Tsai, 2009). Prior work in online learning environments demonstrated improvements in academic achievement from applying SRL strategies, especially time management, metacognition, and effort regulation strategies (Azevedo & Aleven, 2013; Broadbent & Poon, 2015; Niemi, Nevgi, & Virtanen, 2003). These strategies help learners process and retain knowledge in a structured manner (Beishuizen & Steffens, 2011; Dignath & Büttner, 2008; Pintrich, 2004; Zimmerman, 2008). Several studies found that providing scaffolding for these strategies can support SRL and raise achievement (Azevedo, Moos, Greene, Winters, & Cromley, 2008; Kim & Hodges, 2012; Taub et al., 2014).

Studies of attrition in MOOCs suggest that metacognitive strategies and resource and task management strategies are critical for success. In particular, learners' reasons for disengaging from MOOCs can inform which SRL strategies are potentially valuable. Kizilcec & Halawa (2015) examined reasons for disengaging in a sample of 1,698 learners across 20 MOOCs and identified four clusters (in order of significance): time issues, course difficulty, format and content, and goals and expectations. A follow-up study, targeted at learners predicted to have just disengaged from a MOOC, confirmed this pattern using open-ended responses that were coded: 84% of respondents mentioned that they had "not enough time for the course." Half of the 84% who faced time issues also indicated being easily distracted from the course, which suggests that better metacognitive and resource management strategies could have prevented their disengagement. Additionally, satisfaction and relative progress in the MOOC were associated with goal striving, which relates to goal setting and strategic planning strategies. Zheng et al. (2015) interviewed 18 learners about their experiences in MOOCs and the issue of not having enough time also emerged in their analysis. Moreover, the lack of pressure emerged as another factor influencing retention, which relates to task management strategies, such as effort regulation. Learners also missed a sense of community, which would limit the use of help-seeking strategies. Finally, Nawrot and Doucet (2014) present further evidence that the most common reasons for disengaging from MOOCs can be related to task management (e.g., time management) and metacognitive strategies (e.g., strategic planning, goal setting).

Besides reasons for disengaging from MOOCs, what insights can successful learners offer about strategies that were helpful? Interviews of 17 learners who successfully completed a MOOC helped identify several ostensibly effective behavioral patterns (Kizilcec et al., 2016). A number of them were related to task management strategies, such as reserving time in the week for studying (time management), starting and finishing a chapter on the same day (task strategies, effort regulation), and working with others on the course (help seeking). Other patterns reflected metacognitive strategies, such as having clear objectives and planning around those (goal setting, strategic planning), applying what one has learned in the course to internalize it, and creating summaries or mind maps of lecture content (self-evaluation, self-monitoring). This account of success strategies complements the findings on reasons for attrition. Overall, based on the findings above mentioned, in this subsection we focused only in goal setting, strategic planning, self-evaluation, task strategy, elaboration and help seeking SRL strategies that are expected to support learners in MOOCs.

3.2.1.2 Individual differences in self-regulated learning and online course behavior

Prior research has investigated how individual differences between learners might relate to self-reported SRL and behavior in MOOCs. Learners who report higher levels of motivation, commitment to learn, formal education, and relevant prior knowledge also indicate higher levels of SRL (Hood et al., 2015; Littlejohn et al., 2016) and exhibit higher course achievement (Guo & Reinecke, 2014; Jaggars, 2014; Kay et al., 2013; Kizilcec & Halawa, 2015; Laplante, 2013). By contrast, several investigations have found no significant gender differences in terms of SRL in the context of various digital learning environments (Basol & Balgalmis, 2016; Liou & Kuo, 2014; Yukselturk & Top, 2013). Hood et al. (2015) examined how learners' context (i.e., background characteristics) influences their ability to self-regulate their learning in MOOCs. They found higher levels of SRL among learners with a higher level of formal education and among working professionals in domains related to the course content. Littlejohn et al. (2016) found differences between learners with varying levels of SRL in their reported motivations and goals for the course, which apparently shaped their approach to the MOOC and their use of learning strategies. On the basis of in-depth interviews, they identified differences in self-described learning behaviors between learners with low versus high SRL profiles for five SRL sub-processes. Moreover, numerous studies have found individual differences in learners' engagement and achievement in MOOCs.

Empirical investigations have linked variation in course behavior and achievement with various individual differences: learners' demographic and personal background (Evans et al., 2016; Guo & Reinecke, 2014; Hansen & Reich, 2015; Kizilcec & Halawa, 2015); motivations for enrolling and intentions for the course (de Barba et al., 2016; Jordan, 2014; Kizilcec & Halawa, 2015; Kizilcec & Schneider, 2015; Reich, 2014), and self-efficacy

(Wang & Baker, 2015). Guo and Reinecke (2014) analyzed the navigation strategies of course certificate earners by age and country of origin. They found older learners and learners from countries with fewer teachers per student to take less linear paths through the course content—potentially a manifestation of lower SRL skills. Based on a sample of over 67,000 learners across 16 MOOC, Kizilcec and Halawa (2015) found higher grades and levels of persistence among male learners, and those with more formal education, stronger time commitment to the course, prior experience with the course topic, an intent to complete the course, and who were located in the Global North. Across 68 courses, Hansen and Reich (2015) found that U.S. learners with lower socioeconomic resources were also less likely to enroll in and complete MOOCs, especially among adolescents and young adults. To summarize, prior work has identified individual differences in terms of SRL and in terms of behavior and achievement in MOOCs. Thus, in a context with a highly heterogeneous learner population, individual differences warrant further empirical investigation.

3.2.1.3 Research questions

The current literature offers several accounts of SRL in MOOCs and individual differences based on characteristics such as learners' formal education, prior knowledge, and their professional context. This prior work provides a basis for deeper investigations of SRL in large-scale online learning environments. We identified two gaps in our current understanding of SRL in online learning that warrant further investigation.

First, we need to advance our understanding of the relation between self-reported SRL strategies and objective behavioral measures in a large-scale learning environment over time. As noted above, prior work suggests that learners' self-reported SRL strategies have an influence on how they approach MOOCs, and prior studies have examined SRL in small-scale online environments. However, how SRL manifests in the actual interactions with course content in MOOC has received no scholarly attention. Moreover, we found no evidence on the relative efficacy of different SRL strategies to support online learners achieve personal learning goals over time. We identified six SRL strategies that have been related to academic achievement in online learning and MOOCs in prior work (see Section 3.2.1.1). However, the relative extent to which these SRL strategies predict differences in achieving personal goals in MOOCs is unknown. Accordingly, we pose the following sub research question: *Sub-RQ 2.1: Which self-reported SRL strategies are most helpful to achieve personal course goals?*

Second, we need to advance our understanding of individual differences in SRL. Prior work found individual characteristics of learners such as their level of education, gender, age, course intentions, and motivations to be associated with performance in the course. For example, prior investigations have demonstrated that learners with more formal education self-report stronger SRL skills and exhibit higher persistence and achievement. However, there has not been a systematic analysis of individual characteristics that predict learners' self-reported SRL, because this demands a large and diverse survey sample of learners, which is rarely available outside of MOOCs. Insight into individual differences in SRL could improve targeted scaffolding interventions, for example, by informing Bayesian priors in models. We will identify a broad set of individual differences in SRL in terms of characteristics, many of which were examined in prior work (demographics, course intentions, motivations, etc.) to investigate the following research question: *Sub-RQ 2.2: How do self-reported SRL strategies vary by individual learner characteristics?*

3.2.2 Methods

3.2.2.1 Context: Sample and MOOC

This final study sample is a subset of the 6,709 learners who answered the initial course survey about their SRL strategies and various individual characteristics, including demographics, course intentions and motivations. The final study sample included 4,831 online learners in six distinct MOOCs. The courses, offered by Pontificia Universidad Católica de Chile through Coursera, were taught in Spanish and followed a self-paced format, such that course materials were available all at once without deadlines. The courses were concerned with different subjects, including topics in Engineering, Computer

Science, Management, Transportation, and Education. Each course encompassed 6-10 sections, with 5-10 video lectures and several assessment items (e.g., multiple-choice quizzes, peer-review activities) per section. Most course assessments were formative and could be attempted multiple times. The target audiences of these courses were high school & college students and professionals in subject-related industries. To achieve a high level of generalizability, the courses selected for this study cover a wide spectrum of subject domains, which was expected to also attract a highly diverse learner audience. In fact, based on self-reports, the average age was 32.0 (SD = 10.8), 26% were women, 63% held a bachelor's or higher degree (15% a master's or Ph.D.), 60% were employed, and 25% were students. Data was collected between April and December 2015.

3.2.2.2 Measures

Participants completed an optional course survey when entering the course for the first time. The survey included the following standard measures: demographics (age, gender, education, occupation), time commitment (hours per week), course intentions (intend to watch all lectures; intend to complete all assessments), prior experience with the course topic, the number of prior online courses started and the number of completed ones. The survey also included the Online Learning Enrollment Intentions (OLEI) measure (Kizilcec & Schneider, 2015) translated into Spanish and a measure of SRL. In this research, we did not use the MOOC-SRLQ questionnaire developed in chapter 2 subsection 2.2, given that this instrument was built after we run several experiments, testing different questionnaires in order to detect their weaknesses and strengths. For this reason in this subsection and in subsection 3.4 the SRL measure used was adapted from the questionnaires used by Littlejohn and Milligan (2015) and Barnard et al. (2008), which are based on several established instruments (Barnard-Brak, Paton, & Lan, 2010; Pintrich & others, 1991; Rigotti, Schyns, & Mohr, 2008; Schraw & Dennison, 1994; Warr & Downing, 2000).

Based on our review of SRL strategies in online learning environments (see Section 3.2.1.1), we selected six strategy subscales from the original instrument (items previously used by Azevedo et al., 2008; Taub et al., 2014). The resulting questionnaire had participants rate 23 statements about SRL strategies for how characteristic they were for them on a labeled 5-point scale (coded 0 to 4): goal setting strategies (4 statements), strategic planning (4), self-evaluation (3), task strategies (6), elaboration (3), and help seeking (3). The order in which statements were presented in the survey was randomized. The individual score for each strategy was computed by averaging ratings of corresponding statements. Table 3-2 provides descriptive statistics for the collected SRL survey data with an exemplary statement for each strategy and a composite computed by averaging scores for all strategies. The SRL measure had high reliability for all strategy subscales with Cronbach's α of at least 0.75, despite the small number of items used. As shown in Table 3-2, the help-seeking subscale had a lower mean and lower correlation with the composite; this may be partly because it was the only subscale that included a reverse-coded item.

Table 3-2 Descriptive statistics for each SRL strategy and an average SRL composite ($\bar{x}$ with exemplary statements, mean and standard deviation, Chronbach's α, and pairwise Pearson's correlation coefficients

Strategy	**Example Statement**	***M* (*SD*)**	**α**	**2.**	**3.**	**4.**	**5.**	**6.**	**$\bar{x}$**
1. Goal Setting	I set realistic deadlines for learning.	3.0 (0.76)	.86	.70	.48	.57	.46	.29	.78
2. Strategic Planning	I organize my study time to accomplish my goals to the best of my ability.	3.1 (0.65)	.75		.60	.66	.58	.32	.84
3. Self-evaluation	I think about what I have learned after I finish.	3.3 (0.66)	.80			.63	.59	.25	.74
4. Task Strategies	When I study for this course, I make notes to help me organize my thoughts.	3.1 (0.62)	.78				.72	.35	.87
5. Elaboration	When I am learning, I try to relate new information I find to what I already know.	3.3 (0.64)	.77					.32	.77
6. Help Seeking	When I do not understand something, I ask others for help.	2.6 (0.79)	.77						.58
$\bar{x}$ SRL Composite	–	3.0 (0.52)	.92						

3.2.2.3 Analytic Approaches

To addresss ***Sub-RQ 2.1*** about the ***relationship between SRL and achieving personal course goals*****,** we assess associations between each strategy and course outcomes depending on learners' stated course goal. We used non-parametric Spearman correlation coefficients, because the outcome data was either binary or skewed. Additionally, we fitted logistic regression models to evaluate the predictive power of the six SRL strategies simultaneously.

To address ***Sub-RQ 2.2*** about ***individual differences in self-reported SRL strategies***, we considered 27 individual learner characteristics. The self-reported characteristics encompassed learners' demographics (8 predictors) and time commitment, their experience with the course topic, their prior experience with online courses (2 predictors), and their goals for the course (2 predictors) and motivations for enrolling (13 predictors). We used penalized regression to identify individual characteristics that were most predictive of each SRL strategy. The advantage of penalized regression in this context is that it performs variable selection. The algorithm shrinks coefficients on predictor variables that provide little or no improvement to model fit, thereby effectively excluding unimportant predictors from the model. When considering individual characteristics that are correlated, such as age and education, the estimated coefficients characterize the predictor's association with an SRL strategy while adjusting for all other predictors in the model. Continuous predictors (age, online courses started/finished, time commitment) were standardized to zero mean and unit variance. All remaining predictors were binary and dummy-coded for the analysis. Scores for the six SRL strategy outcomes were also standardized for ease of interpretation. We applied an elastic net penalty (Zou & Hastie, 2005) in the regression models, which performs variable selection akin to the LASSO penalty (Tibshirani, 1996), but it is less prone to randomly choosing between highly correlated predictors. We used a 90% LASSO with 10% Ridge penalty and 10-fold cross-validation to identify the parameter value that minimized the prediction mean-squared error (cf. Friedman, Hastie, & Tibshirani, 2001). The penalized regression models yielded six sets of coefficients that are illustrated in Figure 3-1.

3.2.3 Results

We begin with general observations about the survey results. Learners reported an average time commitment of 4.9 hours per week (SD = 3.1; median = 4). The vast majority reported an intention to watch all lectures (95%) and complete all assessments (93%) in the course. Half of the learners reported having prior experience with the course topic and a majority had prior experience with online courses (number of prior online courses started: M = 2.4, SD = 4.0, median = 1; number of completed courses: M = 1.8, SD = 3.2, median = 1). The most pronounced SRL strategies reported were self-evaluation and elaboration, followed by strategic planning, task strategies, and goal setting; the least common strategy was help seeking (Table 3-2). Moreover, several of the SRL strategies were highly correlated, such as goal setting with strategic planning ($r = 0.70$), strategic planning with task strategies ($r = 0.66$), and task strategies with elaboration ($r = 0.72$). Help seeking was the least correlated with the overall SRL composite.

Sub-RQ 2.1: Which self-reported SRL strategies are most helpful to achieve personal course goals?

We evaluated how SRL strategies were related to achieving three different personal course goals: first, earning a course certificate, which requires achieving satisfactory grades on course assessments; second, completing assessments (independent of grades), and third, watching lectures in the course. For each personal goal, we assessed the correlation between self-reported SRL strategies and goal attainment among those who expressed the goal. Results from these pairwise correlations, provided in Table 3-3, indicate that goal setting and strategic planning were significant positive predictors of goal attainment for all three goals. In contrast, help seeking was a significant negative predictor of goal attainment (except for completing lectures, $p = 0.069$). Self-evaluation and task strategies were predictive of completing assessments and lectures, while elaboration was not at all correlated with goal attainment.

In light of high correlations between strategies, we proceeded to fit logistic regression models to evaluate all six SRL strategies simultaneously when predicting goal attainment. Goal setting was a strong positive predictor of goal attainment, while help seeking was a strong negative predictor. Results were consistent across personal course goals and robust to regression adjustment for available covariates (demographics, experience, commitment, etc.), and notably, strategic planning was also a strong positive predictor with goal setting excluded from the model. For example, learners who indicated 1 SD higher levels of goal setting had 54% higher odds of achieving their goal of earning a certificate ($z = 2.68$, $p = 0.007$). By contrast, the same model yielded 27% lower odds of certification ($z = -3.11$, $p = 0.002$) for learners who indicated 1 SD higher levels of help seeking. Likewise, coefficient estimates predicting the other course goals were highly significant and only somewhat smaller. Thus, **learners who engaged in goal setting and avoided help seeking were significantly more likely to achieve their personal course goals**. Although several other SRL strategies were individually associated with goal attainment, goal setting and help seeking emerged as the two key predictors.

Table 3-3 Associations between achieving personal course goals and SRL strategies in terms of Spearman correlation coefficients evaluated for binary certification outcome and continuous proportion of assessments/lectures completed in the course.

Personal Course Goal	**Expressed goal (and attained goal)**	**Goal Setting**	**Strategic Planning**	**Self-evaluation**	**Task Strategies**	**Elaboration**	**Help Seeking**
Earn course certificate	32% (8.9%)	0.08^{**}	0.05^{*}	≈ 0	0.04	0.03	-0.05^{*}
Complete all assessments	93% (7.3%)¶	0.05^{**}	0.05^{**}	0.04^{*}	0.04^{*}	0.03	-0.05^{**}
Complete all lectures	95% (9.1%)¶	0.03^{*}	0.04^{**}	0.03^{*}	0.03^{*}	0.03	-0.03

¶ Goal attainment was evaluated for completing over 80% of assessments and lectures, respectively. $^{*}p < 0.05$; $^{**}p < 0.005$.

Sub-RQ 2.2: How do self-reported SRL strategies vary by individual learner characteristics?

We assessed individual differences in self-reported SRL strategies based on 27 individual characteristics, encompassing demographics, prior experience, time commitment, goals and motivations. Figure 3-1 illustrates the results of six penalized regressions, one for each SRL strategy, with coefficient estimates from each model in each column. Blank entries in Figure 3-1 indicate instances where the penalized regression shrunk a coefficient to zero, thereby excluding the corresponding predictor from the model. Estimates are adjusted for all other predictors in the model; for example, the coefficient on age is estimated adjusting for all other characteristics in the model, such as occupation and level of education.

A number of individual differences emerged for learner demographics. Older learners reported consistently higher levels of SRL, except for help seeking. Women reported lower levels of strategic planning, elaboration, and self-evaluation; however, women reported higher levels of goal setting, task strategies, and especially help seeking. Compared to the 37% of learners in the sample who had not earned at least a bachelor's degree, those with a bachelor's degree reported lower strategic planning, self-evaluation, and help seeking. By contrast, learners with a professional or master's degree, and especially those with a Ph.D. reported higher levels of goal setting, strategic planning, and task strategies. While learners with a Ph.D. reported generally strong SRL skills, they reported being much less inclined seek help. Learners who were also students in school or university reported consistently lower SRL, especially for self-evaluation and task strategies. In contrast, learners who were employed were more inclined to engage in goal setting, strategic planning, and help seeking, despite lower levels of self-evaluation. Individual differences by learners' prior experience were more consistent across strategies. Learners who had started more online courses in the past consistently reported lower SRL, while those who had completed more online courses consistently reported higher SRL, especially goal setting. Those with prior experience with the course topic reported higher levels for most SRL strategies but were less inclined to seek help. Furthermore, learners who were willing to commit more time to the course reported consistently higher SRL. Likewise, SRL skills were substantially higher—up to 0.5 SD—among learners who expressed the goal of either finishing all lectures or finishing all assessments.

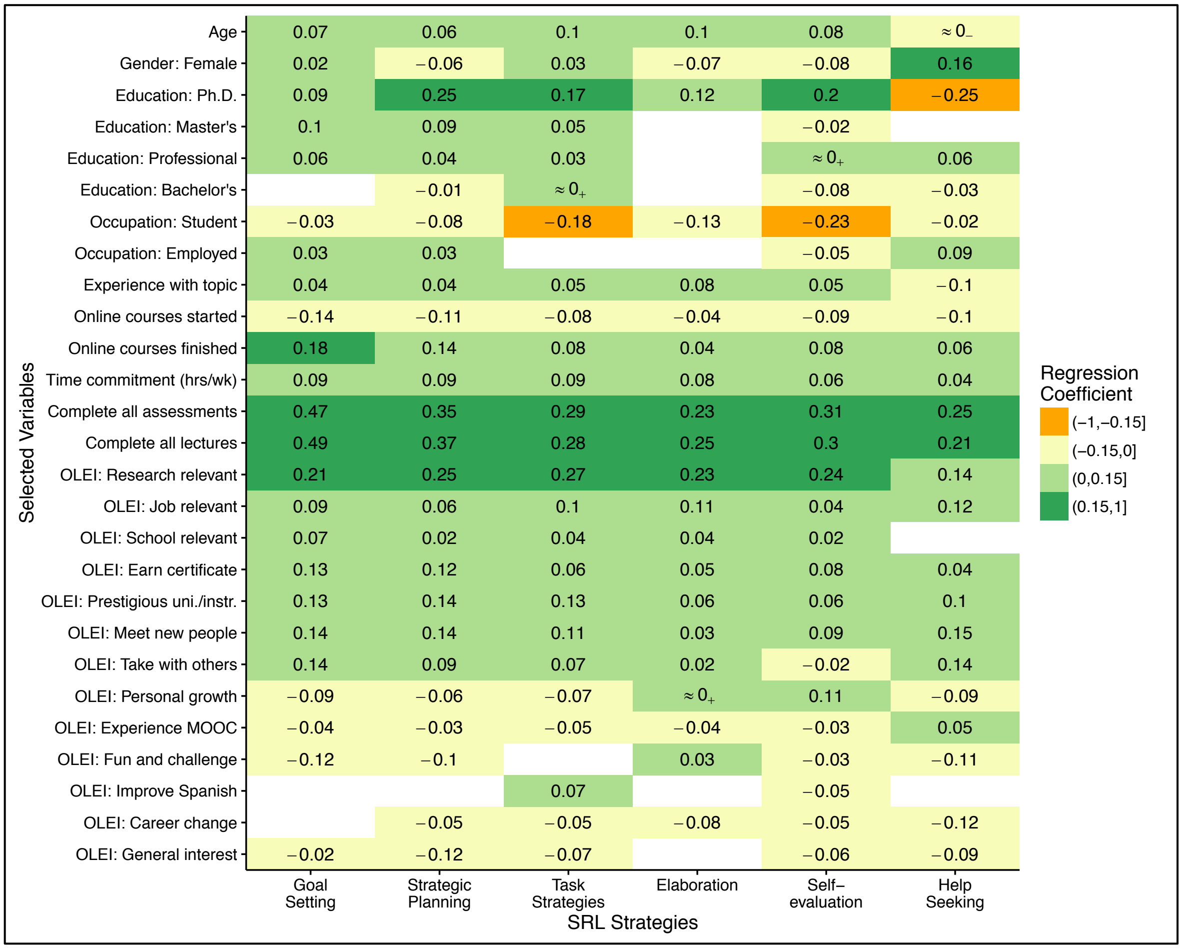


Figure 3-1 Individual differences in SRL examined by demographics, prior experience, time commitment, goals and motivations (marked OLEI). Showing penalized regression coefficients for six models, one for each SRL strategy, with standardized continuous predictors (i.e., age, online courses started/finished, time commitment) and dummy-coded binary predictors (all other predictors). SRL outcome variables were also standardized for ease of interpretation. Blank boxes indicate predictor variables that were excluded by variable selection. Colors indicate the sign and magnitude of coefficients.

3.2.4 Discussion

This study provides a quantitative account of SRL that advances our understanding of which SRL strategies support online learners in MOOCs and how SRL strategies vary across a heterogeneous group of learners. Our results are based on an analysis of survey and platform log data from 4,831 learners across six MOOCs. We briefly summarize the findings pertaining to each of the two sub research questions that we investigated.

First, Sub-RQ 2.1 *which self-reported SRL strategies are most helpful to achieve personal course goals?* Learners who reported engaging more in goal setting and strategic planning were more likely to attain personal course goals, such as earning a certificate. In contrast, help seeking was a negative predictor of goal attainment.

Second, Sub-RQ 2.2 *how do self-reported SRL strategies vary by individual learner characteristics?* A large number of significant individual differences in self-reported SRL were found. Gender, occupation and learners' intentions with the course, commitment and prior knowledge declared by learners indicated stronger SRL skills. Also, motivations for taking the course predicted stronger SRL skills.

The present findings have implications for theory and practice around SRL in the context of MOOCs and similar online learning environments. We discuss three implications of our findings in the context of prior work: (1) supporting goal setting and strategic planning; (2) interpreting the negative results for help seeking, and (3) leveraging insights from individual differences.

First, goal setting and strategic planning stood out as particularly helpful strategies in MOOCs. Learners who reportedly engaged in these metacognitive strategies were more likely to achieve their course goals and engaged more deeply with course assessments, perhaps because they also appreciate the value of assessments for checking their understanding and receiving feedback to support their learning. The results are consistent with accounts from prior work that highlight goal setting and strategic planning as important factors underlying attrition and achievement in MOOCs (Kizilcec & Halawa, 2015; Kizilcec et al., 2016; Nawrot & Doucet, 2014; Zheng et al., 2015).

Second, the finding that help seeking negatively predicts goal attainment can be interpreted several ways. The finding seems surprising considering that prior work has found that learners who report working on the course with someone else, such as a friend, have higher performance (e.g., Breslow et al., 2013; Kizilcec & Schneider, 2015). Although the two constructs were positively related in our analysis of individual differences (see Fig. 3-1, 'OLEI: Take with others'), which suggests some degree of overlap, they differ in several ways. Learners who have coordinated with a 'study buddy' are probably very organized and committed to the course, engage in collaborative learning, and benefit from mutual support and social accountability. Learners who reported being more inclined to seek help were perhaps alone in their educational endeavor and were hoping to enter an active community of learners who support each other during the course.

Third, our findings of individual differences in SRL between learners who expressed different motivations for taking the course provide empirical evidence consistent with recent work. Hood et al. (2015) also found increased self-reported SRL behaviors among learners who were studied or worked in a field related to the course topic compared to those without a topic-relevant role or context. Littlejohn et al. (2016) conducted in-depth interviews with MOOC learners and found consistent evidence for the role of learners' context in shaping their perceptions of their learning process and the purpose of the course.

Note that most instantiations of the SRL strategies considered in this subsection could not be observed directly in the MOOC environment—no data was available about whether learners set clear learning goals, engaged in note-taking while watching lectures, practiced self-explanation, or consulted friends or the Internet for help. Unless SRL strategies are facilitated in the environment or through linked third-party applications, neither self-report nor course log data provides a complete account of online SRL and therefore limits the ability to draw valid conclusions about (the effects of) SRL in these environments.

3.3 Classification of learners based on the relation between their SRL strategies deployed and their performance in MOOCs

In subsections 2.3 and 2.4 of chapter 2, a classification of the learners was presented based on the student's interaction with the contents of the course, which translated into a series of strategies used by the students (behavior patterns) in the MOOC. These strategies were extracted from the data collected by the Coursera and edX platforms, resulting in four groups of students classified based on their actual behavior. These groups were: (1) Sampling learners, (2) Targeting learners, (3) Low Comprehensive learners and (4) Highly Comprehensive learners. To expand the work done, this subsection presents the result of the exploratory study conducted on the differences found in the navigation of students through a MOOC, in which the self-reported data from two questionnaires have been combined, that is the SRL questionnaire and another on Learning Styles (LS). In this case, the study was done in a MOOC deployed on the Open edX platform (i.e., open source version of edX platform). Based on the SRL questionnaire we obtained three types of SRL learners' profiles: low, medium and high. Based on the LS questionnaire we obtained four types of LS learners' profiles: active, pragmatic, theoretical and reflective. The findings suggest that learners with different SRL profiles follow similar navigation paths, but there are differences when differentiating students by their LS, which provide more evidence about the relationship between learners' SRL strategies and their performance in MOOCs.

3.3.1 Related Work

This section provides a review of relevant literature on SRL and LS in online environments and MOOCs.

3.3.1.1 Learning Styles and Self-Regulated Learning

LS is defined as the attitudes and behaviors that characterize a person's way of learning (Honey & Mumford, 1986). They can also be understood as cognitive and affective traits, which serve as indicators of how students perceive, interact and respond to their learning environments (Keefe, 1988). Cognitive traits are linked to the preferences students have to understand and process the information they learn; the affective ones with the motivations and expectations that they have when facing their learning. For Kolb (2005), the LS is the preferential capacities to learn and that are a consequence of hereditary factors, previous experiences, and demands of the current environment in which the individual is inserted. Currently, there is a variety of models that identify different types of LS. Here are some of the most prominent: (1) Dunn & Dunn (1978) created a model that focuses on the perceptual modalities through which students respond in learning tasks: visual style, auditory style, and tactile or kinesthetic style. (2) Felder & Silverman (1988) propose five dimensions to define the LS, these dimensions are linked to the type of information (sensitive/intuitive), preferential stimulation (visual/verbal), the way of organizing the data (inductive/deductive); to process and understand information (sequential/global); to work with information (reflective/active). (3) Myers & Briggs (1977) define LS from four dimensions that describe preferences: orientation to life (introversion/extroversion), perception (sensory/intuition), decision making (rational/emotional) and attitude towards the outside (judgment/perception).

