Author/s:
Malekian, Donia

Title:
Temporal analytics for understanding students' study behaviours in digital educational environments

Date:
2020

Persistent Link:
https://hdl.handle.net/11343/240991

Temporal Analytics for Understanding Students' Study Behaviours in Digital Educational Environments

Donia Malekian

Submitted in total fulfilment of the requirements of the degree of

Doctor of Philosophy

School of Computing and Information Systems
THE UNIVERSITY OF MELBOURNE

December 2019

Abstract

The growth of using technologies in education has motivated the development of research to study and promote students' academic success in the educational environments applying them. Digital educational settings provide a high volume of data for this analytical purpose and enable researches to easily collect and analyse data from students' interactions within the system (audit trails) as they proceed towards their study goals. This has motivated the development of Learning Analytics (LA) approaches in these educational settings that offer innovative applications of analytics methods to understand and promote students' study behaviours. One of the main challenges of LA approaches is making connections between students' data traces and educational assumptions which is necessary to ensure the improvement of education and needs interdisciplinary knowledge. In addition, in digital environments, students' data are available at different levels (i.e., fine-grain to coarse-grain) and can be structured in varied ways (e.g., aggregated, temporal). This data requires appropriate formulation to be able to reveal information regarding specific aspects of students' study behaviours. In this matter, the main focus of LA research is on representing students' study behaviour based on the aggregated measures of their task level interactions within digital environments that helped to identify various patterns of learning processes associated with learning outcomes. However, this suffers from some limitations, mainly due to neglecting the time dimension that could better reveal the effect of processes students used during studying. In addition, investigation of students' behaviour at specific context levels such as session is understudied by research that may reveal novel insight into particular aspects of students' study processes. This thesis provides an understanding of students' study behaviours by considering specific levels of conceptualizing students' data and the level at which their behaviours can be structured.

In the first part of this thesis, a temporal analysis based on clustering and statistical tests is performed in the context of a Massive Open Online Course (MOOCs) where students' study behaviour is investigated at session level; that is, dedicated blocks of time in which learners complete single or multiple contiguous learning tasks without interruption. The concept of "session" has rarely been explicitly examined in relation to learning outcome in online learning. Creating and managing sessions when learning online has been associated with an important factor that is associated with students' time management strategy that subsequently can impact their academic outcome. The result of this study provides insight into varied ways that students organize and prioritise their time

in terms of sessions when learning in a MOOC and how these behaviours impact students' academic outcome.

In the second part, a study is conducted in the context of two offerings of a MOOC, where the impact of sequential representations of students' task level behaviour on their learning outcome is investigated. This study considers assessment task outcomes as a proxy for learning outcome rather than students' final achievement that could provide more insight regarding students' progress over time. For this purpose, temporal and non-temporal prediction models are used to show how the sequential nature of learners' task level behaviour in a MOOC is more informative (predictive) of their assessment outcome rather than aggregated measures examined in most studies. Additionally, it provides insight into variations in behavioural sequences of high and low achieving students when preparing for assessments using a sequential pattern mining approach. The results show that it is possible to successfully predict students' readiness for assessment tasks, particularly if the sequential aspects of students' behaviour are represented in the model. Moreover, the results reveal some behavioural patterns reflecting specific learning strategies that may be more effective in promoting learning.

In the third part of this thesis, a study is performed in the context of a digital word processing software to examine the importance of the temporal nature of students' writing behaviour on their writing outcome. It helps to understand how particular aspects of the writing process at specific moments of writing influence the writing outcome. This view is understudied in writing research using students' audit trails (i.e., keystrokes). For this purpose, a temporal approach is proposed combining classification and local feature interpretation as methods. The results reveal the importance of temporal analysis when studying students' writing behaviour. Findings also reveal that the influence of specific writing behaviours on writing quality is likely determined in combination with other writing characteristics, that emphasise the necessity of using models that capture and take the interrelationship between features into account.

In summary, this work contributes to the learning analytics research by raising awareness regarding the need to account for various levels of conceptualizing data and different dimensions when studying learners' behaviour. Various stakeholders could take benefits from the knowledge discovered in this research to improve learners' study behaviours. In particular, educators can identify which study behaviours require support - and (most importantly) when – so they can select relevant interventions to include in their courses.