For this research, the classification proposed by Alonso, Gallego, & Honey (1994) has been adopted, who, like Kolb (2005), propose that the best learning is generated when students cyclically pass by four phases. These phases are: 1) Act, 2) Reflect, 3) Theorize and 4) Experiment. And based on these phases, these authors define the following learning styles: Active, Reflective, Theoretical and Pragmatic (Alonso et al., 1994). In a very synthetic way, students who tend to have an active style learn best when they are involved in small activities, require an immediate response or a specific action. Students with a tendency to the theoretical style learn best through models, theories and systems of concepts that allow them to read, interpret and interrogate a reality. Reflective style students prefer to learn through analysis, consideration of different perspectives. Students of pragmatic style prefer activities that link theory with practice, allow them to apply or transfer what they have learned to concrete situations and that are linked to their professional performance.

This model was selected for two reasons. First, this model is based on many previous investigations that validate it. Second, there is a questionnaire (CHAEA) that, based on this model, allows to identify the LS. This questionnaire exists in Spanish and has been validated in previous studies (Alonso et al., 1994).

In this context, it is important that a MOOC can address the differences in the students' SRL capabilities. It is also key to know which LS is predominant among students when designing content and activities in a MOOC. These should have the intention to adjust to their preferences and be facilitators of learning. Both LS and SRL profiles have been studied extensively in the last decade from the aptitude perspective (Muñetón, Pinzón, Alarcón, & Bohérquez, 2012). That is, as a set of skills that students believe they have. One of the techniques most commonly used to identify both SRL and LS profiles are the self-report questionnaires (García, Santizo, & Alonso, 2009). However, there are very few studies that analyze these profiles from the processes. That is, as the set of sequences of activities carried out by a student in the course or online platform. For this reason, this subsection will seek to answer the following sub research question: *Sub-RQ 2.3: How do the different SRL and LS profiles manifest themselves in a MOOC in terms of learning sequences?*

3.3.2 Methods

3.3.2.1 Context: MOOC and Participants

This exploratory study was conducted in the context of a MOOC deployed on the Open edX platform and offered by the University of Cuenca and the Ecuadorian Advanced Internet Consortium (CEDIA). The course was given in Spanish on the subject of Learning Objects. It was launched on March 14th, 2016 and ended on April 10th, 2016. The course was offered openly for all teachers of Ecuadorian universities in the framework of a national contest for the production of digital educational materials. The course is structured in 5 modules. Each module is made up of a set of lessons. Each lesson is composed of a collection of readings, videos, and evaluations. In total the course has 75 texts, 17 videos, and five assessments. All modules were available during the four weeks of the course. The readings correspond to 77.32% of the total objects of the course, the videos 17.52% and the evaluations 5.15%. In the course N = 99 students were registered, of which 21 did not record any activity on the course and 24 students abandoned between the second and third week. For the analysis, all the students were considered except the 21 who did not report any activity, leaving a cohort of 78 students. Of the 78 students who started the first week, only 58 answered the SRL questionnaire and the LS questionnaire. 91% of the participants were between 25 and 55 years old, 65% declared to be male, and 76% had postgraduate training, and 24% had a university education.

3.3.2.2 Instruments

To measure the SRL profile of the learners in the MOOC, an instrument constructed based on four well-established questionnaires in the literature was applied. The questionnaires considered were: (1) professional SRL by Littlejohn et al. (Littlejohn, Hood, Milligan, & Mustain, 2016), (2) MSLQ by Pintrich et al. (Pintrich, Smith, Garcia, & Mckeaechie, 1993), (3) OSLQ by Barnard et al. (Barnard, Lan, To, Paton, & Lai, 2009) and (4) LASSI 2ed by Weinstein et al. (Weinstein & Palmer, 2002). All these instruments have been validated and used in different contexts (Barnard et al., 2008; Hood et al., 2015; Magno, 2011). The final questionnaire contains 66 questions, 13 related to the motivations and intentions of the students based on the OLEI scale of the work of Kizilcec and Schneider (Kizilcec & Schneider, 2015). The other 53 questions are related to 13 SRL strategies. These strategies were evaluated on a scale Likert from 1 to 5 (1 means “Nothing true for me” and 5 means “Very true for me”). The strategies included in the questionnaire are: self-efficacy (6 statements), goal setting (4), strategic planning (4), study environment management (6), rehearsal (4), elaboration (3), organization (4), time management (6), help seeking (4), effort regulation (4), self monitoring (4), self evaluation (2) and self satisfaction (2). The questions were presented in a random order. The reliability of the questionnaire was validated, as shown in Table 3-4. The values obtained for Cronbach's alpha, all higher than .6, so according to (George & Mallery, 2003) the values are acceptable, determining that the questionnaire is well designed.

Table 3-4 Statistics for the reliability of the SRL questionnaire

SRL Strategy	**Chronbach's α**
Self efficacy	0.765
Goal setting	0.897
Strategic planning	0.845
Study environment management	0.845
Rehearsal	0.866
Elaboration	0.748
Organization	0.939
Time management	0.803
Help seeking	0.695
Effort regulation	0.687
Self monitoring	0.782
Self evaluation	0.733
Self satisfaction	0.974

To measure the predominant LS of the participants, the validated CHAEA questionnaire of Honey and Alonso was used. It consists of eighty questions (twenty reference items for each learning style: Active, Reflective, Theoretical and Pragmatic) to which we must respond by stating agreement or disagreement (Alonso et al., 1994).

3.3.2.3 Extracting learning sequences in MOOC

We used the PM methodology proposed in subsection 2.3, where the proposed approach was used to extract learning sequences from MOOC following the next four stages:

Stage 1 – Data extraction: to study the activity sequences of the students in the MOOC, the activity data or log files recorded by the Open edX platform were taken. The platform generates a set of data log files in JSON format. This data is obtained from the "track_trackinglog" table and contains information about the course and the interaction of the participants with the different types of course elements. In addition, the table “auth_user” and “student_anonymoususerid” contain information about the student and the id of each one, which allows them to follow up on their activity in the course (Figure 3-2). The data from these tables were exported and converted from JSON to CSV using the “Convert Jason to CSV” web application. After this, the “CSV” files were used on Open Refine. This software was used to clean the data. This cleaning consisted of eliminating the information that was not considered important for the study, and that was discarded for the following reasons: (1) two or more columns presented repeated information, (2) information of the monitoring environment (Browser, Operating System, IP, etc.) not relevant, and (3) encoded information of the platform only useful for the administration of the platform.

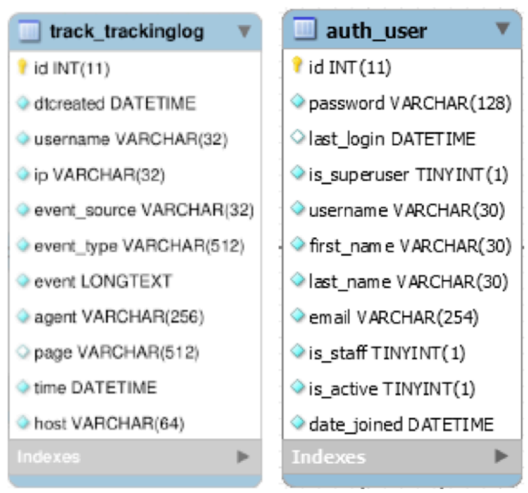


Figure 3-2 Part of the tables extracted from the Open edX database

Stage 2 – Event Log Generation: after the data extraction stage, the event logs are generated from the integration of the data of the activity registered by the students in Open edX, the information collected on the student’s SRL profile (SRL questionnaire), the information collected about the LS of the students (CHAEA questionnaire) and the result of their activity in the course (completed and not completed). Here are how these three elements are integrated. The events of the activity recorded by the students are defined as the set of actions of the interaction of this one with the contents of the MOOC (at the micro level) and with the MOOC lessons (at the macro level).

At the level of interaction with the objects of the course (micro): at this level, events are defined as the result of the student's interactions with the MOOC resources (videos, texts, assessments) in learning sessions (same approach defined in subsection 2.3). The interactions with texts, videos and assessments that are made during this period and are labeled according to their characteristics of completeness, in two types of events: (1) "initiated" - when the interaction with an object begins and is not completed and (2) "completed" - when the interaction with an object that has been initiated in the past is completed (Figure 3-3).

Case_ID	Time Stamp	Actividad	Sesion	Estado Actividad	Nivel SRL	Estilo Aprendizaje
9bd1acf267e0f0365b8527ccf81e28d4	2016-03-14T20:55:41.260914+00:00	Lectura	1	Completado	Medio	Activo
9bd1acf267e0f0365b8527ccf81e28d4	2016-03-14T21:37:47.735182+00:00	Lectura	1	Completado	Medio	Activo
9bd1acf267e0f0365b8527ccf81e28d4	2016-03-14T22:01:57.281524+00:00	Lectura	1	Iniciado	Medio	Activo
9bd1acf267e0f0365b8527ccf81e28d4	2016-04-04T16:18:24.507822+00:00	Evaluación	1	Iniciado	Medio	Activo
9bd1acf267e0f0365b8527ccf81e28d4	2016-03-14T22:16:56.239419+00:00	Evaluación	1	Completado	Medio	Activo
48d98e234099be3d05515d5507203901	2016-03-15T20:11:31.873483+00:00	Evaluación	2	Completado	Alto	Reflexivo
48d98e234099be3d05515d5507203901	2016-03-15T20:14:46.291206+00:00	Video	2	Completado	Alto	Reflexivo
48d98e234099be3d05515d5507203901	2016-03-15T21:45:25.238995+00:00	Evaluación	2	Completado	Alto	Reflexivo

....

Figure 3-3 Fragment of the generated Event Log- Interaction of the student with the objects of the course per session

At the level of interaction with the lessons of the course (macro): at this level, the events are defined as a result of the student's interactions throughout the four weeks with the lessons of the course (Figure 3-4).

Case_ID	Time Stamp	Actividad	Nivel SRL	Estilo Aprendizaje	Estado Curso
9bd1acf267e0f0365b8527ccf81e28d4	2016-03-14T20:55:41.260914+00:00	2.2. Composición Interna	Bajo	Activo	Aprueba
9bd1acf267e0f0365b8527ccf81e28d4	2016-03-14T21:37:47.735182+00:00	2.3. Objetos de aprendizaje	Bajo	Activo	Aprueba
9bd1acf267e0f0365b8527ccf81e28d4	2016-03-14T22:01:57.281524+00:00	2.3. Objetos de aprendizaje	Bajo	Activo	Aprueba
9bd1acf267e0f0365b8527ccf81e28d4	2016-04-04T16:18:24.507822+00:00	1.1. Evaluación (20% Nota Final)	Bajo	Activo	Aprueba
9bd1acf267e0f0365b8527ccf81e28d4	2016-03-14T22:16:56.239419+00:00	1.1. Evaluación (20% Nota Final)	Bajo	Activo	Aprueba
48d98e234099be3d05515d5507203901	2016-03-15T20:11:31.873483+00:00	1.1. Evaluación (20% Nota Final)	Alto	Reflexivo	No Aprueba
48d98e234099be3d05515d5507203901	2016-03-15T20:14:46.291206+00:00	2.7. Actividad Complementaria	Alto	Reflexivo	No Aprueba
48d98e234099be3d05515d5507203901	2016-03-15T21:45:25.238995+00:00	2.8. Evaluación (20% Nota Final)	Alto	Reflexivo	No Aprueba
		....			

Figure 3-4 Fragment of the generated Event Log. Interaction of the student with the lessons of the course in the 4 weeks

Stage 3 and 4 – Discovery and analysis of the model: at this stage, the discovery PM algorithm is applied (as subsection 2.3 recommended). The analysis of the results (graphics and numerical) will allow answering the sub research question *Sub-RQ 2.3: How do the different SRL and LS profiles manifest themselves in a MOOC in terms of learning sequences?* To answer this question, we classified the learners depending their SRL and LS profiles as follows:

First, the students were classified according to their level of SRL. This was calculated from the scores obtained with the questionnaire. Then, three percentiles were calculated, which allowed students to be classified as a) low SRL level if their score was <= 50, b) medium SRL level if their score reached was between 50 and 75, and c) high SRL level if their score reached was >= 75. Second, students were classified according to their learning style as (1) active, (2) reflective, (3) theoretical and (4) pragmatic, based on the use of the CHAEA tool.

Finally, the process models were obtained for each of the LS mentioned and SRL profiles at the level of interaction with the objects of the course and the level of interactions with the lessons of the course.

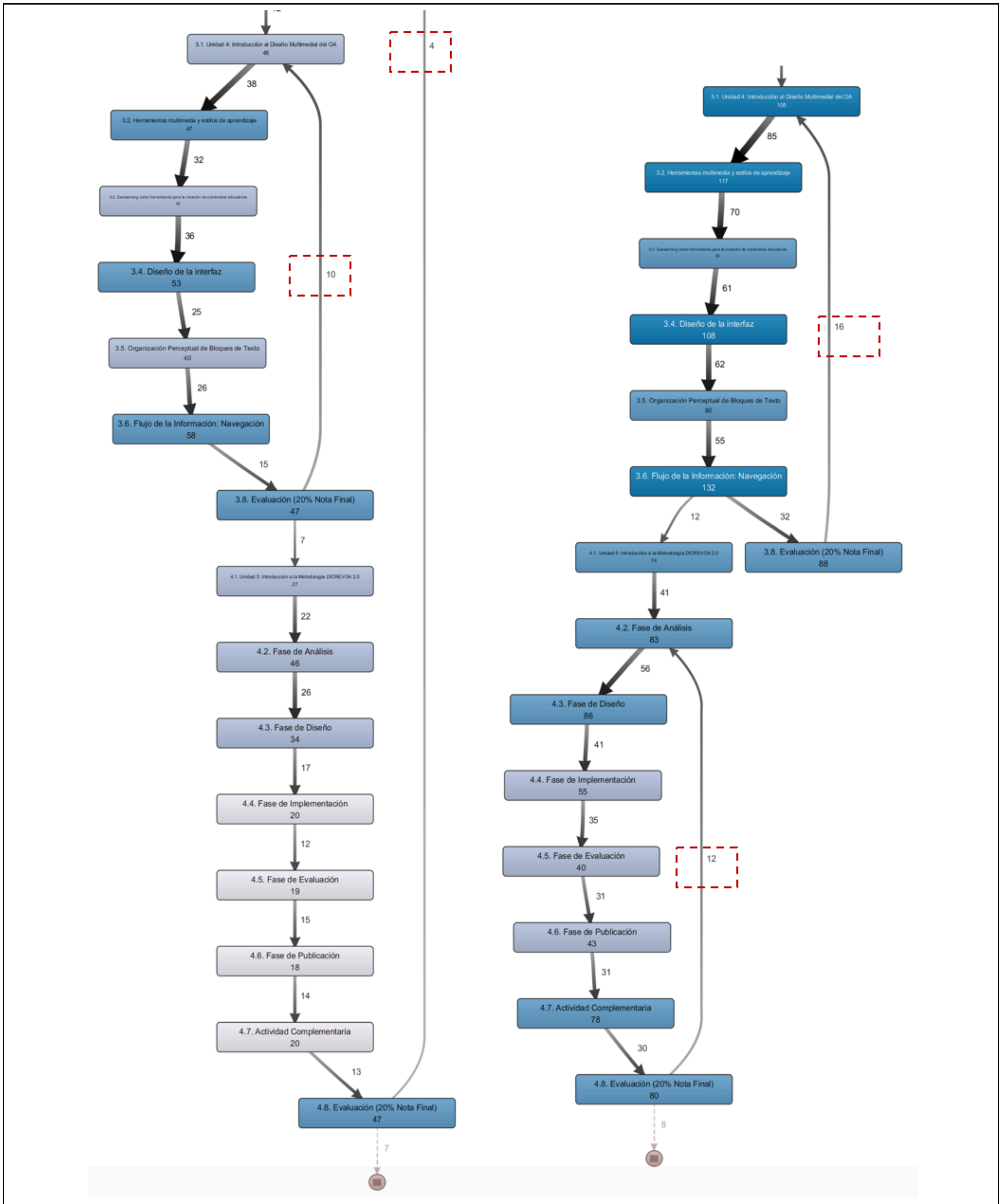
3.3.3 Results

In this section, we present the results obtained from the analysis of the event logs. The results have been organized according to the sub research question below:

Sub-RQ 2.3: How do the different SRL and LS profiles manifest themselves in a MOOC in terms of learning sequences?

Result 1. Students in a MOOC with an average level of SRL tend to return to the beginning of each module after completing an assessment while students with a high level of SRL do not repeat this behavior until the end of the course. Figure 3-5 shows the comparison of the process models obtained from their interaction with the lessons of the course: (a) high SRL level, (b) medium SRL level. Each rectangular box in the process models represents an activity that, in this case, are the lessons that each module has in the MOOC. In this way of the process model (a) – high SRL, it can be seen that there are 10 transitions and, in the model, (b) – medium SRL 16 transitions that return from the evaluation of module 3 to the beginning of this module (red framed). In the model (a) there are 4 transitions from the end of the course to the beginning of this, this is because a reminder was sent using the Open edX platform to the participants in the last week to remind them to complete the questionnaire of EA. Finally, for the model (b), 12 transitions are observed from the evaluation of module 4 to the beginning of this module (red frame).

Result 2. In a simple work session, students in a MOOC with a high SRL level tend to complete a greater number of consecutive readings. If one compares the models generated in Figure 3-6 (a) and 3-6 (b), it can be seen that the transition from LECTURA\\completa→LECTURA\\completa repeats 1,964 times for students with high SRL. On the other hand, for students with a medium level of SRL (Figure 3-6 (b)) the transition LECTURA\\completa→LECTURA\\completa repeats 780 times. Based on the number of repetitions of consecutive sequences of LECTURA\\completa, it can be inferred (based on the number of activities and transitions between activities) that students with a high SRL level would work more intensively.



(a) (b)

Figure 3-5 Sequence of activities of modules 3 and 4 of the MOOC, for students with (a) high SRL level and (b) medium SRL level

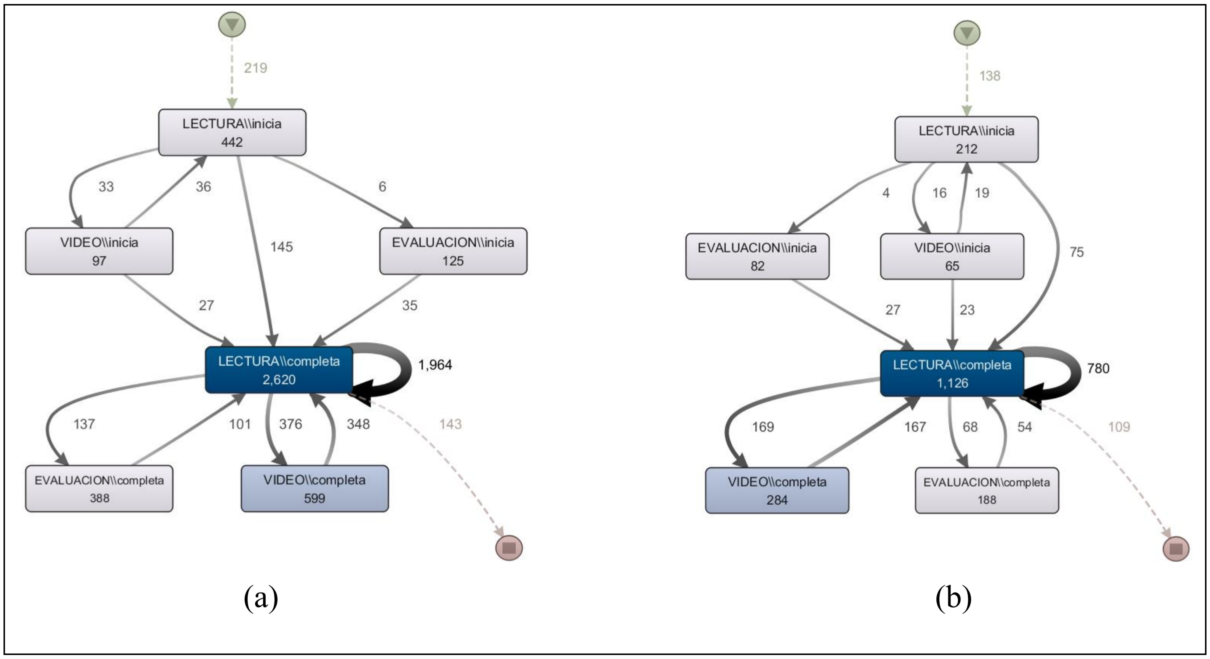


Figure 3-6 Process Model for students with (a) high SRL level and (b) medium SRL level

Result 3. In a study session, students in a MOOC with theoretical, reflective and pragmatic LS tend to initiate an evaluation and then complete it without performing intermediate readings. If we observe the process models generated in Figure 3-7 (a), (b) and (c) that correspond to the theoretical, reflective and pragmatic LS respectively, we observe that there is a complete transition between the activities EVALUACION\\inicia → EVALUACION\\completa (framed in red). In the case of students with theoretical LS, 19 transitions are made, for the case of the reflective 25 transitions, and for the case of pragmatists, 14 transitions are made. Unlike students with active LS, these do not show the sequence of activities that the other LS mentioned above. From the process model of Figure 3-7 (a) it is observed that the transition between the activities EVALUACION\\inicia→EVALUACION\\completa, is mediated by LECTURA\\completa.

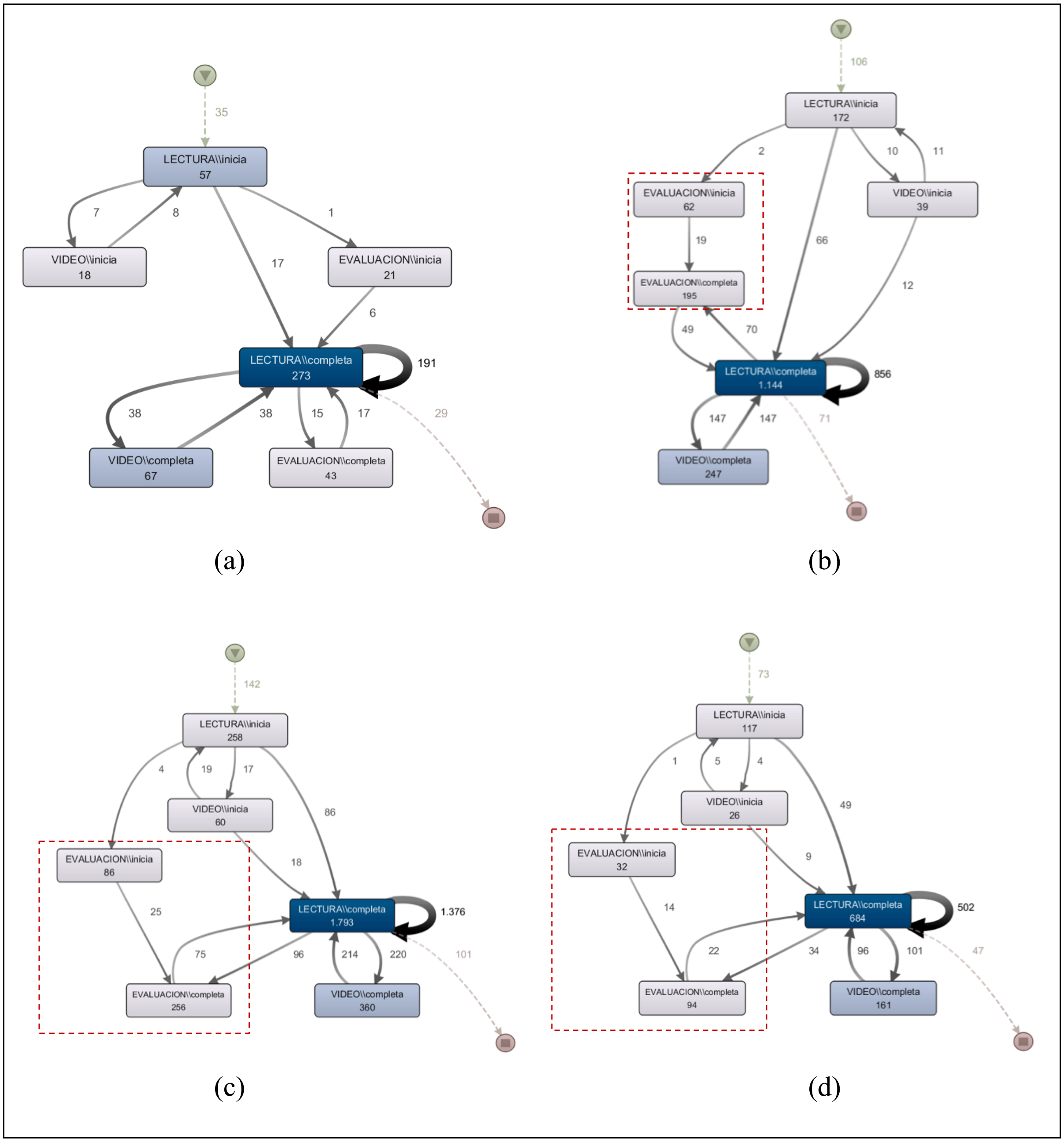


Figure 3-7 Process Models obtained for students with (a) active LS, (b) theoretical LS, (c) reflexive LS and (d) pragmatic LS

Result 4. Students in a MOOC with theoretical and pragmatic LS tend to follow the linear sequence proposed by the course. From Figure 3-8 (a section of the entire model is presented to exemplify the set of sequences of activities obtained), the processes models generated, both the (a) - students with pragmatic LS and (b) - students with Theoretical LS, they show the sequence of activities carried out by the students, from when the MOOC starts until it ends. It can be seen that both groups traverse each of the lessons linearly in the order in which they were structured for the MOOC.

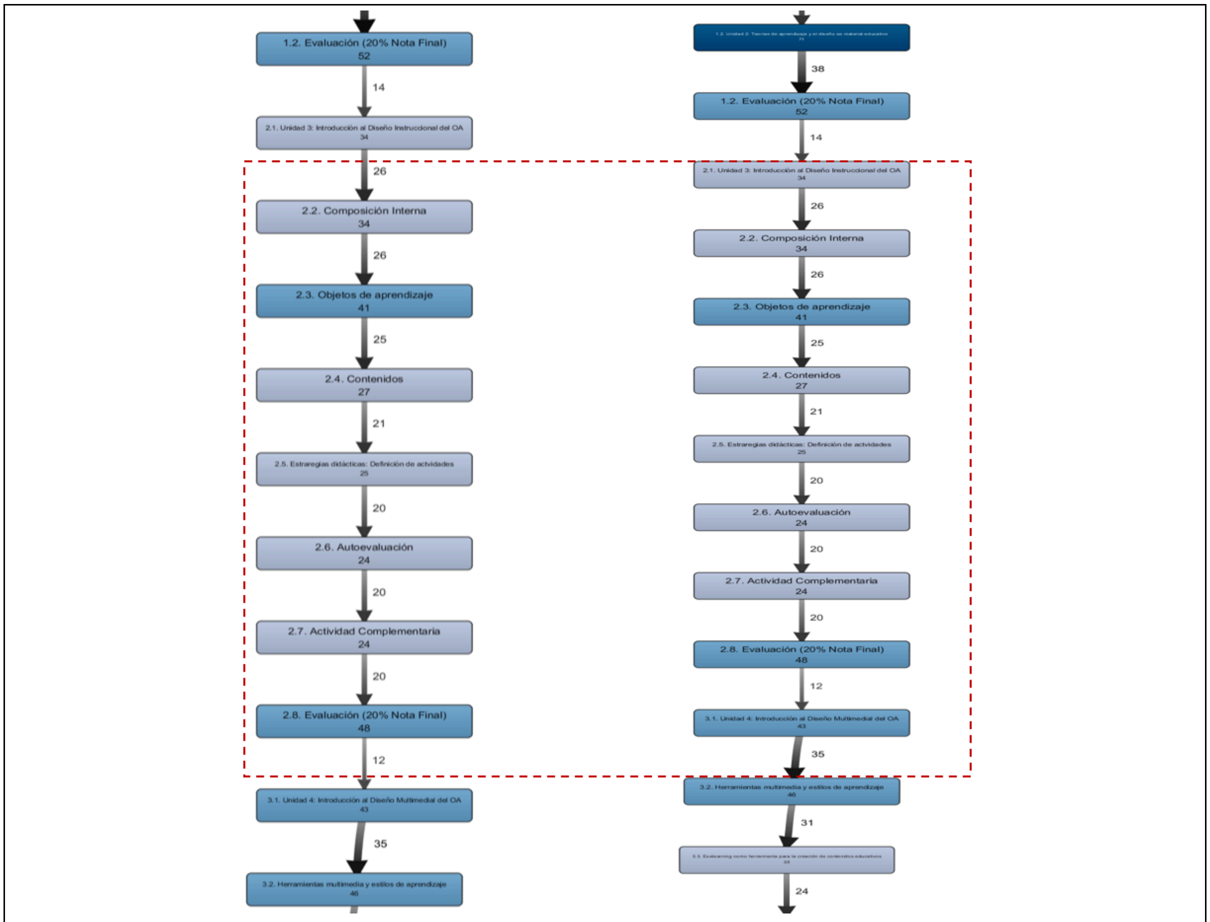


Figure 3-8 Process Models for students with (a) pragmatic LS, (b) Theoretical LS. The area framed in red represents a section of the whole model to exemplify the set of sequences obtained

Result 5. Students in a MOOC with active LS do not follow the linear sequence proposed by the course structure. From Figure 3-9 we can observe the most important part of the generated process model, where students with active LS tend to jump backward (backtrack in the lessons) and go through the course in a non-sequential way. It can be observed in the marked red area, that after making the first reading of the first lesson of the first module (rectangular box of intense blue color framed), they try to solve the initial assessment (rectangular box of less intense blue color and successive to the aforementioned). For this, they carry out a series of sequences of consecutive activities between lesson 1 and evaluation 1 (14 transitions). From these transitions, we can infer that the students were returning to review the lesson to answer the evaluation questions.

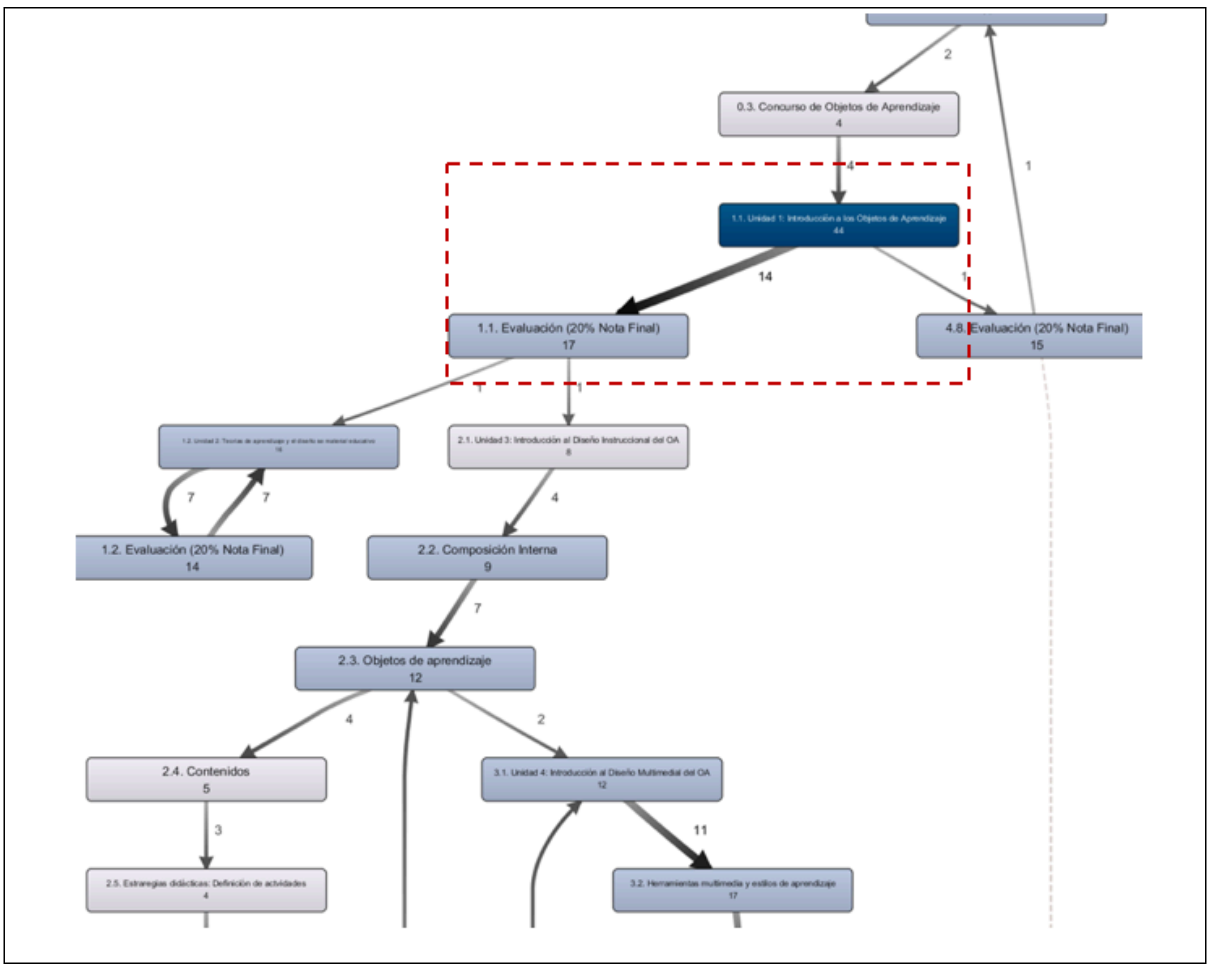


Figure 3-9 Process model for students with active LS

Result 6. Students in a MOOC with reflexive LS tend to follow the linear sequence proposed by the course, but also return to the beginning of each module after completing an evaluation. This was observed for both modules 2, 3 and 4 of the MOOC. Figure 3-10 shows part of the generated process model where the sections framed in red show the sequence of activities mentioned above.

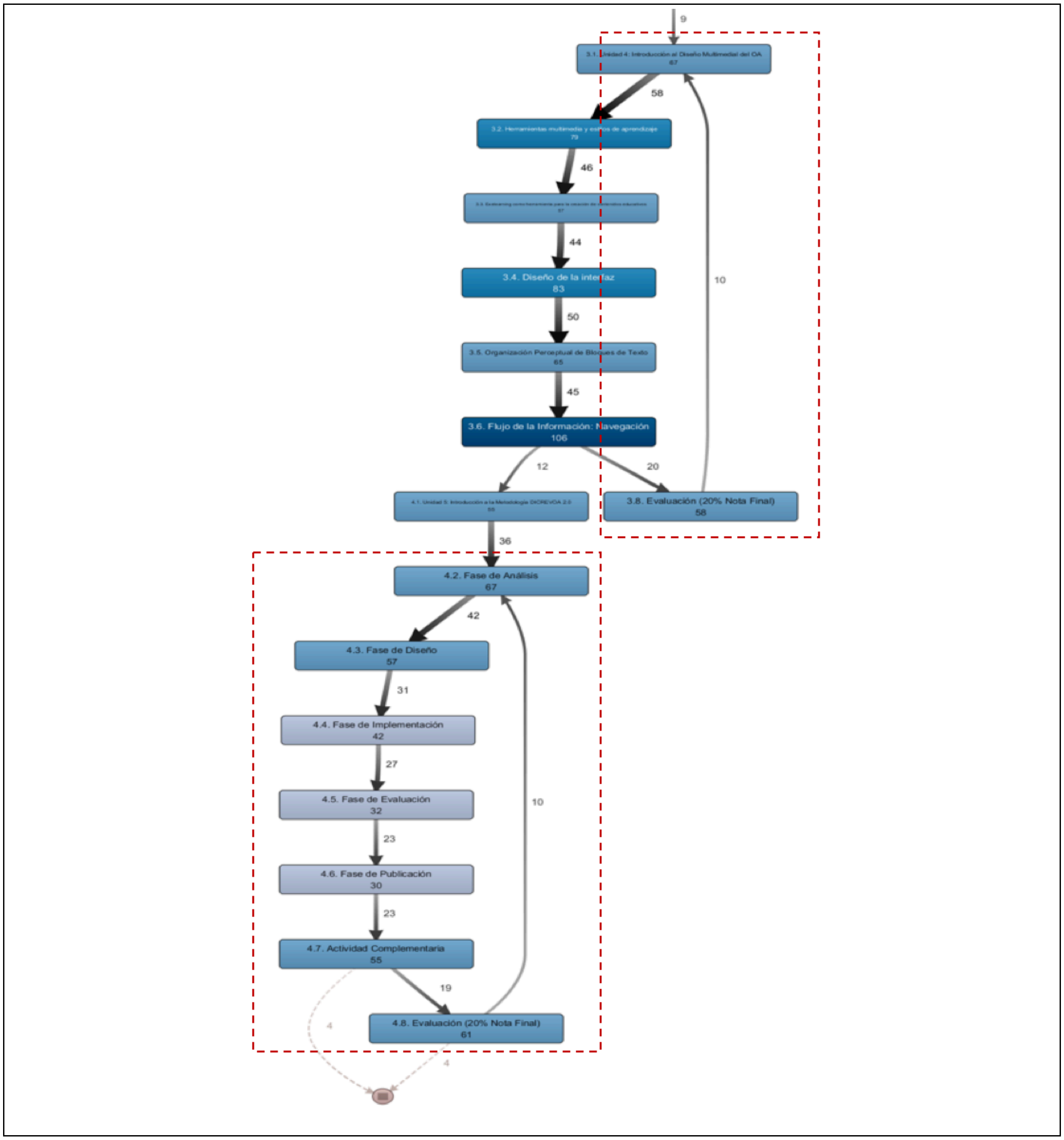


Figure 3-10 PM of modules 2, 3 and 4 of the MOOC for students with reflexive LS

Result 7. Students in a MOOC who complete the course tend to sequentially follow the structure of the course. The process model presented in Figure 3-11 shows the comparison of the sequence of activities carried out by the students who complete the course (a) about those who do not complete the course (b). The section framed in red shows the difference between the transitions made by each group of students.

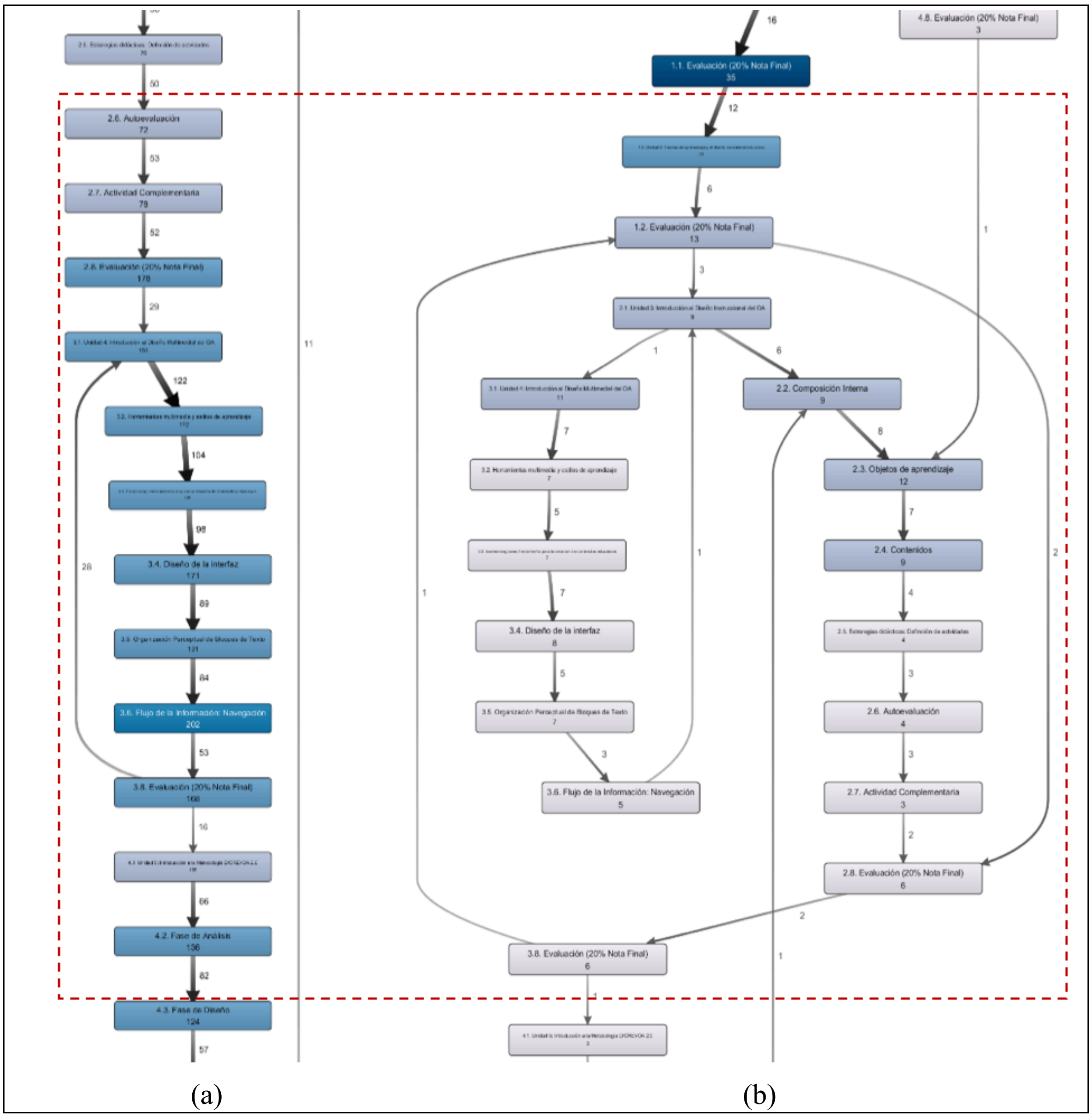


Figure 3-11 Process model for students who (a) complete the course and (b) who do not complete the course

3.3.4 Discussion

Although there is previous work in the literature where the trajectories carried out in a MOOC by students have been explored (Guo & Reinecke, 2014, Mukala, Buijs, & van der Aalst, 2015, Sonnenberg & Bannert, 2015), this is one of the first studies that analyze these trajectories and their relation with the SRL and the LS. For this, the processes (activities) carried out by the students in the MOOC have been studied, contributing, unlike other works (e.g., Alharbi, Paul, Henskens, & Hannaford, 2011), new information. This information allows us to understand how the self-reported characteristics in the questionnaires are manifested in concrete actions within the platform. This complex analysis has allowed advancing in the understanding of the SRL and LS in a MOOC. The results obtained support previous works related to SRL in online environments (Bannert, 2009, Roth et al., 2015, Wirth & Leutner, 2008) and associated with SRL in a MOOC (Beheshitha, Gašević, & Hatala, 2015; Hood, Littlejohn, & Milligan, 2015; Kizilcec, Pérez-Sanagustín, & Maldonado, 2016). In this exploratory study, the data of the self-reported questionnaires by the students (SRL and LS) were combined with the data they left on the Open edX platform (data log). The result of the interaction of students at the micro level (with course objects) and macro level (with the lessons of the course) was derived from the use of PM techniques. The visualization of these models helps to understand how students navigate through the MOOC, based on the particular characteristics of SRL and LS of these.

In the context of this research, it has been shown that both students who complete the course and those who do not tend to follow sequences of similar activities, which is to follow the sequential structure proposed by the course, as shown by similar studies (Guo & Reinecke, 2014; Mukala, Buys, Leemans, & van der Aalst, 2015). These results also reveal that students with a high SRL level work more persistently on the objects in the course compared to those with a medium SRL level. It is also important to mention that many of the activities that are designed for a MOOC are limited by the technological platforms on which the course is implemented.

3.4 SRL strategies that predict learners' success in MOOCs

Most of the learners who enroll in a MOOC decide which parts of the course content they choose to engage with, and eventually only a small proportion of these enrollees complete the course (typically less than the 10%) (Chuang & Ho, 2016). This has aroused the interest on studying the causes why learners complete or drop out a MOOC. Prior research shows that self-regulation is one of the critical skills needed to achieve personal learning goals in a MOOC (Maldonado-Mahauad et al., 2018). Moreover, recent research in self-regulated Learning (SRL) suggests that successful learning and academic achievement are associated with the deployment of regulatory activities such as goal-setting, planning or monitoring (Bannert, 2009).

MOOC enrollees present a diversity of behaviors depending on: learner's previous knowledge, prior experience, intentions and motivations (Littlejohn, Hood, Milligan, & Mustain, 2016; Reich, 2015). In a MOOC platform, this behavior is recorded as the interactions of the learners with the course content, generating a great deal of information that offers an opportunity for identifying patterns and predict trends (Grainger, 2013). Actually, using all these data to run predictions about learner's success in a MOOC is of special relevance. Understanding enrollees' learning behavior can help to detect learners who "probably" will not pass the course (Zhao, Yang, Liang, & Li, 2016). Moreover, this analysis could be used to better understand how learners work in the course and what kind of support he/she may need, anticipating problems which may lead to learners' dropouts.

Several studies have tried to predict attrition, retention and completion in MOOCs. Most of these studies have been carried out in cohort MOOC settings (e.g., instructor based), where time is typically structured, learners follow a fixed schedule, and course materials are released at specific times. However, in self-paced MOOCs, this prediction models may be more critical. On the one hand, the success in self-paced courses, without the support of an instructor, depends on the ability of enrollees to be able to self-regulate their behavior (Maldonado, Palta, et al., 2016). On the other hand, learners' behavior could be more variable, since students do not follow a strict schedule, all materials are released when the course starts, and dates for assessments are flexible (Kocdar, Karadeniz, Bozkurt, & Buyuk, 2018). As a consequence, to detect and predict trends in self-pace MOOCs is still a challenge that have been addressed in prior works with different approaches. For example, authors in (Xu & Yang, 2016) developed a grade predictive method that uses learner activity features to forecast whether or not a learner may get a certificate. Authors in (Brinton, Buccapatnam, Chiang, & Poor, 2016) developed a predicting model to understand when learners will answer a question correctly. In (Sinha, Jermann, Li, & Dillenbourg, 2014), authors analyzed the relationship between interactions and the number of days in which learners interact with the content.

Despite of the predictive power of the models proposed, these models raised some discussions in the community. On the one hand, some researchers argue that frequency and events count are not the best metrics to obtain practical indicators to explain individual differences in online learning (You, 2016). On the other hand, existing models are based on the use of low-level indicators of learners' interaction with the course, but this makes it difficult to obtain meaningful patterns of more complex behaviors, such the use of SRL strategies (You, 2016). Therefore, there is an opportunity to improve these predictive models by considering both, data informing about the heterogeneity of learners (e.g., self-reported data about learning strategies) and more complex behaviors represented by activity sequences instead of individual events.

As a first proposal in this line, in this subsection, we present an exploratory study that uses SRL behavioral patterns related with learners' success as coarse-grained data to predict their behavior in a self-paced MOOC. Specifically, we investigate whether or not learners pass the course based on these patterns together with demographic variables, SRL self-reported strategies and learners' intentions. As a result, we identified new factors to improve predictive models of learners' success in self-paced MOOCs.

3.4.1 Related Work

This section provides a review of relevant literature in prediction in MOOCs and SRL. Then, one sub research question is proposed that is addressed empirically.

3.4.1.1 Prediction in MOOCs and Self-regulated learning

MOOCs have special features that differentiate them from other online courses. First, the big amount of global data that can be collected about learners' activity with the course content. Second, the variety of this data, in which we can identify heterogeneous profiles in terms of personality, learning preferences, education, etc. And third, the number of the interactions related to intensive use of video-lectures and assessments, less frequent in traditional online courses (Moreno-Marcos, Muñoz-Merino, Alario-Hoyos, Estévez-Ayres, & Delgado-Kloos, 2018). All these data have been used to discover predictive patterns of persistence or attrition through MOOC success and completion. Specifically, the data sources used in previous work is usually: (1) learners' demographic data, (2) learners' self-reports data (as intentions regarding the course), (3) clickstream data, (4) forums and social media data and (5) other clickstream traces (Kizilcec et al., 2017).

In the past years, recent studies started considering not only learners' demographic data for predicting behavior, but also self-reported data related with more complex students' learning strategies. For example, studies (Broadbent & Poon, 2015; Broadbent, 2017) found positive relationship between learners' self-reported SRL strategies and academic achievement. According to these studies, the use of SRL strategies affects the learning outcomes achieved and is typically associated with better academic performance in both traditional and online learning situations. In study (Davis, Chen, Hauff, & Houben, 2018) authors found 15 learning strategies were correlate with learners' academic performance (final grades) in online environments, and 5 were found to predict learners' grades. In another example with 50,000 learners (Hood et al., 2015), authors found significant differences in the scores obtained by learners who were already familiar or working in fields related with the MOOC content, with higher self-efficacy, than their counterparts. In another study with 4,831 learners (Kizilcec et al., 2017), authors found that goal setting and strategic planning predicted attainment of personal course goals. Further, in (Corrin et al., 2017) in a study with 2,439 learners, authors found that having a particular help seeking strategy predicts better performance in the course. Regarding clickstream, data with video-lectures, assessments and forums have been used in predictive models. For example, studies (Kizilcec et al., 2017; Maldonado-Mahauad et al., 2018) use video-lectures actions related to pause, play, stop video, watch, complete or review as a method for measuring learners' engagement the course content. Results of these studies showed that the amount of video-lectures intended and completed are predictors of course completion and showed that it is not necessary for learners to watch video-lectures from the beginning to the end to demonstrate its predictive effect (Sinha et al., 2014). In relation to assessment, different types of clickstream such as trying or completing an assessment, have been found to be predictors of course completion (Maldonado-Mahauad et al., 2018). Researchers in (Barba, Kennedy, & Ainley, 2016), for example, found that the number of assessments' attempts is predictor of course completion; even more, those who try the first assessment were 30% less likely to drop out the course. Regarding the activity recorded in forums, the study (Sinha et al., 2014) found that the number of forum pages viewed, or activities within the forum, such as voting up or down, were found as predictors of MOOC completion and persistence. Finally, some others clickstream traces have been found as predictors to MOOC persistence and completion, such as the number of active days that learners spent in a MOOC and the learners' pace through the contents (Moreno-Marcos et al., 2018).

Despite of their demonstrated predictive power, these models have some limitations. On the one hand, the use of these data sources as indicators for predict success in a MOOC are not always the more adequate. Learners' self-reported data captures only the intentions of the learners regarding the course, but not their actual behavior. Since SRL is a continuous process rather than a single picture in time, considering indicators that come from the learners' activity within the course could be a better potential indicator. On the other hand, frequency counts of events from clickstream data and other clickstream traces that are obtained directly from low-level data are limited for detecting learners more complex behavior in a MOOC for suggesting learning guidance. Moreover, as other studies already demonstrated, clickstream data in isolation do not necessarily build better predictive models (Zhao et al., 2016). Therefore, predictive models could be improved by adding variables built on longer activity sequences resulting from learners' interaction with the course content. That is, to propose new indicators that represent how learners adhere to the designed paths of the course, such as activity sequences extracted from coarse-grained data. This idea is built upon previous studies, which investigated the relationships between interaction sequences and learning outcomes using methods such as transition graphs, process mining, sequential pattern analysis, and Markov models (Kizilcec et al., 2017; Maldonado-Mahauad et al., 2018; Pardo et al., 2016). Therefore, and based on prior work, this subsection tackles the following sub research question: *Sub-RQ 2-4: Which indicators of SRL obtained from self-reported questionnaires and activity sequence extracted from trace data can predict course success in self-paced MOOCs?*

3.4.2 Methods

3.4.2.1 Context: Sample and MOOC

This study uses data from one MOOC on Electronics offered by Pontifical University Catholica of Chile in Coursera. The course was taught in Spanish and the materials were organized in four modules. In total the course included 17 lessons, 83 video-lectures and 16 summative assessments. The course followed a self-paced delivery mode in which course materials were available all at once, and without specific predefined deadlines. Data collection occurred between April and December of 2015.

A total of 25,706 learners registered for the MOOC, but the study sample is N = 2,035 which corresponds with those learners who answered a self-reported SRL questionnaire that was introduced at the beginning of the course to define SRL learners' profile. Learners' average age was 30.7 years (SD = 11.06); and the 11% were women.

3.4.2.2 Measures

The instrument used to define learners' SRL profile was already validated in previous studies (see subsection 3.3). It contains 35 questions about learners' intentions with the MOOC content (e.g., hours expected to be dedicated to the MOOC, interest in the topic, etc.), demography (e.g., age, gender, employment status, etc.) and a measure of SRL (Kizilcec et al., 2017). The SRL measure consisted of 24 statements related to six SRL strategies: goal-setting strategies (4 statements), strategic planning (4), self-evaluation (3), task strategies (6), elaboration (3) and help seeking (4). Learners rated statements using a 5-point scale (coded from 0 to 4), where a total average of 4 means a high SRL profile. The SRL measure exhibited high reliability for all strategy subscales with Cronbach's alpha of at least .70. For this study, we also defined success in a self-paced MOOC based on the grades that learners achieve in the course. Therefore, success learners include any enrollee who meets one of the following two conditions:

1. obtains at least the minimum score to pass the course (80%) independently if he/she tackle most of the course materials (most common form of success),
2. obtains at least the minimum score to pass the course attempting at least 50% of the videos in the course materials

3.4.2.3 Extracting learning sequences

In order to extract sequence patterns from a self-paced MOOC, we used the PM methodoly defined in subsection 2.3 and was structured into four stages. As a result we obtained the same six learning sequences as in subsection 2.3: (1) *only video-lectures*, (2) *only assessment*, (3) *explore*, (4) *assessment-try to video-lecture*, (5) *video-lecture-complete to assessment-try*, and (6) *video-lecture to assessment-complete*. Also, three clusters that classify learners according to their interaction sequences patterns and SRL profile were obtained. These clusters are: Sampling learners (cluster 1), Comprehensive learners (cluster 2) and Targeting learners (cluster 3).

3.4.2.4 Applying Predictive Models

Once we identified the learning sequences, we combined these with self-reported SRL strategies, other traditional self-reported variables such as demographics, intentions, and variables that result from the activity of the learner within the platform, in order to identify which of these variables (fine and coarse grained) are predictors of learners' success in self-paced MOOCs. In order to assess whether the variables in Table 3-5 had statistically significant and independent effects for predicting learners' success, we conducted multiple linear regression analyses and logistic regression analysis. Variables used in the predictive model were selected by means of a stepwise regression, using the 23 predictors. Stepwise regression uses an algorithm to select the best grouping of predictor variables that account for the most variance in the outcome (R2); this technique is useful in exploratory studies or when testing for associations. All the predictors are continuous except for gender, employment status, interest in topic, interest in assessment and prior experience, which are dummy-coded binary predictors. Finally, with the self-reported data on SRL strategies as well as the patterns extracted, demographic data about learners, intentions towards the course and activity registered in the course, we built a dataset containing 23 variables that were considered as possible predictors of success. These predictors are presented in Table 3-5.

Table 3-5 Predictors classified by categories

Category	**Predictors**
SRL Strategies	(1a) Goal setting (1b) Strategic planning (1c) Self-evaluation (1d) Task strategies (1e) Elaboration (1f) Help-seeking
Sequence patterns	(2a) Only video-lectures (2b) Only Assessment (2c) Explore (2d) Assessment-try to video-lecture (2e) Video-lecture-complete to assessment-try (2f) Video-lecture to assessment-complete
Demographics	(3a) Age (3b) Gender (3c) Employment status (student) (3d) Employment status (job)

Table 3-5 Predictors classified by categories

Category	Predictors
Intentions	(4a) Time commitment (4b) Interest in topic (4c) Interest in assessment (4d) Prior experience
Activity	(5a) Active days (5b) Time spent (minutes) (5c) Number of sessions

3.4.3 Results

Sub-RQ 2-4: Which indicators of SRL obtained from self-reported questionnaires and activity sequence extracted from trace data can predict course success in self-paced MOOCs?