Declaration

This is to certify that

1. the thesis comprises only my original work towards the degree of Doctor of Philosophy except where indicated in the Preface,
2. due acknowledgement has been made in the text to all other material used,
3. the thesis is fewer than 80,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Donia Malekian

Preface

This thesis has been written at the School of Computing and Information Systems, The University of Melbourne. The major parts of the thesis are Chapters 3, 4 and 5 that are the expanded version of the following papers respectively:

Paula G. de Barba[1], Donia Malekian., Eduardo A. Oliveira, James Bailey, Tracii Ryan, & Gregor Kennedy. The importance and meaning of session behaviour in a massive open online course, *Computers & Education,* 2019. DOI: 10.1016/j.compedu.2019.103772

Donia Malekian, James Bailey, Gregor Kennedy. Predicting Students' Assessment Readiness in Online Learning Environments: The Sequence Matters, To appear in *Proceedings of the 10th International Conference on Learning Analytics and Knowledge (LAK'20)*, Frankfurt, Germany, March 23-27, 2020.

Donia Malekian, James Bailey, Gregor Kennedy, Paula G. de Barba, & Sadia Nawaz, Characterising Students' Writing Processes Using Temporal Keystroke Analysis, *Proceedings of the 12th International Conference on Educational Data Mining (EDM'19),* Montreal, Canada, July 2-5, 2019.

[1] First authorship is shared between Paula G. de Barba and Donia Malekian. Donia Malekian contributed in framing the motivation of the paper, writing and performing all the analytics except for the self-regulated related part that is not provided in this thesis.

Acknowledgements

The completion of this thesis and the incredible journey of PhD would not be possible without the support of my supervisors, my colleagues, my family and friends. Here, I would like to take this opportunity to thank everyone who supported me throughout this lengthy process.

First and foremost, I would like to express my sincere gratitude to my supervisors Prof. James Bailey and Prof. Gregor Kennedy, for their constant support throughout the entire research. I sincerely thank them for giving me this unique opportunity to be their student. Without their significant guidance, patience, motivation and immense knowledge, this thesis would not be possible. They provided me with valuable insights and suggestions on almost every aspect of my research and more importantly, significant freedom to follow my ideas and interests. I am so grateful to have these brilliant academics and wonderful persons as supervisors of my research.

My sincere thank also goes to Prof. Sean Maynard for being the chair of my advisory committee and for the insightful guidance that kept me on track of my research. Meanwhile, I would like to acknowledge the University of Melbourne, the School of Computing and Information Systems, for the excellent workplace environment.

In addition, I would like to thank all my colleagues and friends: Xingjun, Shuo, Florin, Yunzhe, Sadia, Sobia, Anam, Sukarna, Yang, Yamuna, Mohadeseh, Sadaf, Ehsan, Lis, Tracii, and Eduardo for helping me understand many aspects of my PhD path and for all the fun we had over a couple of years. In particular, I would like to express my appreciation to my colleague Paula for all the help and support in my research and for being open to my questions at any time. Thank you all for bringing so much joy into my life!

Finally, I would like to sincerely thank my parents and my brothers for their unlimited support and unconditional love. Of course, finishing this journey and even starting it would not be possible without the help of my dear husband Mostafa, thank him sincerely for believing in me, tremendous support and constant encouragement throughout this journey and my life in general.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Digital technologies have been pervasive in education for over 100 years since the early 20th century (Howard & Mozejko, 2015). While integrating technology into education has not yet led to a revolution, it has resulted in many improvements and advancements in educational settings (Howard & Mozejko, 2015). Modern educational technology encompasses e-learning, multi-media learning, blended learning environments, computer-based instruction, online education, Massive Open Online Courses (MOOCs), virtual learning environments, word processing software, games and many others, often present lower barriers for students in terms of accessibility to learning materials, location, and time (Brooker, Corrin, de Barba, Lodge, & Kennedy, 2018). The growth of using these educational technologies has resulted in many efforts to study, enhance and promote students' academic success in the environments applying them.