We assessed individual differences between three groups: (1) Comprehensive learners as a group (cluster 2), (2) Targeting learners as a group (cluster 3) and (3) all learners as one group (cluster 1, cluster 2 and cluster 3). For this assessment, we used 23 individual characteristics, encompassing SRL strategies, sequence patterns extracted from the behavior of the learner with the course content, demographics, intentions and activity with the course resources. Figure 3-12 illustrates the results of the regressions, one for each group, with estimated standardized coefficients (sign and magnitude) from each model in each column. Blank entries in Figure 3-12 indicate that the corresponding predictor was excluded from the model. These standardized coefficients were obtained after running multiple linear regression and logistic regression. For each group, we have considered grades as a dependent variable. For multiple linear regression, the grades were considered as a continuous variable. For logistic regression, the grades were considered as a binary variable (grade >= 80; grade >= 80 & proportions of video-lectures >= 50%). A number of individual differences emerged for learners who succeed in a MOOC across different set of indicators and depending on the group in which they were classified. For comprehensive learners, the *strategic planning* strategy was associated with success in the course, while *elaboration and help seeking* were the strategies associated with success for targeting learners (grade >= 80; grade >= 80 & proportions of video-lectures >= 50%).

Comprehensive learners who performed the sequence patterns *only assessment, explore, and assessment try to video-lecture* while they were facing the course, were more successful (grade >= 80; grade >= 80 & proportions of video-lectures >= 50%). Targeting learners who performed the sequence patterns *only assessment and assessment try to video-lecture* were more successful (grade >= 80), while for the same group the strategy *assessment try to video-lecture* was associated only with success (proportions of video-lectures >= 50%) if learners passed the course and attempted, at least, 50% of video-lectures. Regarding activity indicators, comprehensive learners who spent more *active days and time* in the MOOC were more successful, while targeting learners only *time spent* was associated with success. To predict the final grade (as continuous), we run a stepwise method. As a result, we obtained 3 models for (1) Comprehensive learners as a group, (2) Targeting learners as a group, and (3) all learners as one group. Table 3-6 describes the regression models obtained for each group.

Table 3-6 Summary of the models using multiple linear regressions for the three groups (grade continuous)

Group	**R^2**	***adj.* R^2**	**df**	**F**	***p***
(1) Comprehensive	0.8296	0.8039	73	32.31	<0.001
(2) Targeting	0.7249	0.7175	408	97.73	<0.001
(3) All	0.8559	0.8552	2026	1202	<0.001

For group (1) Comprehensive learners, the self-reported variable *goal setting*, the sequences patterns *only assessment*, *explore* and *assessment try to video-lecture*, the reported demographics as *young learners*, be *women* and *employment status as student*, the learners' *prior experience* and *interest in assessment* reported, the *active days* and the *time spent* were significant predictors of the final grade. These variables explained 80.39% of the variance in the final grade (R^2 = .8039, F =32.31, p < .001). For group (2) Targeting learners the self-reported variables *strategic planning*, *elaboration* and *help seeking*, the sequences patterns *only assessment*, *video-lecture complete to assessment try*, *explore* and *assessment try to video-lecture*, the reported demographics as young learners, the learners' *prior experience*, the *time spent,* and the *number of sessions* were significant predictors of the final grade. These variables explained 72.49% of the variance in the final grade (R^2 =

.7249, F =97.73, p < .001). For group (3) “All learners as one group”, the self-reported variables *elaboration,* and *help seeking,* the sequences patterns *only assessment*, *video-lecture complete to assessment try*, *explore,* and *assessment try to video-lecture*, and the learners’ *prior experience* reported, the *active days* and the *time spent* were significant predictors of the final grade. These variables explained 85.5% of the variance in the final grade (R^2 = .855, F =1,202, p < .001).

		Grade (continuous)			Grade >= 80 (binary)			Grade >= 80 & prop_lectures >= 0.5		
		Comp	Targ	All	Comp	Targ	All	Comp	Targ	All
SRL Strategies	(1a) Goal_setting	-0.09			-0.28			-0.28		
	(1b) Strategic_planning		-0.05		0.18			0.18		
	(1c) Self_evaluation									
	(1d) Task_strategies									
	(1e) Elaboration		0.07	0.03		0.10	0.04		0.08	0.02
	(1f) Help_seeking		0.12	0.04		0.13	0.03		0.13	0.03
Sequence patterns	(2a) Only Vlecture									0.04
	(2b) Only Assessment	0.17	0.27	0.23	0.18	0.22	0.20	0.18	0.14	0.14
	(2c) Explore	0.27	0.21	0.20	0.24	0.12	0.16	0.24	0.13	0.16
	(2d) Atry_to_Vlecture	0.41	0.33	0.36	0.40	0.28	0.29	0.40	0.26	0.29
	(2e) Vlcomplete_to_Atry		-0.16	-0.05		-0.11	-0.03		-0.07	-0.04
	(2f) VLecture_to_Acomplete					0.06				
Demographics	(3a) Age	-0.13	-0.05							
	(3b) Gender: Female	0.08								
	(3c) Employment status: student	-0.16								
	(3d) Employment status: job									
Intentions	(4a) Time commitment				-0.14		-0.02	-0.14		
	(4b) Interest in topic									
	(4c) Interest in assessment	0.10								
	(4d) Prior experience	0.15	0.06	0.05	0.10			0.10		
Activity	(5a) Active days	0.18		0.07	0.29	-0.18	-0.23	0.29	-0.14	-0.21
	(5b) Time spent (minutes)	0.21	0.48	0.31	0.32	0.47	0.50	0.32	0.50	0.49
	(5c) Number of sessions		-0.10		-0.31			-0.31		

Regression Coefficient
(-1, -0.15]
(-0.15, 0]
(0, 0.15]
(0.15, 1]

Figure 3-12 Individual differences between 3 groups of learners (comprehensive, targeting, all) considering the grade as a continuous and binary variable (grade >= 80; grade >= 80 & proportions of video-lectures >= 50%), examined by SRL strategies, sequence patterns, demographics, intentions and activity. Blank boxes indicate predictor variables that were excluded by variable selection. Colors indicate the sign and magnitude of standardized coefficients. All regression coefficients are significant (p <.001).

The sequence patterns *only assessment*, *explore* and *assessment try to video-lecture,* and the *time spent* were significant positive predictor for the three groups. The magnitude of the standardized coefficient for the predictor *assessment try to video-lecture* for group "Comprehensive" and "All", and the magnitude of the standardized coefficient for the predictor *time spent* for "Targeting" were the highest. It is also worth noting that *video-lecture complete to assessment try* and *employment status as student* were significant negative predictors for "Targeting" and "Comprehensive" respectively. Finally, an evaluation of the models was performed to analyze the predictive power. The dataset was split in train and test sets (80% for training and 20% for testing) and 10-fold Cross Validation (CV) was used within the training set. The first model to predict continuous grades was evaluated through the Root Mean Square Error (RMSE), while the other models to forecast binary variables were assessed through the accuracy, kappa and the Area Under the Curve (AUC) (see Table 3-7).

Table 3-7 Evaluation of the predictive models

Cluster	**Set**	**Grade (continuous)**	**Grade >= 80 (binary)**			**Grade >= 80 & prop_lectures >= 0.5 (binary)**		
		RMSE	*Accuracy*	*Kappa*	*AUC*	*Accuracy*	*Kappa*	*AUC*
All	*CV*	11.30	0.95	0.74	0.98	0.96	0.77	0.98
	Test	11.85	0.95	0.70	0.98	0.95	0.70	0.98
Comprehensive	*CV*	16.62	0.82	0.63	0.84	0.82	0.63	0.84
	Test	11.66	0.94	0.86	0.92	0.94	0.86	0.92
Targeting	*CV*	17.22	0.86	0.70	0.92	0.83	0.63	0.92
	Test	17.86	0.80	0.57	0.90	0.90	0.79	0.92

* *CV* – Cross Validation; AUC – Area Under the Curve

Results show that the predictive power is higher with all learners. This is normal because sampler learners are also included, and their grade is easier to predict given that sampler learners do not do the activities and they fail. As for comprehensive, some differences are encountered between the train and test set. The reason is that there are very few comprehensive learners and data limitations may suppose generalization issues.

Nevertheless, the kappa values indicate at least substantial agreement (Landis & Koch, 1977) in all cases (in all groups) and AUC values are excellent (Mezaour, 2005) (excepting the AUC value for comprehensive learners in CV, which can be considered good). These results entail that the new variables related to self-regulated learning and sequence patterns can be useful for predicting grades, together with the well-known activity variables.

3.4.4 Discussion

This section has presented an exploratory study on the variables that are good predictors of the success (grades) for three groups of learners in a self-paced MOOC: “Comprehensive”, “Targeting” and “All” learners. Comprehensive learners are those who follow the course path designed by the teacher. Targeting learners are those who seek for the information required to pass assessments. For both type of learners, we found a group of variables as the most predictive: (1) the self-reported SRL strategies ‘goal setting’, ‘strategic planning’, ‘elaboration’ and ‘help seeking’; (2) the activity sequences patterns ‘only assessment’, ‘complete a video-lecture and try an assessment’, ‘explore the content’ and ‘ try an assessment followed by a video-lecture’; and (3) learners’ prior experience, together with the self-reported interest in course assessments, and the number of active days and time spent in the platform. The variables analyzed in these groups were extracted from self-reported SRL strategies, mined interaction sequence patterns, traditional self-reported variables such as demographics, intentions, and variables that result from the activity of the learner within the platform. Multiple linear regression models were obtained for each of the three groups of learners, which are statistically significant at 99,9% level of confidence.

The findings of this study are subject to some limitations due to the nature of data, and methodological choices. The study is based on learners’ behavioral data automatically collected by the platform, and self-reported data collected from an optional survey.

3.5 SRL strategies employed by students in a MOOC in a blended context

The ease of access to personal computers and mobile devices has allowed students and teachers to have electronic devices that were once considered a luxury and are now used as a tool for work. Added to this is the penetration of the Internet in households worldwide, and especially in Latin America, which has generated a positive impact on the way in which new generations of students are being educated, giving them the opportunity to access a large amount of digital content (Maldonado, Carvallo, & Siguencia, 2015; Pérez-Sanagustín, Maldonado, & Morales, 2016). As a result, new teaching and learning scenarios have been configured, where the teacher assumes new challenges to incorporate digital teaching (Education 2.0) as a complementary part to traditional teaching (i.e., face to face context). This has led to teachers who teach from anywhere in the world and students who learn timelessly, showing that methodologies focused on the teacher or content are no longer the central axis of the learning process (Hamdan, McKnight, McKnight, & Arfstrom, 2013). Also, instead seeks to make the student protagonist of his instruction and that the proposed scenarios are capable of satisfying his learning demands (Maldonado, Bermeo, & Pacheco, 2016, Maldonado, Fernandez-Pampillon, & Sanz, 2015).

In reference to the above mentioned, Bergmann, Sams and Gudentrath (2015), who are known for their pedagogical proposal of Flipped Classroom (FC), state that it is possible to meet the learning needs of students they demand, when the master class turns out not to be as effective, for example when the number of students in class is numerous. According to Cockrum (2013) the use of the pedagogical model of FC supported with technology, allows to hybridise the learning space and optimise classroom time, allowing students to arrive in a personalised way, adapting to their learning rhythm and covering their demands when learning the contents. In this sense, the Higher Education Institutions (HEIs) have begun to explore and experiment with hybrid initiatives, in which they use MOOCs, either their own (produced by the same HEIs) or from third parties (for example, courses on the Coursera platform) along with the FC model. FC seeks to promote more active learning in the classroom, in which students can develop attitudes and skills such as collaboration, critical thinking, creative thinking, trying to develop more autonomous students during their learning (Pérez-Sanagustín, Hilliger, Alario-Hoyos, Delgado-Kloos, & Rayyan, 2017). Based on the work carried out in (Maldonado, Palta, et al., 2016), which was presented in subsection 3.4, in which the current behavior of students in a MOOC is explored and depending on their SRL profile. It was observed that how they interact with the contents of the course was different and context-dependent (Hood et al., 2015). For this reason, this research seeks to explore and present the differences in the behavior of students in a MOOC when it is used as part of a FC proposal, depending on its SRL profile.

3.5.1 Related Work

3.5.1.1 Flipped Classroom Model and Self-regulated Learning

The traditional face to face class is the central, almost exclusive, element that has marked the history of teaching methods in Higher Education spaces. It is a teaching method focused on the teacher and the transmission of information almost exclusively unidirectional between a teacher who plays an active role and students who are passive recipients of information. As an alternative to this teaching method, the pedagogical model of FC has been raised. This model seeks to turn traditional teaching around so that the contents are previously delivered to the students to give more time to practice and application in classroom hours; that is, the contents are reviewed beforehand by the students in their home or place of study and the tasks are carried out in class with the guidance of the teacher. All of this process accompanied by the appropriate technology (Bergmann & Sams, 2015). In this sense, students receive help from teachers to co-regulate their learning process, becoming a metacognitive facilitator to guide students to achieve learning objectives.

Several authors have developed ways to implement the FC model. Bergmann and Sams (2015) have proposed different methodological proposals that include video recording technology and synchronous and asynchronous models of information exchange (e.g., Live

Recording, Flipped Mastery, Flipped Classroom). Another proposal is presented by Erik Mazur (2012) of Harvard University, who states that assimilation must take place in the classroom and the transfer can be done at home by students, mediating the didactic dialogue by the medium of technology. Currently, some universities have implemented FC to strengthen the learning process on different specialities, from language classes, exact sciences, medicine, etc. For example, the University of Uludag - Turkey, conducted a study with 96 students and revealed that the implementation of FC improved the learning process of students, due to the application of strategies centred on the student, and that was reflected in the final grades (Sengel, 2016). The University of Foreign Languages - China conducted a study with 69 students and revealed greater student satisfaction and the possibility of adjusting their learning spaces and times with flexibility. As a result, students achieved higher academic achievements (Zhonggen & Guifang, 2016).

On the other hand, in an analysis of 28 cases of studies carried out in universities in the United States, Australia, United Kingdom, Taiwan and Malaysia revealed that the implementation of this model helped the students to develop critical and independent thinking, building their capacity for lifelong learning and therefore their preparation for future work contexts (O'Flaherty & Phillips, 2015). However, in Latin America few studies have reported at an academic level what is the impact on the students' grades of the use of the FC model. Moreover, there are no studies that relate the SRL profile of students with their behavior in a MOOC when it is used as part of this type of educational initiatives (FC), or the institutional or teacher effort required to achieve it. For this reason, the following subresearch question is proposed: *Sub-RQ 2.5: How does the behavior of students with different SRL profiles differ when a MOOC is used as part of an FC proposal?*

3.5.2 Methods

This section presents the pilot study that was carried out using the methodological proposal of FC developed in this article and supported by a MOOC, to respond to the *Sub-RQ 2.5* posed.

3.5.2.1 Context: Course and Participants

In the study, first-year students of the Faculty of Engineering of the University of Cuenca took the subject "Algorithm, data and structures I", from the semester September 2016 - February 2017. A total of 149 students participated (N = 149), who were randomly assigned to one of the six defined groups (group 1, 2, and 4 with 25 students each, group 3 and 5 with 24 students each and group 6 with 26 students). The students were between the ages of 18 and 20. 19.46% were female, and 80.54% were male. To accompany the FC methodological proposal, a MOOC entitled "Fundamentals of Programming I" was developed, deployed on the Open edX platform, offered by the University of Cuenca in Spanish language and was opened for all students who take or not the subject

The MOOC presents a first introductory module (with the purpose of providing an appropriate context for it to be taken independently as an online course) and three theoretical-practical modules that provide the contents associated with the foundations of programming (e.g., concepts basic, algorithms, conditional and control structures) and the use of the Pseint tool to assist the student in their first steps in programming. The contents were delivered through readings on the platform (35 readings - 62.5%), video-lectures (18 videos - 32.14%) and evaluations (3 evaluations at the end of each module - 5.36%). This MOOC covered the first half of the contents of the subject and was launched on September 12th, 2016. The MOOC was available throughout the semester of classes.

With groups 1, 2 and 3 (experimental group) the FC model was applied and the MOOC of "Programming Foundations I" was used as support for the activities during the first half of the semester and groups 4, 5 and 6 (control group) used the traditional master class method based on exposures. However, these students were not restricted from accessing the MOOC materials, as well as the teachers assigned to these groups, who could or did not use the material at their need. All the professors who taught the subject (for all the groups) were professionals of the Computer Science, all with reliable knowledge about the related contents. Only the professors assigned to the experimental groups (those of group 1, 2 and

3), were trained in how to use a methodological proposal of FC and voluntarily accepted to use it and develop the planning of half of the subject under this methodological proposal.

3.5.2.2 Instruments and Measures

To study the behavior of the students, PM techniques were used, and the methodology reported in subsection 2.2 of this thesis was followed. Also, it was required to characterize the student profile as part of the event log. For this, the self-regulation questionnaire that contains 53 questions related to 13 self-regulation strategies was applied. These strategies were evaluated on a scale Likert from 1 to 5 (1 means nothing true for me, and 5 means very true for me). The questions were presented in a random order. The reliability and validity of the questionnaire were validated in subsection 3.3 (Maldonado, Palta, et al., 2016).

3.5.3 Results

Sub-RQ 2.5: How does the behavior of students with different SRL profiles differ when a MOOC is used as part of an FC proposal?

Figure 3-13 (a) and 3 -13 (b) shows the comparison in the behavior of students who have a high and low SRL level and who belong to the experimental group. Each rectangular box in the process model represents an activity, which, in this case, are the lessons that each module has in the MOOC. The direction of the arrows indicates the path from one activity to another. The thickness of the trajectory line shows the number of times that the trajectory is repeated. The boxes with the strongest blue colour are the activities that have the highest number of repetitions, while the boxes with the weakest blue colour show the least amount of activities repeated.

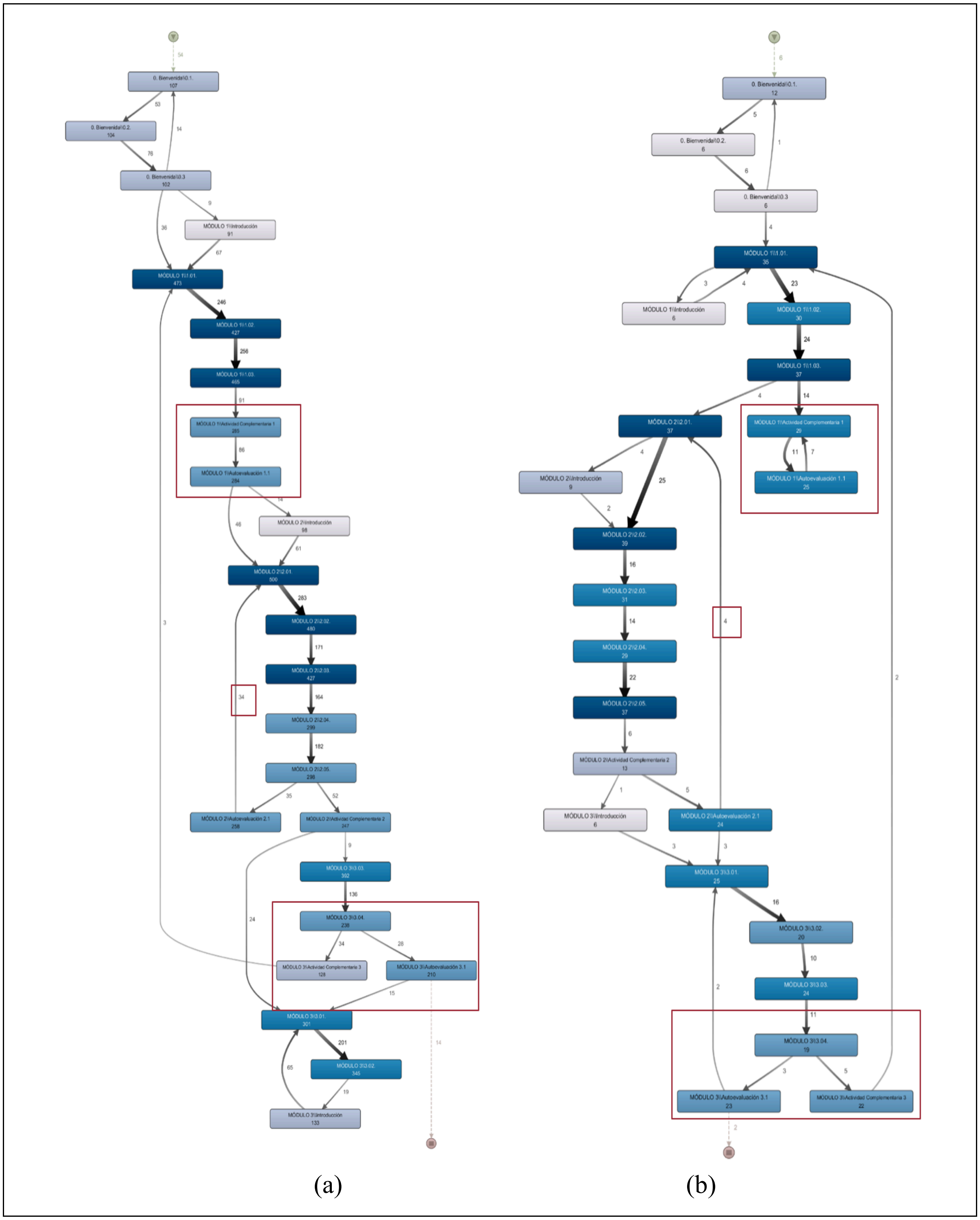


Figure 3-13 Sequence of activities that follow the students in the MOOC with (a) high SRL profile and (b) low SRL profile

In the model of Figure 3-13 (a) the trajectories between the activities of module 1 (1.01→1.02→1.03) and the trajectories between the activities of module 2 (2.01→2.02→2.03) are the most repeated among students who have a high SRL profile. Also, there is a sequential path between the activities of module 1 that correspond to the "complementary activity 1" and the "self-assessment activity 1" while for students with a low SRL profile an iterative cycle between these activities is presented. This would indicate *that these students have to return to look or search for specific content before completing an evaluation activity to continue*.

On the other hand, Figure 3-13 (a) shows that students with a high SRL profile tend to follow the linear sequence proposed by the course, but also return to the beginning of Module 2 after completing the evaluation activity at the end of the module (2.01→2.02→2.03→2.04→2.05→Autoevaluación→ 2.01). *This would indicate that students after self-assessment tend to return to the beginning of the module in order to organize or recapitulate the concepts learned*.

Finally, in the process model of Figure 3-13 (a) students with high SRL profile tend to go after completing self-assessment_module2 directly to activities 3.3 and 3.4 before addressing the self-evaluation_ module3 (Autoevaluación_módulo2 → 3.03 → 3.04 → Autoevaluación_módulo3) while their counterparts tend to make a sequential path (Autoevaluación_módulo2 → 3.01 → 3.02 → 3.03 →3.04 → Autoevaluación_módulo3). *This would indicate the ability of students with a high SRL profile to try to walk the last module autonomously.*

3.5.4 Discussion

This section contributes a new perspective for the study and understanding of how students self-regulated in a MOOC under the FC model. Although, there is previous work in the literature where the trajectories of students in a MOOC have been explored (Guo & Reinecke, 2014, Mukala, Buijs, & van der Aalst, 2015, Sonnenberg & Bannert, 2015), this is one of the first studies that analyze these trajectories and their relation with the SRL when FC model is used (Maldonado, Palta, et al., 2016). If we compare the process model of students in the blended context who have a high SRL profile (Figure 3-13 a) and the process model of students in the online course who have a medium SRL profile (subsection 3.3 in this chapter), it can be seen that in both models the behavioral pattern of the students throughout the course is similar. That is, they tend to return to the beginning of the module with the purpose of organizing or recapitulating the concepts learned. However, the students of the online course did it spontaneously without having received any instruction while taking the course, while the students who used the FC pedagogical model did so after receiving guidance from the teacher. In this case, the teacher was a "metacognitive facilitator". The above would show the difference that exists at the level of learning sequences, and that is evident in the regulation of students, in an online and a hybrid context.

On the other hand, students in the experimental group with high SRL profile tend to: (1) return to look or search for specific content before completing an evaluation activity to continue, (2) back to the beginning of the module for organizing or recapitulating the concepts learned and (3) to try to walk the last module autonomously. The generalization of these results is subject to the limitations of the methodology used in the study. The results obtained through the use of PM techniques are directly related to the data collected as a basis for this study. The process models obtained are closely related to the structure of the course that supports the Open edX platform.

3.6 Conclusions

This chapter has presented the work performed in order to answer the main ***RQ2: What is the relationship between SRL strategies and academic performance, taking into consideration the characteristics of the participants, the MOOC and the course context that influence the use of these strategies?*** To address the RQ2, five sub research questions (Sub-RQ) have been proposed:

- Sub-RQ 2.1: Which self-reported SRL strategies are most helpful to achieve personal course goals?
- Sub-RQ 2.2: How do self-reported SRL strategies vary by individual learner characteristics?
- Sub-RQ 2.3: How do the different SRL and LS profiles manifest themselves in a MOOC in terms of learning sequences?
- Sub-RQ 2.4: Which indicators of SRL obtained from self-reported questionnaires and activity sequence extracted from trace data can predict course success in self-paced MOOCs?
- Sub-RQ 2.5: How does the behavior of the learners with different SRL profiles differ when a MOOC is used as part of a Flipped Classroom proposal?

The work developed to address these five Sub-RQ has allowed to achieve five main contributions in this thesis that are: (1) the identification of the SRL strategies that are most helpful to achieve personal goals and intentions in MOOCs; (2) the identification of the individual learner characteristics that predict the use of SRL strategies in MOOCs; (3) the proposal of a classification of learners based on the relation between their SRL strategies deployed and their performance in MOOCs, depending on SRL profile, LS profile and depending the learning sequences deployed by learners in MOOCs; (4) the identification of the SRL strategies that predict learners' success in MOOCs; and (5) the identification of the SRL strategies employed by students in a MOOC in a blended context.

The first contribution has been to identify the SRL strategies that are most helpful to achieve personal goals and intentions in MOOCs. Specifically, we found that learners who reported engaging more in goal setting and strategic planning were more likely to attain personal course goals, such as earning a certificate.

The second contribution has been the identification of the individual learner characteristics that predict the use of SRL strategies in MOOCs. For example, in gender, in particular women, were more inclined to seek help than men and learners with a Ph.D. were generally more self-regulated, but much less inclined to seek help. In contrast, learners who were also students reported lower SRL skills, that is low levels in self-evaluation and task strategies, while learners with ambitious course intentions, greater time commitment, and prior experience with the topic generally indicated stronger SRL skills. Finally, motivations for taking the course that signaled a supportive life context (i.e., course relevant to job/school/research) predicted stronger SRL skills, while motivations that signaled a less supportive context (taking course for fun and challenge, to experience a MOOC, for career change) predicted weaker SRL skills.

The third contribution has been to propose a classification of learners based on the relation between their SRL strategies deployed and their achievements in MOOCs. Specifically, we found four groups of learners: (1) Sampling learners, who only explore the contents of the course without attempting to finish any video-lecture or assessment and had low activity in the course, (2) Targeting learners, who work intensively with assessments with the objective of certificate the course, but without attempting the video-lectures; (3) low Comprehensive learners, who worked intensively with the course materials, but focus on summative more than formative assessments and (4) highly Comprehensive learners, who work also intensively with course materials and on average perform a large number of study sessions, showing the intention of learners to achieve a deep understanding of the contents and self-evaluate their progress. They follow the course path designed by the teacher.

The fourth contribution has been to identify the SRL strategies that predict learners' success in MOOCs. Specifically, we found the variables that are good predictors of the grades for "Comprehensive" and "Targeting" learners in a self-paced MOOC. For Comprehensive learners (nor low or highly), the self-reported variable *goal setting*, the sequences patterns *only assessment*, *explore* and *assessment try to video-lecture*, the reported demographics as *young learners*, be *women* and *employment status as student*, the learners' *prior experience* and *interest in assessment* reported, the *active days* and the *time spent* were significant predictors of the final grade. For Targeting group, the self-reported variables *strategic planning*, *elaboration* and *help seeking*, the sequences patterns *only assessment*, *video-lecture complete to assessment try*, *explore* and *assessment try to video-lecture*, the reported demographics as young learners, the learners' *prior experience*, the *time spent,* and the *number of sessions* were significant predictors of the final grade.

The fifth contribution has been to identify the SRL strategies employed by students in a MOOC in an online and blended context. Specifically, we found differences between the different SRL profiles. For those learners who reported high SRL profile in online context, we found that they tend to complete most of the video-lectures consecutively and pass the assessments. Also, they tend to return to the beginning of the module with the purpose of organizing or recapitulating the concepts learned; while those with low SRL profiles struggle with the contents of the course and attempt to pass assessment without achieving it and back for revisiting video-lectures that they completed in the past. In the case of blended context, those learners who reported high SRL profile tend to follow the path of the course and return to look or search for specific content before completing an assessment activity to continue. Also, return to the beginning of the module to organize or recapitulating the concepts learned. However, they received instructions from the teacher for recapitulating while learners with high SRL profile performed the same behavior but without received neither instructions. For those learners with medium SRL profile in blended context they tend to follow the path of the course until the end, while learners with high SRL profile tend to pass the assessment before watching the video-lectures.

Chapter **4**

Conclusions and lessons learned

The roots of education are bitter, but the fruit is sweet.

Aristotle

The main contributions of this thesis are framed within the study of Self-regulated Learning in MOOCs. In particular, this thesis contributes with a set of instruments and methods for studying SRL strategies in MOOCs and analyzing how they relate with academic performance. Specifically, subsection 4.1 introduces a summary of the main contributions: (1) the instruments and methods for measuring SRL in MOOCs; and (2) the relationship between SRL strategies and academic performance. Subsection 4.2 reviews the lessons learned from the application of these instruments and methods for the study of SRL in MOOCs. Subsection 4.3 introduces the limitations of this work. Finally, subsection 4.4 introduces the new research avenues derived from this thesis.

4. CONCLUSIONS AND LESSONS LEARNED

4.1 Summary of contributions

The primary motivation of this thesis has been to explore the new opportunities and challenges that LA offers to study learners' SRL strategies in MOOCs. Specifically, two main research questions have been addressed in this thesis: ***RQ1.*** What ***instruments*** and ***methods*** are more appropriate to explore learners' ***self-regulatory strategies*** used in MOOCs? And ***RQ2.*** What is the ***relationship*** between ***SRL strategies*** and ***academic performance,*** taking into consideration the characteristics of the participants, the MOOC and the course context that influence the use of these strategies? We highlight in what follows what the main contributions associated with these research questions are.

[Contribution 1] The first contribution is a self-reported questionnaire for capturing students' SRL strategies in MOOCs. This questionnaire is the result of a literature review on how SRL has been measured in traditional face-to-face and online contexts. This analysis of the literature has shown how the measurements of SRL have evolved with time and emphasizes the lack of current instruments and methods to study SRL as an aptitude and as a process in MOOCs. The proposed self-reported questionnaire assesses 5 SRL strategies (Self-efficacy, Goal setting, Study environment management, Organization, and Help-seeking) through 22 items proposed and evaluated in a 5-point Likert scale in order to study SRL as an aptitude. The questionnaire has been tested and evaluated; and together with the literature review are published online in **[J1] - Journal of Educational Review** (literature review), and (2) [J2] – **Journal of Research on Technology in Education** (the instrument, Under Review) (Chapter 2).

[Contribution 2] The second contribution is a methodological approach based on Process Mining techniques to extract SRL strategies from fine-grained data in MOOCs. This methodological approach is the first contribution in the field that proposes an SRL process perspective of learners' strategies in MOOCs. The proposed methodology consists of four stages: (1) extraction, (2) event log generation, (3) model discovery, and

(4) model analysis stage. These four stages help us to identify and analyze the digital traces of the learners' activities in a MOOC. One of the main novelties of this methodology is that it proposes combining SRL aptitude-based approach with a process-based approach to investigate SRL strategies in MOOCs across contexts, by relying on both a self-report instrument and process mining of behavioral learners' data. Specifically, this methodology allows analyzing unfolded fine-grained data from MOOCs: (1) to identify the most frequent interaction sequence patterns that learners exhibit in a MOOC; (2) to differentiate interaction sequence patterns between learners with different characteristics; (3) to identify learner profiles based on their observed interaction sequence patterns and; (4) to associate observed interaction sequence patterns with SRL strategies established in SRL theory. The methodology was also adapted to be used in a different context of MOOCs providing external validity to the methodology for studying SRL both, as an aptitude and as a process. The results of this methodological approach have been published in **[J3] - Journal of Computers in Human Behavior** and replicated in another paper in **[J4] - Journal of Computing in Higher Education** (under review - Chapter 2).

[Contribution 3] The third contribution is a set of empirical results of applying the above mentioned instrument and methods in different educational contexts with MOOCs for analyzing the relationship between the SRL strategies and academic performance considering three aspects: (1) learners' individual characteristics; (2) characteristics of the MOOC; and (3) characteristics of the context in which the MOOC is deployed. The results of this analysis can be summarized into four main findings (Chapter 3):

> **(1) Goal setting and strategic planning are the self-reported SRL strategies that better correlate with attaining personal course goals.** Learners who reportedly engaged in these metacognitive SRL strategies were more likely to achieve their course goals. Women were more inclined to seek help than men and learners with a Ph.D. were generally more self-regulated but much less inclined to seek help. In addittion, learners who reported lower levels in self-evaluation and task strategies were undergraduate students, while learners who reported ambitious

course intentions, greater time commitment, prior experience with the topic of the course and motivations for taking the course given the relevance for supportive life context indicated stronger SRL skills. Finally, self-reported SRL strategies as goal setting, strategic planning, self-evaluation, and task strategy are associated with better performance in the course.

(2) Learners can be classified into four categories based on their SRL strategies deployed in interaction with the MOOC content (captured in trace data): sampling learners, targeting learners, low comprehensive learners, and highly comprehensive learners. (a) Sampling learners, who only explore the contents of the course without attempting to finish any video-lecture or assessment and had low activity in the course, (b) Targeting learners, who work intensively with assessments with the objective of certificate the course, but without attempting the video-lectures; (c) Low Comprehensive learners, who worked intensively with the course materials, but focus on summative more than formative assessments and (d) Highly Comprehensive learners, who work also intensively with course materials and on average perform a large number of study sessions, showing the intention of learners to achieve a deep understanding of the contents and self-evaluate their progress. Also, they typically follow the path of the course designed by the teacher.

(3) Self-reported SRL strategies (goal setting, strategic planning, elaboration and help-seeking), behavioral sequence patterns (only assessment, explore and assessment try to video-lecture, video-lecture complete to assessment try), learners' demographics (gender, occupation, prior experience) and activity of the learner within the platform (active days, time spent, number of sessions) are good predictors of the grades for "Comprehensive" and "Targeting" learners in a self-paced MOOC. For Comprehensive learners (both low and highly), the self-reported variable goal setting, the sequences patterns only assessment, explore and assessment try to video-lecture, the reported demographics as young learners, be women and employment status as student, the learners' prior experience and interest in assessment reported, the active days and the time spent were significant predictors of the final grade. For Targeting learners, the self-reported variables strategic planning, elaboration and help-seeking, the sequences patterns only assessment, video-lecture complete to assessment try, explore and assessment try to video-lecture, the reported demographics as young learners, the learners' prior experience, the time spent, and the number of sessions were significant predictors of the final grade.

(4) The context in which the MOOC is deployed, either online or blended, influences in learners' SRL strategies. In an online context, those learners who reported high SRL profiles, tend to complete most of the video-lectures consecutively and pass the assessments. Also, they tend to return to the beginning of the module to organize or recapitulate the concepts learned; while those with low SRL profiles struggle with the contents of the course and attempt to pass assessment without achieving it and back for revisiting video-lectures that they completed in the past. In the case of blended context, those learners who reported high SRL profiles tend to follow the path of the course and return to look for or search for specific content before completing an assessment activity to continue. Also, they return to the beginning of the module to organize or recapitulating the concepts learned. Moreover, learners who reported medium SRL profile tend to follow the path of the course designed until the end; that is, they see the video-lectures and attempt the assessment afterward. However, learners with high SRL profile tend to pass the assessment before watching the video-lectures.

4.2 Lessons learned

The contribution of this thesis led to a set of lessons learned for the study of SRL in MOOCs that can be organized into a conceptual framework. This conceptual framework is based on the bibliography analyzed during this thesis, in the current theories of self-regulation and the empirical evidence collected through this research work.

4.2.1 Conceptual Framework

The main aim of this framework is to organize the main lessons learned of this thesis and provide a guideline for guiding other researchers in the analysis of SRL in MOOCs. Figure 4-1 presents the preliminary conceptual framework. The conceptual framework is organized around three key elements (blue boxes) that should be considered for the analysis of SRL strategies (green box with white text) in MOOCs. These key elements are: (1) the characteristics of the learners, (2) the characteristics of the MOOC; and (3) the context in which the MOOC is deployed. Each of these three elements refers to specific student attributes (i.e., gender, occupation and prior-experience/knowledge), MOOC attributes (i.e., pace, certification, activities and length of the course) and the context in which the MOOC is used (i.e., online) that are important to consider when studying SRL strategies in MOOCs. These attributes have emerged from the empirical evidence collected throughout this work. The orange dotted line box on the left, and the green dotted line box on the right refer to the types of analyses that can be conducted when studying SRL strategies in MOOCs (i.e., studying SRL as an aptitude and as a process). When studying SRL as an aptitude, the most common instrument to be used is a self-reported questionnaire to collect the intentions and actitudes of the students with the course, objectives, and strategies of SRL. The learners' answers to this questionnaire can be processed to define student's SRL profiles. When studying the SRL as a process, students' trace data capturing their interaction with the course contents should be analyzed to study learners' actual behavior within the platform. Both approaches combined help to establish the SRL strategies (i.e., cognitive, metacognitive and resource management) that students deploy in a MOOC. Combining the analysis of SRL strategies (as an aptitude and as a process) provides the necessary information to understand better how SRL relates to academic performance. This academic performance can be measured both, as the students' performance within the course (pass or fail), or as the set of strategies that they deploy within the course (targeter, comprehensive and sampler learners).

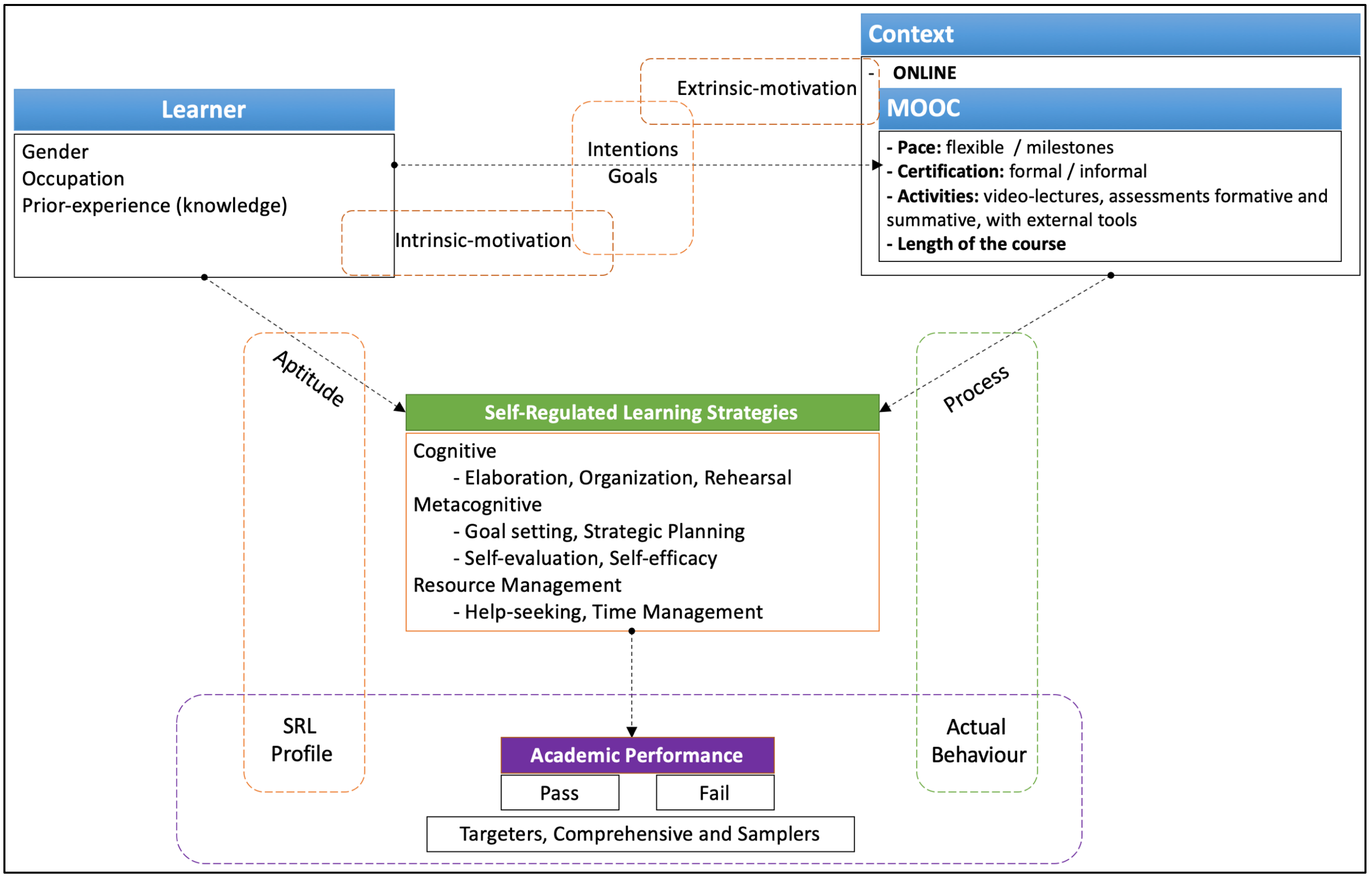


Figure 4-1 Proposed Framework to study SRL in MOOCs

4.2.2 Implications of the conceptual framework regarding the bibliography

Based on the different models available, there have been several articles about SRL measurement. These measurements have evolved with time and have led to what Panadero et al. (2016) called "waves of measuring SRL." Panadero et al. (2016) identify three waves of measurement: the first one is characterized by a static conceptualization to measure SRL, and its emphasis is placed on the use of self-report instruments to describe SRL processes following an aptitude approach. The second wave of measurement conceptualizes SRL as a process and not as a trait. In this case, they propose using measures "online" which are able to study students' self-regulatory processes in a particular moment during the activity. And, the third wave proposes studying SRL while scaffolding students' strategies through technological systems. This approach proposes developing tools to scaffold certain SRL strategies at the same time that it serves for capturing students' "traces" about their actual behavior.

The first wave of SRL measurement, or the study of SRL as an aptitude, has two significant limitations when used in a MOOC context: (1) self-reported SRL strategies are primarily based on students' perspectives and beliefs, which is not a reliable and accurate source of information; and (2) depending on students' intentions and goals with course contents, SRL strategies may vary over time during the course, so this approach to study SRL as an aptitude does not capture changes in students' SRL strategies. These results are aligned with what we found when analyzing the strategies in an online context (sections 2.2 and 3.2), in which we observed that SRL strategies change over time. These findings were published in three papers: in two journal papers **[J1] - Journal of Educational Review; [J5] - Journal of Computers and Education** and conference paper **[C3] - III Conference on Learning at Scale ACM - 2017**

Regarding the second wave of SRL measurement or the study of SRL as a process, this thesis provides empirical evidence that the proposed methods overcome limitations of the aptitude approach, which also aligns with other recent works (Alonso-Mencía et al., 2019; Roth et al., 2015). The process approach led to a set of "events measures" from traces captured by technological tools. This thesis contributes to extending this second wave of SRL measurement by proposing a methodological approach based on PM. This methodological approach has been evaluated with data from MOOCs in Coursera, edX and Open edX platforms. According to Van den Beemt, Buijs, and Van der Aalst (2018), this methodology is appropriate to get a better understanding of underlying educational processes and allows determining patterns from sequences of activities in MOOCs. Aside from previous studies conducted with courses in traditional higher education settings (Jovanovic et al., 2017; Lust, Elen, & Clarebout, 2013), to the best of our knowledge, this thesis presents the first process-based study of SRL in MOOCs. These results were published in six papers as contributions: in two journal papers **[J3] – Journal of Computers and Education; [J4] – Journal of Computers in Human Behavior** and four conference papers **[C1] - XLII IEEE Informatics Latin America Conference - CLEI 2016; [C4] - XII IEEE Latin American Conference on Learning Technologies LACLO - 2017; [C5] - HybridEd Workshop 2018: Succesful and Promising Experiences in Blended Learning with MOOCs; [C6] – I Learning Analytics Latin America Conference 2018.** This way of studying the SRL statically using self-reported data must be complemented with behavioral data that result from the student's interaction with the contents of the MOOC and that are recorded on the platform. The combination of these two data sources allows contrasting the preferences in the use of SRL strategies with the actual use of these strategies in the platform.

This thesis has also contributed to the third wave of SRL measurement, which proposes studying the SRL while scaffolding students' strategies through computational systems. Specifically, the analyses presented in chapter 2 and chapter 3 served as a base for the design of a tool called NoteMyProgress (Pérez-Álvarez, Maldonado-Mahauad, & Pérez-Sanagustín, 2018). NoteMyProgress is a dashboard for learning analytics, aimed at supporting self-regulatory strategies that have been seen to be effective (goal setting, strategic planning, time management, and organization) that students use in MOOCs. This tool serves for capturing students' "traces" about their actual behavior in MOOCs, and contributes to the study of SRL at scale, unlike the proposed tools nStudy and gStudy (Winne & Hadwin, 2013). These results were published in a journal paper **[J6] - Journal of Universal Computer Science.**

The empirical evidence collected in this study has shown that there are specific characteristics of the students, and of the MOOC when used in a non-formal context, that is important and that should be taken into consideration when studying SRL in MOOCs (Kizilcec et al., 2017). Regarding the characteristics of the students, the evidence indicates that gender, occupation, knowledge and or previous experience and intrinsic motivation affect their use of the SRL strategies employed. Specifically, prior knowledge and occupation of students are intrinsic motivating agents (Hood et al., 2015) as well as the characteristics of the course becomes extrinsic motivating agents (Jordan, 2014), so the student manages to establish their intentions and goals with the contents of the course.

Regarding to the characteristics of the MOOC, the evidence shows that the pace (if the course is flexible and with or without milestones), the proposed type of activities (video-lectures, formative or summative assessments, third party activities) and the length of the course influence over the SRL strategies employed by learners in MOOCs (Ferguson & Clow, 2015; Freitas et al., 2015; Guo & Reinecke, 2014). When learners study in a MOOC, the processes of goal-setting, strategic planning, self-monitoring, time management and help-seeking fall directly on the student, where they are expected to be able to establish their learning objectives, plan their study session, monitor their progress and manage their time appropriately, seeking help from other peers or contents of the course to advance.

Furthermore, those who take and finish a self-paced MOOC reported different intentions and objectives concerning the course. On the one hand, students seeking only to pass the course use strategies that are focused on those that allow them to evaluate their knowledge and achieve the certificate of the course (if it is available), leaving aside the content that may present the course employing video-lectures or another type of material. This result suggests students' intentions could guide course design in order to provide them with a personalized experience better adapted to their initial intentions. On the other hand, students who are looking to either finish or certify the course seek to engage with the course contents and learn in depth the contents. These students deploy SRL strategies that allow them to align their objectives with those of the course, follow the structure of the proposed course (which consists of presenting a series of contents and then evaluating them). Moreover, these learners return to review content previously visited in the course to organize the blocks of information as they progress and synthesize what they have learned. This demonstrates the importance of the pedagogical sequencing of the contents and the activities in the proposed course (Joksimović et al., 2015; Zheng et al., 2015; Zhenghao et al., 2015).

In the case of those MOOCs designed with milestones, where students have some flexibility to start and finish the course, we showed that the formative activities help especially to students who commit themselves with the contents of the course. Formative activities promote students' strategies of self-monitoring, self-evaluation, and help-seeking. The above suggests that the evaluation activities in these types of MOOCs should be designed to provide students' with immediate feedback either to reinforce the correctness or to correct based on the error (Alario-Hoyos et al., 2014; Ferguson & Clow, 2015; Hew & Cheung, 2014).

In a MOOC, academic performance is usually measured based on whether students' pass or not pass the course. However, this way of defining academic performance must be rethought by taking into account students' objectives and intentions with the course (Clow, 2013; Taub et al., 2014; Zheng et al., 2015). Based on the empirical evidence, the term academic performance can be redefined and must be related to the student's behavior performance. Also, this redefinition of academic performance can guide the design of tools to support specific SRL strategies of students aligned with their particular objectives, which are not necessarily the objectives of the course (Jordan, 2014; Kizilcec et al., 2016). These tools should be designed to study students' SRL strategies "on the fly", to understand how strategies vary over time and depending on the support provided. In this way, we would have, not only a photograph of the SRL process but a series of photographs that would allow us to understand what strategies are more useful to support and in which moment (Panadero et al., 2016).

4.3 Limitations

The study conducted through this thesis has three main limitations that, although common in this type of research, should be noted when concluding from our findings.

The first limitation concerns the external validity of the samples analyzed. This thesis used samples of mostly Latin American learners engaging in MOOCs that were offered in Spanish (excepting for one study performed over a MOOC in English and run into the edX platform, presented in subsection 2.4). On the one hand, given that most published findings are based on samples from Western-educated industrialized rich democratic countries (Henrich, Heine, & Norenzayan, 2010), this study advances the inclusivity of our science by drawing on a non-traditional sample. On the other hand, as noted here and in prior work, learners' socio-cultural context has consequences for how they perceive and engage with online courses (e.g., Guo & Reinecke, 2014; Ogan et al., 2015). While our findings are consistent with prior work that considered other international populations, future work should replicate and extend the current findings with other samples to test generalizability.

The same argument applies to the specific courses that were studied, which were self-paced MOOCs on the Coursera platform in 2015, and synchronous MOOCs on the edX and open edX platform. Prior work found differences between the staggered versus all-at-once content release format for MOOCs in terms of persistence and completion in the course (Mullaney & Reich, 2015). While the courses covered a wide range of topics, the design, the content narrative and instrumentation of the platform at the time are expected to play an important role in shaping learner behavior and researchers' interpretation of their behavior through the lens of the collected data. This highlights a structural limitation with implications for both the replicability of findings across platforms and time, and the reliability of inferences that can be drawn from meta-analyses of related research findings.

The second limitation concerns construct validity of the self-reported questionnaires. The instruments we used to assess SRL are based on established and validated instruments in the literature. However, we did not employ any complete instrument available in the literature. Instead, we identified thirteen relevant SRL strategies from prior literature and adapted established instruments to measure the five selected constructs (SRL strategies) specifically. This approach made a trade-off between utilizing a complete instrument with many items that are unsuitable in the MOOC context, on the one hand, and creating entirely new survey items to measure established constructs, on the other hand. It is just unreasonable to ask an online volunteer learner population to fill out a lengthy battery of survey questions and expect to receive data that is of high quality. Another consideration regarding measurements is that we translated the entire questionnaires into Spanish (in total we adapted and used 3 versions of the SRL questionnaires), including the measure of SRL. The translation was performed by two native speakers who understood the underlying constructs that were assessed. Valid translation of survey instruments is a non-trivial issue and it warrants empirical validation.

The third limitation concerns regarding the proposed methodology. The proposed methodology in this thesis is subject to some limitations posed by the nature of the data and methodological choices:

(a) First, we conducted an observational field study with automatically recorded behavioral records and data collected from an optional survey. The observations thus occurred in an actual learning environment, which is a relatively uncontrolled research setting. Prior work on SRL and learning processes that were conducted in online environments utilized research platforms developed or adapted to support SRL, for instance by adding functionalities directly associated with a self-regulation strategy (Beheshitha et al., 2015; Sonnenberg & Bannert, 2015). Field studies in MOOCs typically yield higher external validity for lower control over the research process. For example, the optional nature of the self-reported SRL instrument can raise concerns about self-selection bias, because the survey was used as a basis for including learners in the final study sample. This implies that all participants in the different studies run in this thesis, tended to be more motivated than the average learner enrolled in the courses.

(b) Second, we made several methodological choices in this study that may have influenced the results. For example, we computed for all studies in this thesis the session time based on an inactivity threshold of 45 min and we only studied learners' interactions with two learning resources in all of the courses (video-lectures and assessments), excluding interactions on the discussion forums (this decision was made because hardly any forum interactions occurred in some cases and in others we did not have access to the log data of these activities). We highlight three methodological choices in the analysis that may have influenced our findings. First, like in any data mining or machine learning context one cannot assume to have seen all possibilities in the 'training material' (Van Der Aalst, 2016). Processes typically allow for an exponential or even infinite number of different patterns. It is therefore unrealistic to assume that every possibility is represented in the dataset. Instead, the data is considered a sample of learners' potential and observable behavior (Bose, Mans, & Van Der Aalst, 2013). In a future project, we plan to perform the same analysis on other platforms (e.g., FutureLearn, Moodle) and context (e.g., blended context) to understand the extent to which the present findings are contextually bounded to the affordances of the learning environment. Recent evidence suggests the importance of contextual factors on learner behavior, but it has not been analyzed on a process level to date (Conole, 2015). Moreover, complex multidimensional and multi-granular data needs to be 'flattened' in order to be represented by simple process models (Van Der Aalst, 2016). We attempted to retain a fine level of granularity in the behavioral models, but other levels of granularity are also possible.

(c) Finally, process analysis is, by definition, restricted by the expressive power of the process modeling language (Van Der Aalst, 2011). If the modeling language cannot represent something, then it cannot be observed, resulting in representational bias. The simple process maps used to illustrate the interaction patterns in this study were closely aligned with the analysis of SRL, but alternative process modeling notations with more complex patterns could also be possible. However, the discovery of more complex patterns possesses additional challenges. Overall, we used clear definitions of events and described our methodology in detail to provide the necessary accuracy to make this research reproducible. We hope that this research can serve as a reference point for other researchers who would like to analyze their courses using a PM approach combined with self-reported data to advance our scientific understanding of how individuals learn in MOOCs.

4.4 Future work

Apart from the aforementioned contributions and lessons learned, this thesis opens up new research avenues.

(a) **Regarding the contribution in instruments and methods for measuring SRL.** The self-reported questionnaire and the proposed methodology have allowed us to study the SRL strategies of students in MOOCs. However, in this thesis some deficiencies have been evidenced with respect to the existing self-reported questionnaires to be applied in a virtual environment, and that is adapted to the context of a MOOC, moreover, there is still a need to create a self-reported questionnaire for measuring SRL strategies in a context either different or complementary to online (e.g., blended). In this sense, a new research question is opened regarding the development of a new questionnaire to study SRL strategies in a different context than online: ***Which SRL strategies are possible to study in a blended context employing self-reported questionnaires?***

On the other hand, regarding the methodological proposal, this thesis demonstrates that it serves for studying SRL as a process in an online context, by extracting students' self-regulatory processes when interacting with the contents of a MOOC. However, the MOOCs used in this thesis are MOOCs deployed on the Coursera, edX and Open edX platforms, taught mostly in a self-paced format (exepting those run in edX and Open edX platforms), in Spanish and used in an informal context. The above led us to think in two new questions that arise about: 1) the adaptability of the methodology presented in this thesis to study strategies of SRL in a different context, where the features of the MOOC platform (i.e., length, pace, type of activities, sequence of the contents, quality in the video-lectures contents, approach of the technological platform) and the type of use of the MOOC (i.e., as a driver, as a service, as an added value, as a replacement as proposed in Pérez-Sanagustín et al. (2016) are different from the research in this thesis; and 2) the possible improvements that the methodology could have in order to standardize the analysis of trace data using PM techniques agreed by the experts in the area of PM and Educational Sciences. We plan to explore the application of the methodology in datasets of different granularities to study different types of patterns from the behavior of learners at different levels, macro (i.e., week by week) and micro (i.e., by activity or by clickstream) and even combining more metrics such as the time invested.