In traditional learning, indicators of students' success are usually defined based on the summative assessment of the final learning product, such as exam results or writing products. However, in recent decades, the process and behaviour toward completing a study goal (i.e., learning task, course, writing task) have been found to have a major influence on students' academic achievement (Gulbahar & Ilgaz, 2014). The recognized significance of the underlying process has led to the emergence of a wide variety of research exploring the process students performed to accomplish a given study in different educational contexts. Digital learning environments provide a high volume of students' data available for this analytical purpose and enable researches to easily collect, exchange and analyse data from students' learning activities as they proceed towards their study goals. This expanded and motivated the development of learning analytics approaches in these educational settings.

Learning Analytics (LA), a recently emerging field, provides new applications of analytics methods that could reveal useful information about students' learning behaviour and processes with the aim of understanding and optimising learning and the environments in which it occurs (Ferguson, 2012). Combination of LA with a growing research area that applies data mining methods on educational data called Educational Data Mining (EDM), has enabled researchers to obtain meaningful and interpretable insights regarding students' behaviours and experiences in a data-driven way (Bousbia & Belamri, 2014; Koedinger, D'Mello, McLaughlin, Pardos, & Rosé, 2015). EDM and LA together can bring significant benefits in understanding about students' learning processes and thus can lead to their success in educational outcomes by way of adapting education based on their needs (Greller & Drachsler, 2012; Romero, Ventura, Pechenizkiy, & Baker, 2010).

In this view, the way that the students' success is understood is mainly bounded up with intended learning outcomes and the context in which it occurs (Gibson & Lang, 2018). The main focus of this thesis is on two contexts of MOOCs, and digital writing environments as each of these areas has made a significant contribution to the landscape of digital education in the last decade.

MOOCs are online courses which are often offered through a partnership between universities and online learning platforms (e.g., Coursera, EdX and FutureLearn). In these courses, students are not constrained by location, scope, time or cost (Brooker et al., 2018). Anyone in the world with an Internet connection can learn about their topic of interest directly from the experts in any field, often from the best universities in the world. MOOCs provides a high volume of learners' data available for analytic purposes.

Alternatively, digital writing environments have also become more popular due to their accessibilities and the opportunities that they bring for both learners and educators, and also have expanded the possibilities of writing research. Purpose-built software for writing research has been developed which collects all information possible during the writing process. Initially, these initiatives were restricted to laboratories, but recent advances in software development such as the availability of word processing software and the introduction of cloud-based digital writing software like google docs which are now prevalent in educational environments, have now released such software naturalistic settings, allowing for researchers to examine different aspects of students' writing in real educational environments (Trezise et al., 2017).

There has been considerable focus on the use of LA and EDM approaches to investigate various learners' behaviours in these educational contexts and the related impact on their learning outcome (Diana, Eagle, Stamper, & Koedinger, 2016; You, 2016). In understanding students' behaviour, it is important to consider different levels of data available and the levels at which the data are structured (Nguyen, Huptych, & Rienties, 2018). Data can be represented in various granularity levels from fine-grain (i.e. actions) to coarse-grain (i.e. students, course) (Romero & Ventura, 2013), and can be structured in various formats like aggregated and temporal. In addition, such data can be analysed at different context levels from micro (i.e. task level) to macro (i.e. course level). In this matter, analysis of the previous literature shows that most studies focused on aggregated measures of students' task level interactions within digital environments. However, examining other levels of data with different structure may reveal the effect of novel behaviours which were not identified before. In this regard, focusing on students' behaviour at session level could reveal their time management skill. In addition, considering the temporal aspect of students' behaviour might better reveal the effect of the process students used during accomplishment of a given study. These issues are the focus of this thesis.

This thesis utilises LA and EDM approaches to gain insight into students' study behaviour at different levels (i.e. task, session, and course) and relates that behaviour to their study outcome by incorporating the effect of the level at which their behaviours are structured.

1.1. Motivation for this research

Learning processes are accomplished through a one-way or two-way communication activities where learners interact with educators, content and other learners (Gulbahar & Ilgaz, 2014). This interaction has a direct impact on learning outcome and thus plays a crucial role in the effectiveness of any learning process (Gulbahar & Ilgaz, 2014), particularly in MOOCs environment which is the context of this thesis. In many digital learning environments such as MOOCs, students are often relatively solitary learners, and their success relies more heavily on their abilities to become master of their own learning processes (Zimmerman, 2015).