(b) Regarding the relationship between SRL strategies and academic performance. The identification of the relationship between the SRL strategies that students use in a MOOC with their academic achievements allowed us to understand which strategies are essential during the SRL process, and how a student reaches a certain "status" based on their behavior in an online context. However, as future work, we plan to study more in-depth the SRL strategies in MOOCs when used in a formal context (i.e., as a replacement of the official course in HE) in order to: (a) unveil those strategies that are most helpful to achieve personal goals; (b) identify new learners' behavior based on the interaction within the MOOC; (c) identify the SRL strategies that predict academic achievements in MOOCs and relate with learners' characteristics in formal context.

In addition, the identification of SRL strategies that are related to the academic performance of the students can serve as a basis for proposing LA tools that support specific SRL strategies of students in online environments. For this, indicators of SRL strategies could be presented as visualizations to support students' self-regulatory processes and study how these tools impact or not on their learning process and analyze the behavior of the learner, even beyond the MOOC. In this line, we have started exploring the behavior of the learners applying the proposed methodology in chapter 2 to study the trace data recorded by the tool NoteMyProgress, designed to support SRL strategies in MOOCs (Pérez-Álvarez, Maldonado-Mahauad, & Pérez-Sanagustín, 2018). The study of SRL strategies in MOOCs and its relationship with academic achievements in an online (even more in a blended) context and the tools that can be developed to support SRL strategies in this type of context is still a line of research that remains open, and that should be further explored.

BIBLIOGRAPHY

Agrawal, R., & Srikant, R. (1995). Mining sequential patterns. In *icde* (p. 3).

Alario-Hoyos, C., Estévez-Ayres, I., Pérez-Sanagustín, M., Delgado-Kloos, C., & Fernández-Panadero, C. (2017). Understanding learners' motivation and learning strategies in MOOCs. *The International Review of Research in Open and Distributed Learning*, *18*(3).

Alario-Hoyos, C., Muñoz-Merino, P., Pérez-Sanagustin, M., Delgado-Kloos, C., & Parada, G. (2016). Who are the top contributors in a MOOC? Relating participants' performance and contributions. *Journal of Computer Assisted Learning*.

Alario-Hoyos, C., Pérez-Sanagustín, M., Delgado-Kloos, C., Gutiérrez-Rojas, I., Leony, D., & Parada, G. (2014). Designing Your First MOOC from Scratch: Recommendations After Teaching "Digital Education of the Future." *ELearning Papers, From the Field*, *37*(3), 1–7.

Alario-Hoyos, C., Pérez-Sanagustin, M., Delgado-Kloos, C., & Munoz-Organero, M. (2014). Delving into participants' profiles and use of social tools in MOOCs. *IEEE Transactions on Learning Technologies*, *3*(7), 260–266.

Alharbi, A., Paul, D., Henskens, F., & Hannaford, M. (2011). An investigation into the learning styles and self-regulated learning strategies for computer science students. In *Proceedings ascilite*.

Alonso-Mencía, M. E., Alario-Hoyos, C., Maldonado-Mahauad, J., Estévez-Ayres, I., Pérez-Sanagustín, M., & Delgado Kloos, C. (2019). Self-regulated learning in MOOCs: lessons learned from a literature review. *Educational Review*, 1–27.

Alonso, C., Gallego, D., & Honey, P. (1994). *Los Estilos de Aprendizaje, Procedimiento de Diagnostico y Mejora*. Ediciones Mensajero. Universidad de Deusto. Bilbao.

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with massive online courses. *WWW '14 Proceedings of the 23rd International Conference on World Wide Web*, 687–698. https://doi.org/10.1145/2566486.2568042

Arias Chaves, M., & Rojas Cordoba, E. (2014). Deciphering event logs in SharePoint Server: A methodology based on process mining. In *Computing Conference (CLEI), 2014 XL Latin American* (pp. 1–12).

Azevedo, R., & Aleven, V. (2013). Metacognition and learning technologies: an overview of current interdisciplinary research. In *International handbook of metacognition and learning technologies* (pp. 1–16). Springer New York.

Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. G. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development*, *56*(1), 45–72.

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology*, *96*(3), 523.

Bahamón Muñetón, M. J., Vianchá Pinzón, M. A., Alarcón Alarcón, L. L., & Bohérquez Olaya, C. I. (2012). Estilos y estrategias de aprendizaje : una revisión empírica y conceptual de los últimos diez años. *Pensamiento Psicológico*, *10*(1), 129–144. https://doi.org/10.11144/183

Banerjee, A. V., & Duflo, E. (2014). (Dis)Organization and success in an economics MOOC. In *American Economic Review* (Vol. 104, pp. 514–518). https://doi.org/10.1257/aer.104.5.514

Bannert, M. (2009). Promoting self-regulated learning through prompts. *Zeitschrift Für Pädagogische Psychologie*, *23*(2), 139–145.

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

Barba, P. G. de, Kennedy, G. E., & Ainley, M. D. (2016). The role of students' motivation and participation in predicting performance in a MOOC. *Journal of Computer Assisted Learning*, *32*(3), 218–231.

Barnard-Brak, L., Paton, V. O., & Lan, W. Y. (2010). Profiles in self-regulated learning in the online learning environment. *The International Review of Research in Open and Distributed Learning*, *11*(1), 61–80.

Barnard, 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, L., Paton, V., & Lan, W. (2008). Online self-regulatory learning behaviors as a mediator in the relationship between online course perceptions with achievement. *The International Review of Research in Open and Distributed Learning*, *9*(2).

Basol, G., & Balgalmis, E. (2016). A multivariate investigation of gender differences in the number of online tests received-checking for perceived self-regulation. *Computers in Human Behavior*, *58*, 388–397.

Baumert, J., & Köller, O. (1996). Lernstrategien und schulische Leistungen. *Emotionen, Kognitionen Und Schulleistung*, 137–154.

Beheshitha, S. S., Gašević, D., & Hatala, M. (2015). A process mining approach to linking the study of aptitude and event facets of self-regulated learning. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*, 265–269. https://doi.org/10.1145/2723576.2723628

Beishuizen, J., & Steffens, K. (2011). A conceptual framework for research on self-regulated learning. In *Self-Regulated Learning in Technology Enhanced Learning Environments* (pp. 3–19). Springer.

Bergmann, J., & Sams, A. (2015). *Flipped Learning for Math Instruction* (Vol. 2). International Society for Technology in Education.

Biggs, J. (1999). What the student does: Teaching for enhanced learning. *Higher Education Research & Development*, *18*(1), 57–75.

Biswas, G., Jeong, H., Kinnebrew, J. S., Sulcer, B., & Roscoe, R. O. D. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology Enhanced Learning*, *5*(02), 123–152.

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.

Boekaerts, M. (1999). Self-regulated learning: where we are today. *International Journal of Educational Research*, *31*, 445–457.

Boekaerts, M. (2002). The on-line motivation questionnaire: A self-report instrument to assess students' context sensitivity. *Advances in Motivation and Achievement*, *12*, 77–120.

Boekaerts, M., & Cascallar, E. (2006). How far have we moved toward the integration of theory and practice in self-regulation? *Educational Psychology Review*, *18*(3), 199–210.

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology*, *54*(2), 199–231.

Boud, D. (1995). *Enhancing learning through self-assessment*. Routledge.

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G., Ho, A. D., & Seaton, D. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment*, *8*(March 2012), 13–25. Retrieved from http://www.mendeley.com/catalog/studying-learning-worldwide-classroom-research-edxs-first-mooc/

Briggs, K. C., & Myers, I. B. (1977). *The Myers-Briggs Type Indicator: Form G*. Consulting Psychologists Press.

Brinton, C. G., Buccapatnam, S., Chiang, M., & Poor, H. V. (2016). Mining MOOC Clickstreams: Video-Watching Behavior vs. In-Video Quiz Performance. *IEEE Transactions on Signal Processing*, *64*(14), 3677–3692.

Broadbent, J. (2017). Comparing online and blended learner's self-regulated learning strategies and academic performance. *The Internet and Higher Education*, *33*, 24–32.

Broadbent, J., & Poon, W. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, *27*, 1–13. https://doi.org/10.1016/j.iheduc.2015.04.007

Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3).

Chang, M.-M. (2007). Enhancing web-based language learning through self-monitoring. *Journal of Computer Assisted Learning*, *23*(3), 187–196.

ChanLin, L.-J. (2012). Learning strategies in web-supported collaborative project. *Innovations in Education and Teaching International*, *49*(3), 319–331.

Chaves, M. A., & Córdoba, E. R. (2014). Deciphering event logs in SharePoint Server: A methodology based on process mining. In *2014 XL Latin American Computing Conference (CLEI)* (pp. 1–12).

Chen, G., Davis, D., Lin, J., Hauff, C., & Houben, G.-J. (2016). Beyond the MOOC platform: gaining insights about learners from the social web. In *Proceedings of the 8th ACM Conference on Web Science* (pp. 15–24).

Cheng, G., & Chau, J. (2013). Exploring the relationship between students' self-regulated learning ability and their ePortfolio achievement. *The Internet and Higher Education*, *17*, 9–15.

Cho, M.-H., & Shen, D. (2013). Self-regulation in online learning. *Distance Education*, *34*(3), 290–301.

Cho, M.-H., & Summers, J. (2012). Factor validity of the Motivated Strategies for Learning Questionnaire (MSLQ) in asynchronous online learning environments. *Journal of Interactive Learning Research*, *23*(1), 5–28.

Chuang, I., & Ho, A. D. (2016). HarvardX and MITx: Four Years of Open Online Courses--Fall 2012-Summer 2016.

Clow, D. (2013). MOOCs and the funnel of participation. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 185–189).

Cockrum, T. (2013). *Flipping your English class to reach all learners: Strategies and lesson plans*. Routledge.

Cooper, S., & Sahami, M. (2013). Reflections on Stanford's MOOCs. *Communications of the ACM*, *56*(2), 28–30.

Correa, T. (2010). The Participation Divide Among "Online Experts''': Experience, Skills and Psychological Factors as Predictors of College Students' Web Content Creation." *Journal of Computer-Mediated Communication*, *16*(1). https://doi.org/10.1111/j.1083-6101.2010.01532.x

Corrin, L., de Barba, P. G., & Bakharia, A. (2017). Using learning analytics to explore help-seeking learner profiles in MOOCs. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 424–428).

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

Daradoumis, T., Bassi, R., Xhafa, F., & Caballe, S. (2013). A Review on Massive E-Learning (MOOC) Design, Delivery and Assessment. *2013 Eighth International*

Conference on P2P, Parallel, Grid, Cloud and Internet Computing, 208–213. https://doi.org/10.1109/3PGCIC.2013.37

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

Davis, D., Seaton, D., Hauff, C., & Houben, G.-J. (2018). Toward large-scale learning design: categorizing course designs in service of supporting learning outcomes. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale* (p. 4).

Davis, D., Chen, G., Hauff, C., & Houben, G.-J. (2016). Gauging MOOC Learners' Adherence to the Designed Learning Path. *EDM*, *16*, 9th.

de Barba, P. G., Kennedy, G. E., & Ainley, M. D. (2016). The role of students' motivation and participation in predicting performance in a MOOC Motivation and participation in MOOCs. *Journal of Computer Assisted Learning*, (March 2015). https://doi.org/10.1111/jcal.12130

De Waard, I., Abajian, S., Gallagher, M. S., Hogue, R., Keskin, N., Koutropoulos, A., & Rodriguez, O. C. (2011). Using mLearning and MOOCs to understand chaos, emergence, and complexity in education. *The International Review of Research in Open and Distributed Learning*, *12*(7), 94–115.

Dietze, S., Siemens, G., Taibi, D., & Drachsler, H. (2016). Editorial: Datasets for learning analytics. *Journal of Learning Analytics*, *3*(2), 307–311.

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, *3*(3), 231–264.

Dunn, R. S., & Dunn, K. J. (1978). *Teaching students through their individual learning styles: A practical approach*. Prentice Hall.

Effeney, G., Carroll, A., & Bahr, N. (2013). Self-regulated learning: Key strategies and their sources in a sample of adolescent males1. *Australian Journal of Educational and Developmental Psychology*, *13*, 58–74.

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist*, *46*(1), 6–25.

Endedijk, M. D., Brekelmans, M., Sleegers, P., & Vermunt, J. D. (2016). Measuring students' self-regulated learning in professional education: bridging the gap between event and aptitude measurements. *Quality & Quantity*, *50*(5), 2141–2164.

Erdogan, T., & Senemoglu, N. (2016). Development and validation of a scale on self-regulation in learning (SSRL). *SpringerPlus*, *5*(1), 1686. https://doi.org/10.1186/s40064-016-3367-y

Mazur, E. (2012). Flipped classroom; teaching model gains an online community | Harvard John A. Paulson School of Engineering and Applied Sciences. Retrieved April 30, 2017, from https://www.seas.harvard.edu/news/2012/02/flipped-classroom-teaching-model-gains-online-community

Evans, B. J., Baker, R. B., & Dee, T. S. (2016). Persistence Patterns in Massive Open Online Courses (MOOCs). *Journal of Higher Education*, *87*(2), 206–242. https://doi.org/10.1353/jhe.2016.0006

Felder, R. M., & Silverman, L. K. (1988). Learning and teaching styles in engineering education. *Engineering Education*, *78*(7), 674–681.

Ferguson, R., & Clow, D. (2015). Consistent Commitment: Patterns of Engagement across Time in Massive Open Online Courses (MOOCs), *2*, 55–80. https://doi.org/10.18608/jla.2015.23.5

Ferguson, R., Clow, D., Beale, R., Cooper, A. J., Morris, N., Bayne, S., & Woodgate, A. (2015). Moving through MOOCS: pedagogy, learning design and patterns of engagement. In *Design for teaching and learning in a networked world* (pp. 70–84). Springer.

Fincham, O. E., Gasevic, D. V, Jovanovic, J. M., & Pardo, A. (2018). From Study Tactics to Learning Strategies: An Analytical Method for Extracting Interpretable Representations. *IEEE Transactions on Learning Technologies*.

Floden, R. E. (1981). Chapter 2: The Logic of Information-Processing Psychology in Education. *Review of Research in Education*, *9*(1), 75–109.

Fontana, R. P., Milligan, C., Littlejohn, A., & Margaryan, A. (2015). Measuring self-regulated learning in the workplace. *International Journal of Training and Development*, *19*(1), 32–52.

Franklin, S. B., Gibson, D. J., Robertson, P. A., Pohlmann, J. T., & Fralish, J. S. (1995). Parallel analysis: a method for determining significant principal components. *Journal of Vegetation Science*, *6*(1), 99–106.

Freitas, S. I., Morgan, J., & Gibson, D. (2015). Will MOOCs transform learning and teaching in higher education? Engagement and course retention in online learning provision. *British Journal of Educational Technology*, *46*(3), 455–471.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Springer s). Berlin, Germany: Springer.

Garavalia, L. S., & Gredler, M. E. (2002). Prior achievement, aptitude, and use of learning strategies as predictors of college student achievement. *College Student Journal*, *36*(4), 616.

García-Peñalvo, F. J., Cruz-Benito, J., Borrás-Gené, O., & Blanco, Á. F. (2015). Evolution of the Conversation and Knowledge Acquisition in Social Networks related to a MOOC Course. In *Learning and collaboration technologies* (pp. 470–481). Springer.

García Cué, J., Santizo Rincón, J., & Alonso García, C. (2009). Instrumentos De Medicion De Estilos De Aprendizaje. *Revista Estilos de Aprendizaje*, *4*(4), 1–23. http: //www.uned.es/revistaestilosdeaprendizaje/numero_4/Artigos/lsr_4_articulo_1.pdf

Gardner, J., Brooks, C., Andres, J. M. L., & Baker, R. (2018). Replicating MOOC predictive models at scale. In *L@ S* (p. 1).

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84.

Gasevic, D., Kovanovic, V., Joksimovic, S., & Siemens, G. (2014). Where is research on massive open online courses headed? A data analysis of the MOOC Research Initiative. *The International Review Of Research In Open And Distributed Learning*, *15*(5).

George, D., & Mallery, P. (2003). SPSS for Windows Step by Step: Answers to Selected Exercises. *A Simple Guide and Reference*, 63. https://doi.org/9780335262588

Grainger, B. (2013). Massive open online course (MOOC) report 2013. *University of London*.

Günther, C. W., & Aalst, W. M. P. Van Der. (2007). Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics. *Business Process Management - Lecture Notes in Computer Science*, *4714*, 328–343. https://doi.org/10.1007/978-3-540-75183-0

Günther, C. W., & Rozinat, A. (2012). Disco: Discover Your Processes. *BPM (Demos)*, *940*, 40–44.

Guo, P. J., & Reinecke, K. (2014). Demographic differences in how students navigate through MOOCs. *Proceedings of the First ACM Conference on Learning @ Scale Conference - L@S '14*, 21–30. https://doi.org/10.1145/2556325.2566247

Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. *Handbook of Self-Regulation of Learning and Performance*, *30*, 65–84.

Hadwin, A. F., & Winne, P. H. (2012). Promoting learning skills in undergraduate students. *Enhancing the Quality of Learning: Dispositions, Instruction, and Mental Structures*, 201–227.

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, *2*(2–3), 107–124. https://doi.org/10.1007/s11409-007-9016-7

Hadwin, A., Winne, P., Stockley, D., Nesbit, J., & Woszczyna, C. (2001). Context moderates students' self-reports about how they study. *Journal of Educational Psychology*, *93*(3), 477.

Hamdan, N., McKnight, P., McKnight, K., & Arfstrom, K. M. (2013). *The flipped learning model: A white paper based on the literature review titled a review of flipped learning*.

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

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

Hodges, C. B., & Kim, C. (2010). Email, self-regulation, self-efficacy, and achievement in a college online mathematics course. *Journal of Educational Computing Research*, *43*(2), 207–223.

Honey, P., & Mumford, A. (1986). *The Manual of Learning Styles*. Maidenhead, Berkshire: Peter Honey Publications.

Hood, N., Littlejohn, A., & Milligan, C. (2015). Context counts: How learners' contexts influence learning in a MOOC. *Computers and Education*, *91*, 83–91. https://doi.org/10.1016/j.compedu.2015.10.019

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Articles*, 2.

Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Sample Size, and Advanced Topics*.

Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Educational Technology & Society*, *15*(1), 38–52.

Jaggars, S. S. (2014). Democratization of Education for Whom? Online Learning and Educational Equity. *Association of American Colleges and Universities. Accessed October*, *9*.

Jakešová, J., & Kalenda, J. (2015). Self-regulated learning: Critical-realistic conceptualization. *Procedia-Social and Behavioral Sciences*, *171*, 178–189.

Jansen, R. S., van Leeuwen, A., Janssen, J., Kester, L., & Kalz, M. (2016). Validation of the self-regulated online learning questionnaire. *Journal of Computing in Higher Education*, 1–22. https://doi.org/10.1007/s12528-016-9125-x

Jivet, I. (2016). *The Learning Tracker. A Learner Dashboard that Encourages Self-Regulation in MOOC Learners*. TU Delft. Retrieved from http://repository.tudelft.nl/

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, *101*(3), 621.

Johnson, R. D., Gueutal, H., & Falbe, C. M. (2009). Technology, trainees, metacognitive activity and e-learning effectiveness. *Journal of Managerial Psychology*, *24*(6), 545–566.

Joksimović, S., Kovanović, V., Jovanović, J., Zouaq, A., Gašević, D., & Hatala, M. (2015). What do cMOOC participants talk about in social media?: a topic analysis of discourse in a cMOOC. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (pp. 156–165).

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).

Jovanović, J., Gašević, D., Dawson, S., Pardo, A., Mirriahi, N., & others. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education*, *33*, 74–85.

Juhanák, L., Zounek, J., & Rohliková, L. (2017). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Computers in Human Behavior*.

Karabenick, S. A., Woolley, M. E., Friedel, J. M., Ammon, B. V, Blazevski, J., Bonney, C. R., … others. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist*, *42*(3), 139–151.

Kay, J., Reimann, P., Diebold, E., & Kummerfeld, B. (2013). MOOCs: So Many Learners, So Much Potential... *IEEE Intelligent Systems*, (3), 70–77.

Keefe, J. (1988). *Profiling and Utilizing Learning Style*. Natl Assn of Secondary School.

Kenny, D. A. (2014). Measuring model fit.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, *10*(3), 333–351.

Kim, C., & Hodges, C. B. (2012). Effects of an emotion control treatment on academic emotions, motivation and achievement in an online mathematics course. *Instructional Science*, *40*(1), 173–192.

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, *33*(2004), 1–26.

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

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

Kizilcec, R. F., & Brooks, C. (2016). Diverse big data and randomized field experiments in MOOCs. *Handbook of Learning Analytics*, 211–222.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2016). Recommending Self-Regulated Learning Strategies Does Not Improve Performance in a MOOC. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale* (pp. 101–104). New York, NY, USA: ACM. https://doi.org/10.1145/2876034.2893378

Kizilcec, R. F., & Halawa, S. (2015). Attrition and Achievement Gaps in Online Learning. In *Proceedings of the Second ACM Conference on Learning@Scale* (pp. 57–66).

Kizilcec, R., & Schneider, E. (2015). Motivation as a Lens to Understand Online Learners: Toward Data-Driven Design with the OLEI Scale. *Transactions on Computer-Human Interactions (TOCHI)*, *22*(2), 24.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 170–179).

Klingsieck, K. B., Fries, S., Horz, C., & Hofer, M. (2012). Procrastination in a distance university setting. *Distance Education*, *33*(3), 295–310.

Delgado-Kloos, C., Muñoz-Merino, P. J., Alario-Hoyos, C., Ayres, I. E., & Fernández-Panadero, C. (2015). Mixing and blending MOOC Technologies with face-to-face pedagogies. In *2015 IEEE Global Engineering Education Conference (EDUCON)* (pp. 967–971).

Kocdar, S., Karadeniz, A., Bozkurt, A., & Buyuk, K. (2018). Measuring Self-Regulation in Self-Paced Open and Distance Learning Environments. *The International Review of Research in Open and Distributed Learning*, *19*(1).

Köck, M., & Paramythis, A. (2011). Activity sequence modelling and dynamic clustering for personalized e-learning. *User Modeling and User-Adapted Interaction*, *21*(1–2), 51–97.

Kolb, A. Y., & Kolb, D. a. (2005). The Kolb Learning Style Inventory — Version 3 . 1 2005 Technical Specifi cations. *LSI Technical Manual*, 1–72. https://doi.org/10.1016/S0260-6917(95)80103-0

Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., & Hatala, M. (2015). Penetrating the black box of time-on-task estimation. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*, (October), 184–193. https://doi.org/10.1145/2723576.2723623

Lajoie, S., & Azevedo, R. (2006). Teaching and learning in technology-rich environments. *Handbook of Educational Psychology*, *2*, 803–821.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.

Laplante, P. A. (2013). Courses for the Masses? *IT Professional*, (2), 57–59.

Laurillard, D. (2013). *Teaching as a design science: Building pedagogical patterns for learning and technology*. Routledge.

Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in

EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research & Evaluation*, *12*(2), 1–11.

Lee, T.-H., Shen, P.-D., & Tsai, C.-W. (2008). Applying Web-Enabled Problem-Based Learning and Self-Regulated Learning to Add Value to Computing Education in Taiwan’s Vocational Schools. *Educational Technology & Society*, *11*(3), 13–25.

Li, N., Kidzinski, L., Jermann, P., & Dillenbourg, P. (2015). How Do In-video Interactions Reflect Perceived Video Difficulty? In *Proceedings of the European MOOCs Stakeholder Summit 2015* (pp. 112–121).

Liang, J.-C., & Tsai, C.-C. (2008). Internet Self-Efficacy and Preferences Toward Constructivist Internet-based Learning Environments: A Study of Pre-school Teachers in Taiwan. *Educational Technology & Society*, *11*(1), 226–237.

Liou, P.-Y., & Kuo, P.-J. (2014). Validation of an instrument to measure students’ motivation and self-regulation towards technology learning. *Research in Science & Technological Education*, *32*(2), 79–96. https://doi.org/10.1080/02635143.2014.893235

Little, R. J. A. (1988). Missing-Data Adjustments in Large Surveys. *Journal of Business & Economic Statistics*, *6*(3), 287–296. https://doi.org/10.2307/1391881

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

Littlejohn, A., & Milligan, C. (2015). Designing MOOCs for profesional learners: Tools and patterns to encourage self_regulated learning. *Design Paper*, *42*(June), 1–10. Retrieved from http://www.openeducationeuropa.eu/en/article/Design-Patterns-for-Open-Online-Teaching-and-Learning-Design-Paper-42-4

Liu, Z., He, J., Xue, Y., Huang, Z., Li, M., & Du, Z. (2015). Modeling the Learning Behaviors of Massive Open Online Courses. *Big Data (Big Data), 2015 IEEE*

International Conference On, 2883–2885. https://doi.org/10.1109/BigData.2015.7364110

Lodge, J. M., & Corrin, L. (2017). What data and analytics can and do say about effective learning. *NPJ Science of Learning*, *2*(1), 5.

Lorenzo-Seva, U. (1999). Promin: A method for oblique factor rotation. *Multivariate Behavioral Research*, *34*(3), 347–365.

Lorenzo-Seva, U. (2003). A factor simplicity index. *Psychometrika*, *68*(1), 49–60.

Lynch, R., & Dembo, M. (2004). The relationship between self-regulation and online learning in a blended learning context. *The International Review of Research in Open and Distributed Learning*, *5*(2).

Magno, C. (2011). Validating the Academic Self-regulated Learning Scale with the Motivated Stategies for Learning Questionnaire (MSLQ) and Learning and Study Strategies Inventory (LASSI). *The International Journal of Educational and Psychological Assessment*, *7*(2), 56–73. https://doi.org/10.1037/t09161-000.

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.

Maldonado, J., Carvallo, J. P., & Siguencia, J. (2015). Metodologías y Propuestas Metodológicas Para el Diseño de Objetos de Aprendizaje: Un Estado del Arte en Iberoamérica. *Conferencias LACLO*, *6*(1), 36.

Maldonado, J. J., Bermeo, J. L., & Pacheco, G. (2016). Assessing a Methodological Proposal for the Design, Creation and Evaluation of Learning Objects Oriented to Educators with Diverse Educational and Technological Competencies. *CLEI Electronic Journal*, *19*(1), 3.

Maldonado, J. J., Fernandez-Pampillon, A. M., & Sanz, C. V. (2015). Analysis framework for tailored selection of learning objects methodologies. In *Collaboration Technologies and Systems (CTS), 2015 International Conference on* (pp. 148–158).

Maldonado, J. J., Palta, R., Vázquez, J., Bermeo, J. L., Pérez-Sanagustín, M., & Munoz-Gama, J. (2016). Exploring differences in how learners navigate in MOOCs based on self-regulated learning and learning styles: A process mining approach. In *Computing Conference (CLEI), 2016 XLII Latin American* (pp. 1–12).

Margaryan, A., Bianco, M., & Littlejohn, A. (2015). Instructional quality of massive open online courses (MOOCs). *Computers & Education*, *80*, 77–83.

Martinez-Lopez, R., Yot, C., Tuovila, I., & Perera-Rodríguez, V.-H. (2017). Online Self-Regulated Learning Questionnaire in a Russian MOOC. *Computers in Human Behavior*.

Meyer, J. H. F., & Muller, M. W. (1990). Evaluating the quality of student learning. I—An unfolding analysis of the association between perceptions of learning context and approaches to studying at an individual level. *Studies in Higher Education*, *15*(2), 131–154.

Mezaour, A.-D. (2005). Filtering web documents for a thematic warehouse case study: eDot a food risk data warehouse (extended). In *Intelligent Information Processing and Web Mining* (pp. 269–278). Springer.

Michinov, N., Brunot, S., Le Bohec, O., Juhel, J., & Delaval, M. (2011). Procrastination, participation, and performance in online learning environments. *Computers & Education*, *56*(1), 243–252.