The growth of using digital learning environments has resulted in many efforts to study students' learning processes and their subsequent impact on academic success. Previous research has widely examined learners' task level learning interactions within online learning environments, such as lecture viewing (Guo, Kim, & Rubin, 2014) and quiz taking (Gikandi, Morrow, & Davis, 2011) that helped to identify various patterns of learning processes associated with learning outcomes. However, one aspect of the learning process that has not been given much attention is how students organize their time when studying online. Highly autonomous settings, such as MOOCs, put additional pressure on learners in this regard. The few studies that have focused on this matter found that online learners typically divide their interactions with the course into discrete study sessions (Vavoula & Sharples, 2002; Whipp & Chiarelli, 2004); that is, dedicated blocks of time in which learners complete single or multiple contiguous learning tasks (e.g., watching a lecture, completing a quiz) without interruption (Tough, 1979). Although the concept of "session" has been previously considered in research as a unit of analysis (e.g., Mccardle & Hadwin, 2015; Morgan, 1987; Winne & Hadwin, 1998), it rarely has been explicitly examined in relation to learning outcome in online learning. This gap has been addressed in Chapter 3 of this thesis in which an exploratory temporal analysis is utilised to better understanding of the importance and meaningfulness of session behaviour in online learning and to gain insight into how students organise their session along a MOOC over time and which types of activities (at task level) they prioritised within these sessions.

Besides the investigation of students’ learning behaviours at different context levels that may reveal novel insights into students’ learning processes, it also is important to consider the varied ways that behaviours could be structured. Even though there exist numerous ways to formulate the learning behaviour, most studies represented it based on the simple frequency measures of students’ interaction within digital learning environments. This suffers from some limitations, mainly due to neglecting the time dimension that could reveal the effect of processes students used during learning. Sequential representation of students’ learning interactions might be more useful in showing how the process of students’ learning over time impacts on other key learning processes and outcomes. Addressing this hypothesis, Chapter 4 of this thesis investigates the importance of sequential representation of students’ task level behaviour on learning outcome compared to aggregated representation. In addition, the sequential behavioural patterns associated with students’ learning outcome are explored.

While MOOCs are one kind of digital learning environment, another is related to the specific activity of students’ academic writing. Like MOOCs, learning analytics has increasingly been applied to academic writing given the critical role that academic writing plays in students’ academic achievement (Hyland, 2013). During recent decades the emphasis of teaching writing has shifted from focusing on the final product of writing (text features) to focusing on the process of developing a written text (Latif, 2008). The significance of the underlying writing process has led to the emergence of a wide variety of research exploring the process of students’ academic writing. The writing process, broadly including planning, writing and revising phases, is a non-linear process (Hayes & Gradwohl Nash, 1996). Efforts in writing research have been made to identify behavioural features that could be an indication of these phases and their related association with writing outcome.

With this view, one stream of writing research has focused on using writers’ keystroke logs from digital writing environments to better understand the phases that comprise the writing process and their relation to writing outcome

(Guo, Deane, van Rijn, Zhang, & Bennett, 2018; Zhang, Hao, Deane, & Li, 2018). Statistics and features collected by most studies examining the relationship between keystroke logs and the writing process mainly rely on aggregated or distribution-based representations of the overall students' interactions with systems (i.e. keystroke logs) (Guo et al., 2018). Even when the data has been summarized to a high standard, neglecting the time dimension may hide the effect of specific behaviours at particular moments of the writing. Temporal analysis has been suggested as a better way to uncover the writing process and its related stages (Van den Bergh & Rijlaarsdam, 1996). It may be that writing quality is related to when writers engage with certain writing phases over time (Van den Bergh & Rijlaarsdam, 1996). However, there are only a few studies that combine writing process and temporal analysis, and these are mostly focused on think-aloud procedures and offline measurements. These have been criticized as being inaccurate representations of the underlying writing process (Tillema, Van den Bergh, Rijlaarsdam, & Sanders, 2011). This challenge is addressed in Chapter 5 of this thesis, by employing a temporal analysis that examines the importance of considering the temporal dimension in characterizing students' writing processes and the respective impact on their writing outcome.