Montalvo, F. T., & Torres, M. C. G. (2004). Self-Regulated Learning : Current and Future Directions. *Electronic Journal of Research in Educational Psychology*, *1*, 1–34.

Moreno-Marcos, P. M., Muñoz-Merino, P. J., Alario-Hoyos, C., Estévez-Ayres, I., & Kloos, C. D. (2018). Analysing the predictive power for anticipating assignment grades in a massive open online course. *Behaviour & Information Technology*, *0*(0), 1–16. https://doi.org/10.1080/0144929X.2018.1458904

Mukala, P., Buijs, J., & Van Der Aalst, W. M. P. (2015). Exploring students' learning behaviour in moocs using process mining techniques. *Eindhoven. BPM Center Report BPM-15-10. BPMCenter. Org, Eindhoven Google Scholar*.

Mukala, P., Buys, J., Leemans, M., & van der Aalst, W. M. . (2015). Learning Analytics on Coursera Event Data: A Process Mining Approach. *BPM*, 18–32. Retrieved from http://ceur-ws.org/Vol-1527/paper2.pdf

Muñoz-Merino, P. J., Ruipérez-Valiente, J. A., Moreno, J. L. S., & Kloos, C. D. (2015). Assessment Activities in Massive Open On-Line Courses. *Furthering Higher Education Possibilities through Massive Open Online Courses*, 165.

Nawrot, I., & Doucet, A. (2014). Building engagement for MOOC students: introducing support for time management on online learning platforms. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion* (pp. 1077–1082).

Nesterko, S. O., Seaton, D. T., Kashin, K., Han, Q., Reich, J., Waldo, J., … Ho, A. D. (2014). Education Levels Composition (HarvardX Insights).

Niemi, H., Nevgi, A., & Virtanen, P. (2003). Towards self-regulation in web-based learning. *Journal of Educational Media*, *28*(1), 49–71. https://doi.org/10.1080/1358165032000156437

O'Flaherty, J., & Phillips, C. (2015). The use of flipped classrooms in higher education: A scoping review. *The Internet and Higher Education*, *25*, 85–95.

Panadero, E. (2017). A Review of Self-regulated Learning: Six Models and Four Directions for Research. *Frontiers in Psychology*, *8*.

Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: when measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research*, *60*(6), 723–735.

Panadero, E., & Alonso-Tapia, J. (2015). Teorías de autorregulación educativa: una comparación y reflexión teórica. *Psicología Educativa*, *20*(1), 11–22. https://doi.org/10.1016/j.pse.2015.08.005

Pardo, A., Mirriahi, N., Martinez-Maldonado, R., Jovanovic, J., Dawson, S., & Gašević, D. (2016). Generating actionable predictive models of academic performance. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 474–478).

Pashler, H., & Wagenmakers, E.--J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*(6), 528–530.

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaiane, O. R. (2009). Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(6), 759–772.

Pérez-Álvarez, R., Maldonado-Mahauad, J., & Pérez-Sanagustín, M. (2018). Design of a tool to support self-regulated learning strategies in MOOCs. *Journal of Universal Computer Science*, *24*.

Pérez-Sanagustín, M., Hilliger, I., Alario-Hoyos, C., Delgado-Kloos, C., & Rayyan, S. (2017). H-MOOC framework: reusing MOOCs for hybrid education. *Journal of Computing in Higher Education*, *29*(1), 47–64.

Pérez-Sanagustín, M., Hilliger, I., Alario-Hoyos, C., Delgado-Kloos, C., Rayyan, S. (2016). Describing MOOC-based Hybrid initiatives: The H-MOOC Framework. In *E-moocs* (p. In Press).

Pérez-Sanagustín, M., Maldonado, J. J., & Morales, N. (2016). *WPD 1.1 Estado del arte de adopción de MOOCs en la Educación Superior en América Latina y Europa,* *Proyecto Europeo Erasmus + MOOC-Maker Construction of Management Capacities of MOOCs in Higher Education (561533-EPP-1-2015-1-ES-EPPKA2-CBHE-JP)*.

Perna, L. W., Ruby, A., Boruch, R. F., Wang, N., Scull, J., & Evans, C. (2014). Moving Through MOOCs: Understanding the Progression of Users in Massive Open Online Courses. *Educational Researcher*, (December), 421–432. https://doi.org/10.3102/0013189X14562423

Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, *21*(2), 381–391.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, *16*(4), 385–407. https://doi.org/10.1007/s10648-004-0006-x

Pintrich, P. R. (2000). The Role of Goal Orientation in Self-Regulated Learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 451–502). San Diego: Academic Press. https://doi.org/http://dx.doi.org/10.1016/B978-012109890-2/50043-3

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, *31*(6), 459–470.

Pintrich, P., Smith, D., Garcia, T., & Mckeaechie, W. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*, 801–813.

Pintrich, P. R. R., Smith, D., Garcia, T., & McKeachie, W. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ). *Ann Arbor. Michigan*, *48109*(August 2016), 1259. https://doi.org/ED338122

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33.

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

Puzziferro, M. (2008). Online technologies self-efficacy and self-regulated learning as predictors of final grade and satisfaction in college-level online courses. *The Amer. Jrnl. of Distance Education*, *22*(2), 72–89.

Reich, J. (2015). Rebooting MOOC research. *Science*, *347*(6217), 34–35.

Reich, J. (2014). MOOC Completion and Retention in the Context of Student Intent. Retrieved from http://er.educause.edu/articles/2014/12/mooc-completion-and-retention-in-the-context-of-student-intent

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.

Rigotti, T., Schyns, B., & Mohr, G. (2008). A short version of the occupational self-efficacy scale: Structural and construct validity across five countries. *Journal of Career Assessment*, *16*(2), 238–255.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, *130*(2), 261–88. https://doi.org/10.1037/0033-2909.130.2.261

Roediger III, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27.

Rojas, E., Munoz-Gama, J., Sepúlveda, M., & Capurro, D. (2016). Process Mining in Healthcare: A literature review. *Journal of Biomedical Informatics*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1532046416300296

Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, *2*, 7–12.

Romero, C., Cerezo, R., Bogarín, A., & Sánchez-Santillán, M. (2016). Educational process mining: a tutorial and case study using moodle data sets. *Data Mining and Learning Analytics: Applications in Educational Research*, 1.

Roth, A., Ogrin, S., & Schmitz, B. (2015). Assessing self-regulated learning in higher education: a systematic literature review of self-report instruments. *Educational Assessment, Evaluation and Accountability*, 1–26.

Saint, J., Gašević, D., & Pardo, A. (2018). Detecting Learning Strategies Through Process Mining. In *European Conference on Technology Enhanced Learning* (pp. 385–398).

Salehi, M., & Jafari, H. (2015). Development and Validation of an EFL Self-Regulated Learning Questionnaire. *Southern African Linguistics and Applied Language Studies*, *33*(1), 63–79. https://doi.org/10.2989/16073614.2015.1023503

Sampieri, H. (2008). *Mendoza.(2008). Metodologia de investigación. México. Mc. Gran Hill*.

Samruayruen, B., Enriquez, J., Natakuatoong, O., & Samruayruen, K. (2013). Self-regulated learning: A key of a successful learner in online learning environments in Thailand. *Journal of Educational Computing Research*, *48*(1), 45–69.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, *19*(4), 460–475.

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

Seaton, D. T., Bergner, Y., Chuang, I., Mitros, P., & Pritchard, D. E. (2014). Who does what in a massive open online course? *Communications of the ACM*, *57*(4), 58–65. https://doi.org/10.1145/2500876

Sengel, E. (2016). To FLIP or not to FLIP: Comparative case study in higher education in Turkey. *Computers in Human Behavior*, *64*, 547–555.

Shah, D. (2018). By the numbers: MOOCS in 2017 (2018).

Shah, D. (2015, December). MOOCs in 2015: Breaking Down the Numbers. *EdSurge*. Retrieved from https://www.edsurge.com/news/2015-12-28-moocs-in-2015-breaking-down-the-numbers

Sharples, M., Kloos, C. D., Dimitriadis, Y., Garlatti, S., & Specht, M. (2015). Mobile and Accessible Learning for MOOCs. *Journal of Interactive Media in Education*, *2015*(1).

Siadaty, M., Gašević, D., & Hatala, M. (2016). Associations between technological scaffolding and micro-level processes of self-regulated learning: A workplace study. *Computers in Human Behavior*, *55*, 1007–1019. https://doi.org/10.1016/j.chb.2015.10.035

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. *ArXiv Preprint ArXiv:1407.7131*.

Snow, R. E. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning.

Soffer, T., & Cohen, A. (2015). Implementation of Tel Aviv University MOOCs in academic curriculum: A pilot study. *The International Review of Research in Open and Distributed Learning*, *16*(1).

Sonnenberg, C., & Bannert, M. (2015). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, *2*(1), 72–100.

Stanley, D. J., & Spence, J. R. (2014). Expectations for replications: Are yours realistic? *Perspectives on Psychological Science*, *9*(3), 305–318.

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.

Sunar, A. S., Abbasi, R. A., Davis, H. C., White, S., & Aljohani, N. R. (2018). Modelling MOOC learners' social behaviours. *Computers in Human Behavior*. https://doi.org/https://doi.org/10.1016/j.chb.2018.12.013

Tallent-Runnels, M. K., Thomas, J. A., Lan, W. Y., Cooper, S., Ahern, T. C., Shaw, S. M., & Liu, X. (2006). Teaching courses online: A review of the research. *Review of Educational Research*, *76*(1), 93–135.

Taub, M., Azevedo, R., Bouchet, F., & Khosravifar, B. (2014). Can the use of cognitive and metacognitive self-regulated learning strategies be predicted by learners' levels of prior knowledge in hypermedia-learning environments? *Computers in Human Behavior*, *39*, 356–367. https://doi.org/10.1016/j.chb.2014.07.018

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *58*(1), 267–288.

Trevors, G., Feyzi-Behnagh, R., Azevedo, R., & Bouchet, F. (2016). Self-regulated learning processes vary as a function of epistemic beliefs and contexts: Mixed method evidence from eye tracking and concurrent and retrospective reports. *Learning and Instruction*, *42*, 31–46.

Trigwell, K., & Prosser, M. (1991). Improving the quality of student learning: the influence of learning context and student approaches to learning on learning outcomes. *Higher Education*, *22*(3), 251–266.

Tsai, C.-C., Chuang, S.-C., Liang, J.-C., & Tsai, M.-J. (2011). Self-efficacy in Internet-based Learning Environments: A Literature Review. *Educational Technology & Society*, *14*(4), 222–240.

Tsai, C.-C. (2009). Conceptions of learning versus conceptions of web-based learning: The differences revealed by college students. *Computers & Education*, *53*(4), 1092–1103.

Valle, A., Núñez, J. C., Cabanach, R. G., González-Pienda, J. A., Rodríguez, S., Rosário, P., … Muñoz-Cadavid, M. A. (2008). Self-regulated profiles and academic achievement. *Psicothema*, *20*(4), 724–731.

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. *International Review of Research in Open and Distributed Learning*, *19*(5).

van den Boom, G., Paas, F., & van Merriënboer, J. J. G. (2007). Effects of elicited reflections combined with tutor or peer feedback on self-regulated learning and learning outcomes. *Learning and Instruction*, *17*(5), 532–548.

van Der Aalst, W. (2011). *Process mining: discovery, conformance and enhancement of business processes*. Springer Science & Business Media.

van Der Aalst, W. (2016). *Process mining: discovery, conformance and enhancement of business processes* (Second Edi). Springer Science & Business Media.

van Der Aalst, W. (2016). *Process mining: data science in action*. Springer.

van Der Linden, D., Sonnentag, S., Frese, M., & Van Dyck, C. (2001). Exploration strategies, performance, and error consequences when learning a complex computer task. *Behaviour & Information Technology*, *20*(3), 189–198.

van Eck, M., Lu, X., Leemans, S., & van der Aalst, W. (2015). PM^ 2: A Process Mining Project Methodology. In *Advanced Information Systems Engineering* (pp. 297–313).

Veletsianos, G., Reich, J., & Pasquini, L. A. (2016). The life between big data log events: learners' strategies to overcome challenges in MOOCs. *AERA Open*, *2*(3), 1–10. https://doi.org/10.1177/2332858416657002

Vermunt, J. D. (1998). The regulation of constructive learning processes. *British Journal* *of Educational Psychology*, *68*(2), 149–171.

Wang, C.-H., Shannon, D. M., & Ross, M. E. (2013). Students' characteristics, self-regulated learning, technology self-efficacy, and course outcomes in online learning. *Distance Education*, *34*(3), 302–323.

Wang, S.-L., & Wu, P.-Y. (2008). The role of feedback and self-efficacy on web-based learning: The social cognitive perspective. *Computers & Education*, *51*(4), 1589–1598.

Wang, Y., & Baker, R. (2015). Content or platform: Why do students complete MOOCs? *MERLOT Journal of Online Learning and Teaching*, *11*(1), 191–218.

Warr, P., & Downing, J. (2000). Learning strategies, learning anxiety and knowledge acquisition. *British Journal of Psychology*, *91*(3), 311–333.

Weinstein, C. E., Acee, T. W., & Jung, J. (2011). Self-Regulation and Learning Strategies. *New Directions for Teaching and Learning*, *2011*(126), 45–53. https://doi.org/10.1002/tl.443

Weinstein, C. E., & Palmer, D. R. (2002). *LASSI User's Manual: For Those Administering the Learning and Study Strategies Inventory*. H & H Pub.

Winne, P. H. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition and Learning*, *9*(2), 229–237. https://doi.org/10.1007/s11409-014-9113-3

Winne, P. H., & Baker, R. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *JEDM-Journal of Educational Data Mining*, *5*(1), 1–8.

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

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

Winne, P. H., & Nesbit, J. C. (2010). The psychology of academic achievement. *Annual Review of Psychology*, *61*, 653–678.

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist*, *45*(4), 267–276.

Winne, P. H. (2006). How software technologies can improve research on learning and bolster school reform. *Educational Psychologist*, *41*(1), 5–17.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. *Metacognition in Educational Theory and Practice*, *93*, 27–30.

Wirth, J., & Leutner, D. (2008). Self-regulated learning as a competence: Implications of theoretical models for assessment methods. *Zeitschrift Für Psychologie/Journal of Psychology*, *216*(2), 102–110.

Wong, L.-H., & Looi, C.-K. (2012). Enculturing self-directed seamless learners: Towards a facilitated seamless learning process framework mediated by mobile technology. In *Wireless, Mobile and Ubiquitous Technology in Education (WMUTE), 2012 IEEE Seventh International Conference on* (pp. 1–8).

Xu, B., & Yang, D. (2016). Motivation classification and grade prediction for MOOCs learners. *Computational Intelligence and Neuroscience*, *2016*, 4.

You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *Internet and Higher Education*, *29*. https://doi.org/10.1016/j.iheduc.2015.11.003

Yukselturk, E., & Top, E. (2013). Exploring the link among entry characteristics, participation behaviors and course outcomes of online learners: An examination of learner profile using cluster analysis. *British Journal of Educational Technology*, *44*(5), 716–728. https://doi.org/10.1111/j.1467-8535.2012.01339.x

Zhao, C., Yang, J., Liang, J., & Li, C. (2016). Discover learning behavior patterns to predict certification. In *Computer Science & Education (ICCSE), 2016 11th International Conference on* (pp. 69–73).

Zheng, S., Rosson, M. B., Shih, P. C., & Carroll, J. M. (2015). Understanding Student Motivation, Behaviors and Perceptions in MOOCs. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*, 1882–1895. https://doi.org/10.1145/2675133.2675217

Zhenghao, C., Alcorn, B., Christensen, G., Eriksson, N., Koller, D., & Emanuel, E. J. (2015). Who's Benefiting from MOOCs, and Why. Retrieved October 13, 2015, from https://hbr.org/2015/09/whos-benefiting-from-moocs-and-why

Zhonggen, Y., & Guifang, W. (2016). Academic achievements and satisfaction of the clicker-aided flipped business English writing class. *Educational Technology & Society*, *19*(2), 298–313.

Zimmerman, B. (2015). *Self-Regulated Learning: Theories, Measures, and Outcomes*. *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier. Retrieved from http://www.sciencedirect.com/science/article/pii/B9780080970868260601

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166–183.

Zimmerman, B., & Kitsantas, A. (2007). Reliability and validity of self-efficacy for learning form (SELF) scores of college students. *Zeitschrift Für Psychologie, 215(3)*, 157–163.

Zimmerman, B., Boekarts, M., & Pintrich, P. (2000). A social cognitive perspective. In *of Self-Regulation* (Vol. 13, pp. 451–502). Academic Press San Diego, CA.

https://doi.org/10.1016/B978-012109890-2/50031-7

Zimmerman, B. J. (2000). Chapter 2 - Attaining Self-Regulation: A Social Cognitive Perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 13–39). San Diego: Academic Press. https://doi.org/https://doi.org/10.1016/B978-012109890-2/50031-7

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, *81*(3), 329.

Zimmerman, B. J., & Pons, M. M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal*, *23*(4), 614–628. https://doi.org/10.2307/1163093

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *67*(2), 301–320. https://doi.org/10.1038/203024b0

APPENDIX A – SRL QUESTIONNAIRE FOR MOOCS

Goal of the Questionnaire: establish learner self-regulation profile

[QUESTIONNAIRE]

In this section, answer the questions with a true assessment of yourself and not your personal ambitions. There are no right or wrong answers. Please choose one (1) response out of the following possible responses:

a) Not true at all for me
b) Not very true for me
c) Somewhat true for me
d) Quite true for me
e) Very true for me

[SRL STRATEGY] {SELF EFFICACY}
1. I feel that I am capable of studying and learning in a MOOC because I am confident in my skills
2. When I'm faced with a challenge, I am able to look at it from different angles to overcome it
3. I feel capable of learning everything that I'm presented with
4. My prior experience has prepared me to face new challenges posed by learning online through a MOOC
5. I am able to identify if the objectives of a MOOC are the same or at least similar to the objectives I had set for myself for this course
6. I feel prepared for the demands and requirements of this MOOC

[SRL STRATEGY] {GOAL SETTING}
7. When I study, I set performance standards for myself during the learning process
8. When I study, I set short-term goals (daily or weekly) or long-term goals (for the whole course) for myself
9. When I study, I set goals to help me manage my study time
10. When I study, I set realistic deadlines that help me achieve my learning goals
11. When I start a study session, I establish a fixed time period to try to reach my goals and do the best I can

[SRL STRATEGY] {STUDY ENVIRONMENT MANAGEMENT}
12. When I study, I choose a place that is conducive to learning and distraction-free
13. When I study, I choose a place that is comfortable
14. When I study, I try to isolate myself from noisy places

[SRL STRATEGY] {ORGANIZATION}
15. When I do the video lectures in a MOOC, I make outlines or summaries of the material to help me organize my ideas
16. When I learn using a MOOC, I review the video lectures and the notes I've taken in order to find the most important ideas
17. When I study, I create simple charts, maps, diagrams, or tables to help me organize the material I learn in the MOOC
18. When I learn using a MOOC, I review my notes and make an outline of the most important concepts

[SRL STRATEGY] {HELP SEEKING}
19. I write in the course forum when I need to ask for help with something
20. I look for help using external online materials when the course materials do not satisfy my concerns regarding the content
21. I review the assessments I have previously passed in order to find answers to questions I have about course content
22. I rewind or fast-forward videos in a MOOC to look for specific information on the course topics

APPENDIX B – PUBLICATIONS BY THE AUTHOR

A. Publication 1

Alonso-Mencía, M. E., Alario-Hoyos, C., Maldonado-Mahauad, J., Estévez-Ayres, I., Pérez-Sanagustín, M., & Delgado Kloos, C. (2019). Self-regulated learning in MOOCs: lessons learned from a literature review. *Educational Review*, 1-27.

EDUCATIONAL REVIEW
https://doi.org/10.1080/00131911.2019.1566208

Routledge
Taylor & Francis Group

Check for updates

Self-regulated learning in MOOCs: lessons learned from a literature review

M. Elena Alonso-Mencía[a], Carlos Alario-Hoyos[a], Jorge Maldonado-Mahauad[b,c], Iria Estévez-Ayres[a], Mar Pérez-Sanagustín[b,d] and Carlos Delgado Kloos[a]

[a]Department of Telematics Engineering, Universidad Carlos III de Madrid, Madrid, Spain; [b]Department of Computer Science, Pontificia Universidad Católica de Chile, Santiago, Chile; [c]Department of Computer Science, Universidad de Cuenca, Cuenca, Ecuador; [d]Institute de Recherche Informatique de Toulouse (IRIT), Université Paul Sabatier Toulouse III, Toulouse, France

ABSTRACT

Learners in massive open online courses (MOOCs) are required to be autonomous during their learning process, and thus they need to self-regulate their learning to achieve their goals. According to existing literature, self-regulated learning (SRL) research in MOOCs is still scarce. More studies which build on past works regarding SRL in MOOCs are required, as well as literature reviews that help to identify the main challenges and future research directions in relation to this area. In this paper, the authors present the results of a systematic literature review on SRL in MOOCs, covering all the related papers published until the end of 2017. The papers considered in this review include real experiences with at least a MOOC (other learning scenarios sometimes claimed as MOOCs, such as blended courses, or online courses with access restrictions, are out of the scope of this analysis). Most studies on SRL in MOOCs share some common features: they are generally exploratory, based on one single MOOC and tend not to specify in which SRL model they are grounded. The results reveal that high self-regulators engage in non-linear navigation and approach MOOCs as an informal learning opportunity. In general, they prefer setting specific goals based on knowledge development and control their learning through assignments.

ARTICLE HISTORY
Received 17 July 2018
Accepted 20 December 2018

KEYWORDS
Open education; lifelong learning; literature review; massive open online course; self-regulation; self-regulated learning

Introduction

Massive open online course (MOOC) environments, characterised by their massive and open nature, attract a diversity of learners who differ in learning paces and engagement levels according to their backgrounds and motivations. MOOCs demand learners to be autonomous and create their own learning path with little or no tutoring to help them (Min & Jingyan, 2017). In other words, learners in MOOCs need to be able to self-regulate their learning.

Self-regulated learning (SRL) has been defined differently depending on the theoretical models taken as a reference. However, according to a recent literature review on SRL by Panadero (2017), there is a general agreement in the community for defining SRL as a process that involves cognitive, metacognitive, behavioural, motivational and

CONTACT Carlos Alario-Hoyos calario@it.uc3m.es Department of Telematics Engineering, Universidad Carlos III de Madrid, Avda. Universidad 30, Leganés, Madrid 28911, Spain

B. Publication 2

Maldonado-Mahauad, J., Pérez-Sanagustín, M., & Beyle, C. (2019). A questionnaire for measuring Self-regulated Learning in Massive Open Online Courses based on a Systematic Review – Under Review.

Manuscript Details

Manuscript number	EDUREV_2018_92
Title	A Questionnaire for Measuring Self-Regulated Learning in Massive Open Online Courses Based on a Systematic Review.
Article type	Research paper

Abstract

The bibliography in the area of the Massive Open Online Courses (MOOCs) highlights the importance of students' self-regulation ability on their success in reaching their learning objectives. One of the most widely used tools for measuring self-regulated learning strategies (SRL) in online environments is the questionnaire. However, the questionnaires existing in the bibliography have some limitations when applied in MOOC learning environments: (1) they do not consider strategies that according to the bibliography are relevant in this type of courses and that are related to academic achievements; (2) they are too long regarding the number of questions and, as a consequence, the response rates are low, limiting the use that can later be given to the data; and (3) the statements of the instruments developed for online contexts are based on instruments developed in the 90's and they do not reflect the effect that the context has, specifically, the technological for MOOC courses. To overcome these shortcomings, this study presents the development of a questionnaire to measure SRL strategies in MOOCs - the MOOC-SRLQ. This instrument is built upon an analysis of all existing SRL questionnaires used in traditional face to face, online learning and a bibliographic review that analyzed which SRL strategies relate to successful learning outcomes in MOOCs. This questionnaire was validated through an exploratory and confirmatory factor analysis developed over a sample of 4627 (N = 4627) learners of 6 MOOCs. As a result, the final questionnaire consists of 22 items to assess 5 SRL strategies in MOOCs. Results indicate the MOOC-SRLQ is an acceptable instrument to measure self-regulation in MOOCs, as it considers all the relevant strategies that have been associated with effective learning outcomes in online learning environments.

Keywords	Self-regulated learning; Massive Open Online Course; Online learning; Questionnaire; Learning strategies
Taxonomy	Computer Science, e-Learning, Self-regulated Learning

Corresponding Author	Jorge Maldonado Mahauad
Corresponding Author's Institution	Universidad de Cuenca / Pontificia Universidad Católica de Chile
Order of Authors	Jorge Maldonado Mahauad, Mar Pérez-Sanagustín, Christian Beyle

Submission Files Included in this PDF

File Name [File Type]

Cover Letter.docx [Cover Letter]

Highlights.docx [Highlights]

Title page with author details.docx [Title Page (with Author Details)]

Manuscript without author details without tables and figures.docx [Manuscript (without Author Details)]

Acknowledgments.docx [Acknowledgement]

Figures.docx [Figure]

Tables.docx [Table]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

C. Publication 3

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.

Computers in Human Behavior 80 (2018) 179–196

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses

Jorge Maldonado-Mahauad [a, b, *], Mar Pérez-Sanagustín [a], René F. Kizilcec [c], Nicolás Morales [a], Jorge Munoz-Gama [a]

[a] Pontificia Universidad Católica de Chile, Department of Computer Science, Avda. Vicuña Mackenna 4860, Macul, Santiago, Chile
[b] Universidad de Cuenca, Department of Computer Science, Av. 12 Abril, Cuenca, Ecuador
[c] Stanford University, Graduate School of Education, 485 Lausen Mall, Stanford, CA 94305, USA

A R T I C L E I N F O

Article history:
Received 28 March 2017
Received in revised form
9 November 2017
Accepted 10 November 2017
Available online 13 November 2017

Keywords:
Self-regulated learning
Learning strategies
Process mining
Massive open online courses

A B S T R A C T

Big data in education offers unprecedented opportunities to support learners and advance research in the learning sciences. Analysis of observed behaviour using computational methods can uncover patterns that reflect theoretically established processes, such as those involved in self-regulated learning (SRL). This research addresses the question of how to integrate this bottom-up approach of mining behavioural patterns with the traditional top-down approach of using validated self-reporting instruments. Using process mining, we extracted interaction sequences from fine-grained behavioural traces for 3458 learners across three Massive Open Online Courses. We identified six distinct interaction sequence patterns. We matched each interaction sequence pattern with one or more theory-based SRL strategies and identified three clusters of learners. First, Comprehensive Learners, who follow the sequential structure of the course materials, which sets them up for gaining a deeper understanding of the content. Second, Targeting Learners, who strategically engage with specific course content that will help them pass the assessments. Third, Sampling Learners, who exhibit more erratic and less goal-oriented behaviour, report lower SRL, and underperform relative to both Comprehensive and Targeting Learners. Challenges that arise in the process of extracting theory-based patterns from observed behaviour are discussed, including analytic issues and limitations of available trace data from learning platforms.

1. Introduction

In recent years, masses of fine-grained educational records have become available to researchers and accelerated the nascent field of learning analytics (Dietze, Siemens, Taibi, & Drachsler, 2016). Digital learning platforms collect detailed records of each learner's behaviour, performance, and other types of interaction. In particular, Massive Open Online Courses (MOOCs) are a major source of data on learner behaviour and they enable research to gain a better understanding of how individuals learn in online learning environments (Breslow et al., 2013; Cooper & Sahami, 2013; Daradoumis, Bassi, Xhafa, & Caballe, 2013).