Overall, the aim of this research is to investigate different aspects of students' study behaviour at varied data levels and structures, in two digital educational contexts: MOOCs and digital writing environments. This research contributes to the field of educational technology research with the use of temporal approach (combining learning analytics and educational data mining methods) as a way to investigate students' study behaviour. In this regard, it raises the awareness of research regarding the need to account for various levels of conceptualizing data and the temporal dimension when studying learners' behaviour.

1.2. Research contributions

The main contributions of this thesis are as follows:

Chapter 3 of this thesis investigates how students organise their MOOC workload into study sessions along the course weeks and examine the related impact on their learning outcome. For this purpose, an exploratory temporal analysis is employed using clustering and statistical tests as methods. The main contributions in this work include:

- A temporal approach is employed to get insight into how students organize their time in terms of sessions across the course. In particular, how do they distribute their time in sessions.
- A temporal approach is utilised to get insight into how learners organize and prioritise activities within sessions across the course.
- The patterns of sessions distributions and sessions activities are explored and discussed in relation to learning outcome (course achievement).
- Results show that successful students had more frequent and longer sessions across the course, mixed up activities within sessions and changed the focus of activities within sessions across the course. Further, the findings based on educational theories are discussed.

Chapter 4 of this thesis investigates the impact of sequential and aggregated representations of students' task level behaviour on their learning outcome (assessment outcome) in a MOOC using state-of-the-art data mining methods. The main contributions in this work include:

- A range of prediction models is employed to evaluate the importance of considering the sequential nature of students' activities on learning outcome rather than aggregated measures. In this study, students' learning behaviour is associated to their assessment outcome rather than the final achievement to provide more insight regarding students' progress over time.

- Using the prediction models, the impact of the most recent activities prior to assessment submission on assessment outcome is explored compared to incorporating the more historic activities applied by students. This could reveal the importance of considering the overall process used by students throughout the course rather than the most recent activities prior to assessment submissions.
- A sequential pattern mining approach is utilised to investigate which sequences of behaviours were used by students to be prepared for assessments and whether these sequences differed between high or low level of learning outcome. These factors could not only illustrate the different ways in which students prepare for assessment tasks but could also identify behaviours that need support to improve learners' future preparation.
- The results show how considering the sequential nature of students' behaviour can be used to better predict their learning outcome for assessment tasks compared to aggregated measures.
- The findings confirm the importance of the overall process students applied in the course as a significant indicator of learning outcome rather than only the most recent activities prior to submissions.
- Based on the results students with a high level of learning outcome in assessments had the most sequences related to viewing and reviewing the lecture material, whereas those with low learning outcome had the most sequences related to multiple failed attempts in assessments without doing any activity in between.
- Based on the findings, implications for supporting specific behaviours to improve learning in MOOCs are discussed.

Chapter 5 of this thesis characterises students' writing processes over time based on their keystroke logs to explore the impact of the temporal dimension of their behaviour on their writing outcome. For this purpose, a temporal approach is proposed combining classification and local feature interpretation as methods. The main contributions in this work include:

- Students' writing processes are characterised connected to isolated writing phases, and their influence on writing outcome are examined using a complex machine learning model that consider the interaction of features into account.
- A temporal analysis is applied to show whether, during specific moments of writing, the significance of specific writing behaviours (that predict students' writing outcome) varies as students progress in their writing task. It helps to understand how particular aspects of the writing process at specific moments of writing influence the writing outcome.
- The result suggests that characterizing students' writing based on isolated writing phases is more predictive of writing outcome compared to a baseline model.
- The findings reveal that the effect of several writing characteristics on writing quality changes by taking the time dimension into account. This reveals the importance of temporal analysis when studying students' writing behaviour.