Nevertheless, despite the large amount of data that MOOCs are collecting, this information may not be sufficient to build on educational theories and develop new ones. In particular, access to critical information about learners' behaviour and learning processes is frequently limited (Lodge & Lewis, 2012). Data-driven methods can rapidly extract patterns in what learners do throughout a course, but it remains a challenge to interpret the patterns and understand how they relate to theory. One approach to increase the interpretability of large amounts of clickstream data is to triangulate with other data sources (i.e., taking a mixed-methods approach). For example, clickstream data from MOOCs, which capture learners' actual interactions, can be combined with data from self-report instruments such as questionnaires or think-aloud sessions (Bannert, Reimann, & Sonnenberg, 2014; Eynon, 2013), or data from external sources like eye-tracking (Trevors, Feyzi-Behnagh, Azevedo, & Bouchet, 2016). To get a better

* Corresponding author. Pontificia Universidad Católica de Chile, Department of Computer Science, Avda. Vicuña Mackenna 4860, Macul, Santiago, Chile.
E-mail addresses: jjmaldonado@uc.cl (J. Maldonado-Mahauad), mar.perez@uc.cl (M. Pérez-Sanagustín), kizilcec@stanford.edu (R.F. Kizilcec), nvmorale@uc.cl (N. Morales), jmun@uc.cl (J. Munoz-Gama).

https://doi.org/10.1016/j.chb.2017.11.011
0747-5632/

D. Publication 4

Maldonado-Mahauad, J., Davis, D., Alario-Hoyos, C., Delgado-Kloos, C., & Pérez-Sanagustín, M. (2018). Adaptation of a Process Mining Methodology to Analyse Learning Strategies in a Synchronous Massive Open Online Course – Under Review

Journal of Computing in Higher Education

Adaptation of a Process Mining Methodology to Analyse Learning Strategies in a Synchronous Massive Open Online Course

--Manuscript Draft--

Manuscript Number:		
Full Title:	Adaptation of a Process Mining Methodology to Analyse Learning Strategies in a Synchronous Massive Open Online Course	
Article Type:	Original Research	
Keywords:	Learning Analytics, Learning Behaviour, Learning Strategies, Process Mining, Massive Open Online Courses	
Corresponding Author:	Jorge Maldonado-Mahauad, Phd(c) Pontificia Universidad Catolica de Chile Cuenca, Azuay CHILE	
Corresponding Author Secondary Information:		
Corresponding Author's Institution:	Pontificia Universidad Catolica de Chile	
Corresponding Author's Secondary Institution:		
First Author:	Jorge Maldonado-Mahauad, Phd(c)	
First Author Secondary Information:		
Order of Authors:	Jorge Maldonado-Mahauad, Phd(c)	
	Dan Davis, Phd	
	Carlos Alario-Hoyos, Phd	
	Carlos Delgado-Kloos, Phd	
	Mar Pérez-Sanagustín, Phd	
Order of Authors Secondary Information:		
Funding Information:	Comisión Nacional de Investigación Científica y Tecnológica (CONICYT/DOCTORADO NACIONAL 2016 [grant No.21160081])	Phd(c) Jorge Maldonado-Mahauad
	Erasmus+ (586120-EPP-1-2017-1-ES-EPPKA2-CBHE-JP)	Not applicable
	Universidad de Cuenca (No. 007-2015)	Phd(c) Jorge Maldonado-Mahauad

E. Publication 5

Kizilcec, R. F., Pérez-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.

Computers & Education 104 (2017) 18–33

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses

René F. Kizilcec [a, *, 1], Mar Pérez-Sanagustín [b, 2], Jorge J. Maldonado [b, c, 2]

[a] *Department of Communication, Stanford University, USA*
[b] *Department of Computer Science, Pontificia Universidad Católica de Chile, Chile*
[c] *Department of Computer Science, Universidad de Cuenca, Ecuador*

A R T I C L E I N F O

Article history:
Received 9 June 2016
Received in revised form 31 August 2016
Accepted 1 October 2016
Available online 4 October 2016

Keywords:
Online learning
Learning analytics
Individual differences
Self-regulated learning
Massive open online course

A B S T R A C T

Individuals with strong self-regulated learning (SRL) skills, characterized by the ability to plan, manage and control their learning process, can learn faster and outperform those with weaker SRL skills. SRL is critical in learning environments that provide low levels of support and guidance, as is commonly the case in Massive Open Online Courses (MOOCs). Learners can be trained to engage in SRL and actively supported with prompts and activities. However, effective implementation of learner support systems in MOOCs requires an understanding of which SRL strategies are most effective and how these strategies manifest in online behavior. Moreover, identifying learner characteristics that are predictive of weaker SRL skills can advance efforts to provide targeted support without obtrusive survey instruments. We investigated SRL in a sample of 4,831 learners across six MOOCs based on individual records of overall course achievement, interactions with course content, and survey responses. Results indicated that goal setting and strategic planning predicted attainment of personal course goals, while help seeking appeared to be counterproductive. Learners with stronger SRL skills were more likely to revisit previously studied course materials, especially course assessments. Several learner characteristics, including demographics and motivation, predicted learners' SRL skills. We discuss implications for theory and the development of learning environments that provide adaptive support.

1. Introduction

A primary goal of Massive Open Online Courses (MOOCs) is to provide more people with opportunities for personal and intellectual growth. Between late 2011 and 2015, 550 institutions created 4,200 courses that reached over 35 million people worldwide, according to data collected by Class Central (Shah, 2015). Most learners who enroll in MOOCs selectively engage with parts of the course content and a small proportion eventually completes the course (Anderson, Huttenlocher, Kleinberg, & Leskovec, 2014; Breslow et al., 2013; Evans, Baker, & Dee, 2016; Ho et al., 2015; Kizilcec, Piech, & Schneider, 2013; Perna et al., 2014; Seaton, Bergner, Chuang, Mitros, & Pritchard, 2014). This variation in behavior can be partly attributed to the

* Corresponding author.
E-mail addresses: kizilcec@stanford.edu (R.F. Kizilcec), mar.perez@ing.puc.cl (M. Pérez-Sanagustín), jjmaldonado@uc.cl (J.J. Maldonado).
[1] Present address: Department of Communication, 450 Serra Mall, Stanford University, Stanford, CA 94305, USA.
[2] Present address: Department of Computer Science, Pontificia Universidad Católica de Chile, Avda. Vicuña Mackenna 4860, Macul, Santiago, Chile.

http://dx.doi.org/10.1016/j.compedu.2016.10.001
0360-1315/

F. Publication 6

Pérez-Álvarez, R., Maldonado-Mahauad, J., & Pérez-Sanagustín, M. (2018). Design of a Tool to Support Self-Regulated Learning Strategies in MOOCs. *Journal of Universal Computer Science*, *24*(8), 1090-1109.

Design of a tool to support self-regulated learning strategies in MOOCs

Ronald Pérez-Álvarez
(Pontificia Universidad Católica, Santiago de Chile, Chile
Universidad de Costa Rica, Puntarenas, Costa Rica
raperez13@uc.cl)

Jorge Maldonado-Mahauad
(Pontificia Universidad Católica, Santiago de Chile, Chile
Universidad de Cuenca, Cuenca, Ecuador
jjmaldonado@uc.cl)

Mar Pérez-Sanagustín
(Pontificia Universidad Católica, Santiago de Chile, Chile
Université Tolouse III Paul Sabatier, Tolouse, France
mar.perez@uc.cl; mar.perez-sanagustin@irit.fr)

Abstract: The massive and open nature of MOOCs contribute to attracting a great diversity of learners. However, the learners who enroll in these types of courses have trouble achieving their course objectives. One reason for this is that they do not adequately self-regulate their learning. In this context, there are few tools to support these strategies in online learning environment. Also, the lack of metrics to evaluate the impact of the proposed tools makes it difficult to identify the key features of this type of tools. In this paper, we present the process for designing *NoteMyProgress*, a web application that complements a MOOC platform and supports self-regulated learning strategies. For designing *NoteMyProgress* we followed the Design Based Research methodology. For the evaluation of the tool, we conducted two case studies using a beta version of *NoteMyProgress* over three MOOCs offered in Coursera. The findings of these two case studies are presented as a set of lessons learned that inform about: (1) a list of requirements to inform the design of a second version of the tool; (2) a list of requirements that could serve as a reference for other developers to design new tools that support self-regulated learning in MOOCs.

Keywords: Self-Regulated Learning, SRL, Massive Open Online Courses, MOOC, Tool, Learning Analytics, Dashboard.
Categories: K.3.1, K.3.2

1 Introduction

One of the most relevant characteristics of MOOCs is their massive number of learners. This massiveness makes it difficult for teachers to monitor learners' performance and support them in achieving their goals. In this context, one of the keys for learners to reach their goals is their capacity for self-regulated learning (SRL). Self-regulation is defined as "an active, constructive process whereby learners set goals for their learning and then attempt to monitor, regulate and control their cognition, intentions and behavior, guided and constrained by their goals and the contextual features of the environment" [Pintrich, 99].

G. Publication 7

Maldonado, J. J., Palta, R., Vázquez, J., Bermeo, J. L., Pérez-Sanagustín, M., & Munoz-Gama, J. (2016, October). Exploring differences in how learners navigate in MOOCs based on self-regulated learning and learning styles: A process mining approach. In *2016 XLII Latin American Computing Conference (CLEI)* (pp. 1-12). IEEE.

Exploring Differences in How Learners Navigate in MOOCs Based on Self-Regulated Learning and Learning Styles

A Process Mining Approach

[1,2]Jorge J. Maldonado, [2]René Palta, [2]Jorge Vázquez, [2]Jorge L. Bermeo, [1]Mar Pérez-Sanagustín, [1]Jorge Munoz-Gama
[1]Departamento de Ciencias de la Computación, Pontificia Universidad Católica de Chile, Santiago, Chile
[2]Departamento de Ciencias la Computación, Universidad de Cuenca, Cuenca, Ecuador
jjmaldonado@uc.cl;{randres.paltam; jorge.vazquezm}@ucuenca.ec
jorge.bermeo@ucuenca.edu.ec; {mar.perez; jmun}@ing.puc.cl

***Abstract*— Study in a Massive Open and Online Courses (MOOCs) is challenging, since participants take the course withouth the support of a teacher. Taking a MOOC require the students to have the hability to self-regulate their learning. However, every person has its own learning style and the way each one interact and self-regulate in a MOOC varies. In this work we present an exploratory study from a process-oriented perspective to study whether students with different learning styles and SRL profiles show differences in navigating through a MOOC. Specificaly, we investigate using Process Mining Techniques to analyze logfiles recording the course behaviour of 99 learners accross an Open edX MOOC combined with data from self-reported surveys. Our findings show that learners with different SRL profiles follow similar navigation paths, but there are differences when differentiating students by their learning styles.**

Keywords*— *MOOCs; self regulation; learning styles; process mining.

I. Introducción

La expansión del acceso a internet a nivel mundial a generado un impacto positivo en la educación de las personas. Los MOOCs se han convertido en una fuente de contenidos digitales para todos que pueden ser abordados de forma atemporal y desde cualquier lugar. En este sentido, los MOOC ofrecen contenidos de calidad a millones de personas en todo el mundo, proporcionando nuevas oportunidades para aprender.

Las investigaciones más recientes reportan que entre 2012 y 2015 se inscribieron a un MOOC cerca de 25 millones de personas [1]. De hecho, de acuerdo con un reporte reciente del proyecto MOOC-Maker[1], los MOOCs ya no solo se producen en USA y Europa, sino que América Latina se ha sumado a la ola a gran velocidad [2]. Sin embargo, solo una pequeña parte de quienes inician un MOOC logran terminar el curso completo, dejando a miles de estudiantes comprometidos sin alcanzar los objetivos propuestos. Esto se debe principalmente a tres razones. En primer lugar, el carácter masivo y abierto de un MOOC atrae a una gran diversidad de estudiantes, cada uno con diferentes motivaciones, objetivos, intenciones, creencias y estilos de aprendizaje. Esta variedad de público dificulta la creación de experiencias de aprendizaje adaptadas a la heterogeneidad de perfiles [3]. En segundo lugar, la estructura de los MOOCs actuales proponen diseños instruccionales secuenciales, que fomenta que el estudiante avance de forma lineal en los contenidos del curso. Sin embargo, este diseño y la forma en la que se disponen los elementos del curso con los que interactúa el estudiante, no siempre favorece el desarrollo y uso de estrategias cognitivas adaptadas a su estilo de aprendizaje [4]. Y, en tercer lugar, el aprendizaje en un entorno en línea, y especialmente en un MOOC, requiere que los estudiantes sean capaces de enfrentarse al proceso de aprendizaje de forma autónoma, sin el apoyo de un profesor o tutor. En este contexto, su capacidad para autorregular el aprendizaje es clave para conseguir los objetivos personales y terminar con éxito el curso [5].

La autorregulación del aprendizaje (SRL en adelante) puede ser entendida como un proceso interactivo organizado en tres fases: una fase preparatoria, una fase de completitud de la tarea y una fase de adaptación. Los estudiantes autorregulados se caracterizan por su habilidad de iniciar procesos cognitivos, meta cognitivos, afectivos, motivacionales y de comportamiento, con el fin de tomar las acciones necesarias que les permitan alcanzar las metas y perseverar hasta lograrlo [6]. Sin embargo, la forma en que se autorregulan los estudiantes está estrechamente ligada a su estilo de aprendizaje (EA en adelante). Los EA se definen como las actitudes y comportamientos que caracterizan la forma de aprender de una persona [7]. En este contexto, es importante que un MOOC pueda atender las diferencias en las capacidades de SRL de los estudiantes. También resulta clave, conocer cuáles son los EA predominantes entre los estudiantes al momento de diseñar contenidos y actividades en un MOOC. Estos deberían tener la intencionalidad de ajustarse a sus preferencias y ser facilitadores del aprendizaje.

Tanto los EA como perfiles de SRL han sido estudiados ampliamente en la última década desde la perspectiva de las aptitudes [4]. Esto es, como un conjunto de habilidades que

[1] Enlace al proyecto MOOC-Maker: http://www.moocmaker.org/

H. Publication 8

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., & Delgado-Kloos, C. (2018, September). Predicting learners' success in a self-paced MOOC through sequence patterns of self-regulated learning. In *European conference on technology enhanced learning* (pp. 355-369). Springer, Cham.

Predicting Learners' Success in a Self-paced MOOC Through Sequence Patterns of Self-regulated Learning

Jorge Maldonado-Mahauad[1,3](✉), Mar Pérez-Sanagustín[1,4], Pedro Manuel Moreno-Marcos[2], Carlos Alario-Hoyos[2], Pedro J. Muñoz-Merino[2], and Carlos Delgado-Kloos[2]

[1] Department of Computer Science, Pontificia Universidad Católica de Chile, Santiago, Chile
{jjmaldonado,mar.perez}@uc.cl
[2] Department of Telematics Engineering, Universidad Carlos III de Madrid, Madrid, Spain
{pemoreno,calario,pedmume,cdk}@it.uc3m.es
[3] Department of Computer Science, University of Cuenca, Cuenca, Ecuador
[4] Institut de Recherche en Informatique de Toulouse, Université Tolouse III Paul Sabatier, Toulouse, France

Abstract. In the past years, predictive models in Massive Open Online Courses (MOOCs) have focused on forecasting learners' success through their grades. The prediction of these grades is useful to identify problems that might lead to dropouts. However, most models in prior work predict categorical and continuous variables using low-level data. This paper contributes to extend current predictive models in the literature by considering coarse-grained variables related to Self-Regulated Learning (SRL). That is, using learners' self-reported SRL strategies and MOOC activity sequence patterns as predictors. Lineal and logistic regression modelling were used as a first approach of prediction with data collected from N = 2,035 learners who took a self-paced MOOC in Coursera. We identified two groups of learners: (1) Comprehensive, who follow the course path designed by the teacher; and (2) Targeting, who seek for the information required to pass assessments. For both type of learners, we found a group of variables as the most predictive: (1) the self-reported SRL strategies 'goal setting', 'strategic planning', 'elaboration' and 'help seeking'; (2) the activity sequences patterns 'only assessment', 'complete a video-lecture and try an assessment', 'explore the content' and 'try an assessment followed by a video-lecture'; and (3) learners' prior experience, together with the self-reported interest in course assessments, and the number of active days and time spent in the platform. These results show how to predict with more accuracy when students reach a certain status taking in to consideration not only low-level data, but complex data such as their SRL strategies.

Keywords: Self-regulated learning · Prediction
Massive Open Online Courses · Sequence patterns · Achievement
Success

V. Pammer-Schindler et al. (Eds.): EC-TEL 2018, LNCS 11082, pp. 355–369, 2018.
https://doi.org/10.1007/978-3-319-98572-5_27

I. Publication 9

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2016, April). Recommending self-regulated learning strategies does not improve performance in a MOOC. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale* (pp. 101-104). ACM.

Recommending Self-Regulated Learning Strategies Does Not Improve Performance in a MOOC

René F. Kizilcec
Stanford University
Stanford, CA, USA
kizilcec@stanford.edu

Mar Pérez-Sanagustín
Pontificia U. Católica de Chile
Santiago, Chile
mar.perez@ing.puc.cl

Jorge J. Maldonado
Pontificia U. Católica de Chile & Universidad de Cuenca, Ecuador
jjmaldonado@uc.cl

Author Keywords
Massive Open Online Course; Self-Regulated Learning.

ACM Classification Keywords: K.3.1. COMPUTERS AND EDUCATION: Computer uses in Education.

L@S 2016, April 25-26, 2016, Edinburgh, Scotland, UK
ACM 978-1-4503-3726-7/16/04.
http://dx.doi.org/10.1145/2876034.2893378

Abstract
Many committed learners struggle to achieve their goal of completing a Massive Open Online Course (MOOC). This work investigates self-regulated learning (SRL) in MOOCs and tests if encouraging the use of SRL strategies can improve course performance. We asked 17 highly successful learners about their strategies for how to succeed in a MOOC. Their responses were coded based on a SRL framework and synthesized into seven recommendations. In a randomized experiment, we evaluated the effect of providing the recommendations to learners in the same course (N = 653). Although most learners rated the study tips as *very helpful*, the intervention did not improve course persistence or achievement. Results suggest that a single SRL prompt at the beginning of the course provides insufficient support. Instead, embedding technological aids that adaptively support SRL throughout the course could better support learners in MOOCs.

Introduction
A primary goal of MOOCs is to provide people with an opportunity to learn. Although only a small number of those who start go on to complete the entire course, many online learners selectively engage with parts of the content [3,5,7]. This variation in behavior can be attributed in part to differences in motivation [8,15].

J. Publication 10

Maldonado, J. J., Pérez-Sanagustín, M., Bermeo, J. L., Muñoz, L., Pacheco, G., & Espinoza, I. (2017, October). Flipping the classroom with MOOCs. A pilot study exploring differences between self-regulated learners. In *2017 Twelfth Latin American Conference on Learning Technologies (LACLO)* (pp. 1-8). IEEE.

Flipping the Classroom with MOOCs.

A Pilot Study Exploring Differences between Self-Regulated Learners.

Jorge J. Maldonado[1,2], Mar Pérez-Sanagustín[1]
[1]Departamento de Ciencias de la Computación
Pontificia Universidad Católica de Chile
Santiago, Chile
{jjmaldonado; mar.perez}@uc.cl

Jorge L. Bermeo[2], Lissette Muñoz[2], Guillermo Pacheco[3], Isabel Espinoza[3]
[2]Departamento de Ciencias de la Computación
[3]Instituto Universitario de Lenguas
Universidad de Cuenca
Cuenca, Ecuador
{jorge.bermeo; lissette.munoz; guillermo.pacheco; isabel.espinoza}@ucuenca.edu.ec

***Abstract*—The use of Flipped Classroom (FC) model supported with technology has allowed to optimize classroom time, covering the students' learning demands and adapting to their work pace. However, learners perform at different pace and use external resources during their learning, which requires from them the ability to self-regulate. In this paper, a pilot study is presented and investigates how students with different self-regulation profiles navigate through a MOOC when it is used as a part of FC methodology. To meet the study objectives, by using Process Mining techniques, it has been investigated over log files recording the course behaviour of *N*= 149 learners across an Open edX MOOC used to support FC. The findings show that learners who were exposed to the FC obtained better grades than their counterparts. Also, learners with different SRL profiles follow different navigation paths. This study opens up the possibility to other researchers to explore how learners perform in a MOOC using FC.**

Keywords — Massive Open Online Courses; Self-regulation, Learning strategies; Flipped classroom

I. Introducción

La adopción de las Tecnologías de la Información y Comunicación (TIC) en la actualidad, están suponiendo nuevos retos en todos los contextos (incluido el educativo) que hace algunas décadas atrás, aún eran impensados. La facilidad en el acceso a ordenadores personales y dispositivos móviles, ha permitido que estudiantes y profesores puedan contar con artefactos electrónicos que antes eran considerados como un lujo y sean utilizados hoy como una herramienta más de trabajo. A esto se suma la penetración del Internet en los hogares a nivel mundial, y en especial en Latinoamérica, lo que ha generado un impacto positivo en la forma en cómo se están educando las nuevas generaciones de estudiantes, brindándoles la oportunidad de poder acceder a una gran cantidad de contenidos digitales [1,29]. Como resultado, se han configurado nuevos escenarios de enseñanza y aprendizaje, donde el profesor asume nuevos retos de cara a incorporar la enseñanza digital (Educación 2.0) como parte complementaria a la enseñanza tradicional (cara a cara). Esto ha derivado en profesores que enseñan desde cualquier lugar del mundo y estudiantes que aprenden de forma atemporal, evidenciando que las metodologías centradas en el profesor o en los contenidos ya no son el eje central del proceso de aprendizaje [2] y en su lugar se busca que el estudiante sea protagonista de su instrucción y que los escenarios propuestos sean capaces de satisfacer sus demandas de aprendizaje [3,28].

En referencia a lo anterior, Bergmann, Sams y Gudentrath [4], quienes son conocidos por su propuesta pedagógica de Clase Invertida (CI - Flipped Classroom en inglés), manifiestan que es posible cubrir las necesidades de aprendizaje que los estudiantes demandan, cuando la clase magistral resulta no ser tan efectiva, como por ejemplo cuando el número de estudiantes en clase es numeroso. Según Cockrum [5] el uso del modelo pedagógico de CI apoyado con tecnología, permite hibridar el espacio de aprendizaje y optimizar el tiempo presencial de clases, permitiendo llegar de forma personalizada a los estudiantes, adaptándose a su ritmo de aprendizaje y cubriendo sus demandas al momento de aprender los contenidos.

En este sentido, las Instituciones de Educación Superior (IES) han empezado a explorar y experimentar con iniciativas híbridas, en las cuales haciendo uso de Cursos Abiertos Masivos y en Línea (MOOCs en inglés), ya sean propios (producidos por las mismas IES) o de terceros (por ejemplo cursos sobre la plataforma Coursera) junto al modelo de CI, se busca promover un aprendizaje más activo en el aula de clase, en la que los estudiantes sean capaces de desarrollar actitudes y habilidades como la colaboración, el pensamiento crítico, el pensamiento creativo, intentando desarrollar así estudiantes más autónomos durante su aprendizaje [6].

K. Publication 11

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Pacheco, G., Espinoza, I., & Bermeo, J. Analyzing students' SRL strategies when using a MOOC as a Book.
Link: http://educate.gast.it.uc3m.es/wp-content/uploads/2018/06/HybridEd_2018_paper_11.pdf

Analyzing students' SRL strategies when using a MOOC as a Book

Jorge Maldonado-Mahauad[1,2], Mar Pérez-Sanagustín[1]
Guillermo Pacheco[2], Isabel Espinoza[2], Jorge Bermeo[2],
[1]Departamento de Ciencias de la Computación, Pontificia Universidad Católica de Chile, Chile
[2]Departamento de Ciencias de la Computación, Universidad de Cuenca, Ecuador
{jjmaldonado, mar.perez}@uc.cl
{guillermo.pacheco, isabel.espinoza, jorge.bermeo}@ucuenca.edu.ec

Abstract. This paper presents a pilot study that shows the use of MOOC as a book as part of a Flipped Learning (FL) approach devoted to promote active learning in class. An analysis of the students' behaviour within the MOOC platform indicates that learners' with a different self-regulatory profile show different activity patterns.

Keywords: MOOC, Self-regulation, Learning strategies, Flipped Classroom

1 Introduction

Higher Education Institutions have started to explore and experiment with hybrid initiatives that encompass MOOCs, either self-produced or produced by third-party institutions. Most of current initiatives described in the literature propose the Flipped Classroom model (FC), to promote active learning and students' self-regulatory skills. According to Bergmann, Sams and Gudentrath [1] the use of the Flipped Classroom pedagogical approach, allows to hybridize the learning space and optimize classroom time, facilitating teachers to reach students in a more personalized way [3]. In addition, students can adapt the course to their own pace and personal learning needs [2]. In order to see how different students organize their own learning in this type of experiences, this paper presents a pilot study of a Flipped Classroom. The pilot study was conducted in a first year course of 8 weeks for learning the fundamentals of algorithm and data structures with N=149 learners in which a MOOC about foundations of programming in Python was used to complement the content delivered in class. Two questions were addressed for the analysis: i. *What is the impact on students' academic performance when adopting a FL approach as part of the class proposal?* ii. *How does different profile self-regulation students' behavior differ when a MOOC is used as part of a FL approach proposal.*

2 Pilot Study

A MOOC entitled "Foundations of Programming was designed entirely in Spanish for the Open edX[1] platform. The contents delivered, encompassed 35 readings, 18 video lessons and 3 evaluations at the end of each module. This MOOC was used in the context of the subject "Algorithm, data, and structures I" with 149 students of the Engineering Faculty, University of Cuenca. These students were randomly appointed,

[1] Link to MOOC in Open edX platform in the next link: https://educacionvirtual.cedia.edu.ec/courses/course-v1:UniversidadDeCuenca+UDC001+2016_T1/about

L. Publication 12

Sapunar-Opazo, D., Pérez-Álvarez, R., Maldonado-Mahauad, J., Alario-Hoyos, C., & Pérez-Sanagustín, M. Analyzing learners' activity beyond the MOOC.
Link: http://ceur-ws.org/Vol-2231/LALA_2018_paper_21.pdf

Analyzing learners' activity beyond the MOOC

Diego Sapunar-Opazo[1], Ronald Pérez-Álvarez[1], Jorge Maldonado-Mahauad[1], Carlos Alario-Hoyos[2], Mar Pérez-Sanagustín[1]

[1] Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 486, Santiago, Chile
{dasapunar, jjmaldonado, raperez13, mar.perez}@uc.cl
[2] Universidad Carlos III de Madrid, Avda. de la Universidad, 30, Leganés, España
calario@it.uc3m.es

Abstract. Research on help seeking in MOOCs has mainly focused on analyzing learners' traces within the course forum, or in external social tools which are directly associated to the course. However, little research has been done on the external supplementary websites and digital resources that learners consult outside the MOOC as a way of help seeking. In this working paper, we present the results of an exploratory study with 61 learners from 3 MOOCs in which we analyzed what type of information learners visit outside the MOOC during their study sessions. The results show that learners spent 2% of the time in their study sessions outside the MOOC, being social networking sites, search engines and sites related to the course content the most visited.

Keywords: Learning Analytic, Massive Open Online Courses, MOOCs, Exploratory study.

1 Introducción

De acuerdo con la bibliografía de autorregulación de los últimos 30 años, saber buscar ayuda cuando lo necesitas es una de las estrategias más importantes para lograr sus objetivos de aprendizaje [6] [11]. Esta ayuda puede provenir tanto de otras personas, como también de fuentes de información (búsqueda de información). Debido a la falta de guía por parte de un profesor en los Cursos Masivos en Línea (del inglés *Massive Open Online Courses*), la habilidad de buscar ayuda por parte del estudiante para enfrentar dificultades y lograr los objetivos de aprendizaje es crítica [8].

Investigadores han estudiado la búsqueda de ayuda por parte de los alumnos de MOOCs mediante dos perspectivas: (1) búsqueda de ayuda desde otras personas, y (2) búsqueda de ayuda desde fuentes de información. Respecto a la primera perspectiva, hay estudios que se centran en analizar las interacciones entre los distintos estudiantes dentro del MOOC, generalmente mediante el foro de discusión del curso. Por ejemplo, los autores en [5], proponen diferentes métodos para investigar el intercambio de conocimiento que ocurre en los foros de discusión de un MOOC de Coursera, con el objetivo de ver cómo la estructura de comunicación va cambiando con el transcurso del tiempo. Los autores en [14], analizaron quiénes son los estudiantes más influyentes en los foros