Overall this thesis provides a better understanding of the importance and meaningfulness of students' study behaviours at different levels and structures when engaging with digital educational environments. This helps to build an evidence-base regarding students' learning processes and to develop adaptive learning environments that foster or prevent the use of specific behaviours. It also provides an interesting framework for educators to identify which behaviours require support and provide relevant interventions to include in their courses. Due to the temporal aspect of the analyses, the findings of this research can inform both types of possible effective interventions (e.g., proposing a successful behaviour) and when the intervention would be most useful (e.g., throughout a course or only at the outset of a course).

1.3. Outline of the chapters

This thesis comprises 7 chapters. Chapters 2 provides a detailed introduction to the relevant background of this thesis and literature review. Then, the next three chapters (Chapter 3, 4, and 5) present the three analytic studies conducted for understanding students' study behaviour associated with their academic outcome as well as the obtained results. Chapter 6 provides the key contributions, limitations and potential future directions of this thesis. The thesis chapters are briefly described below:

Chapter 2 provides the relevant background and literature for this research. It starts from providing background on digital learning and writing environments with the main focus on the two contexts of this study (MOOCs and digital writing environments), followed by the relevant background on learning and writing processes, and a detailed literature review on learning analytics and educational data mining accompanied by the relevant methods, data levels, and applications.

Chapter 3 presents the first study focusing on the use of a temporal approach combining learning analytics and learning design as a way to show how students' session behaviour is an important factor related to learning outcome in a MOOC. This chapter provides a meaningful interpretation of how online students manage their time across a MOOC and the relation it has with the learning outcome as well as a detailed explanation of analytics approaches, dataset, along with the obtained results.

Chapter 4 presents the second study focusing on the use of sequential and non-sequential prediction models to show how the sequential nature of learners' task level behaviour in a MOOC is more informative (predictive) of their learning outcome in assessment tasks, rather than aggregated measures. Additionally, it offers more insight into variations in behavioural sequences of high and low achieving students when preparing for assessments using a sequential pattern mining approach. This chapter provides a detailed explanation of analytics approaches, dataset, along with the obtained results.

Chapter 5 presents the third study focusing on using temporal analysis to characterise students' writing processes based on their keystroke logs to show whether, during specific moments of writing, the influence of extracted characteristics on writing quality varies. This reveals the importance of considering the temporal dimension in writing analytics research. This chapter provides a detailed explanation of analytics approaches, dataset, along with the obtained results.

Chapter 6 provides discussion and conclusion over the findings of the three studies, explains the relevant limitations and proposes future research directions.

Chapter 2

Background

In this chapter, the relevant background and literature of this program of research are reviewed which lie at the intersection of digital education, student study behaviour and data analytics. As a starting point, a background on digital education and its growing effect on higher education are provided. Within digital environments, this thesis has focused on MOOCs and digital writing environments as the contexts of the studies. In addition to these fundamental backgrounds, within educational contexts and theories, the main focus is on students learning and writing processes. Thus, it is important to introduce these fundamental concepts by a review of the educational models in these fields. Then Learning Analytics (LA) and Educational Data Mining (EDM) are introduced as two relatively emerging fields in technology-assisted education. In addition to these fundamental concepts, data type, as well as the methods developed for analysing processes of students learning and writing are reviewed. Finally, the current literature is reviewed based on their application in educations focusing on the impact of learning and writing behaviours on study outcome as the backbone of the main contributions of this thesis.

The structure of this chapter is as follows. First, the background, benefits and challenges of digital educational environments, particularly MOOCs and digital writing software, are described in Section 2.1. Then the concepts of LA and EDM, their benefits and challenges and several methods of analytics are provided in

Section 2.2. Further, concepts associated with student learning and writing processes are explained in Section 2.3 and 2.4 respectively. Then the common data features and applied methods of LA and EDM approaches (Section 2.5) and their applications (Section 2.6) are explained. Finally, this chapter is summarised in Section 2.7.

2.1. Digital learning and writing environments

During the last two decades, digital educational environments (e.g., online courses, blended learning environments, MOOCs, writing software) have become pervasive in higher education. These environments refer to the use of digital resources such as computers, software, and systems to enable and manage learning, and have made a significant contribution to the field of education (Peters, 2000). In particular, MOOCs have made a major contribution to the landscape of online learning.

According to Daniel (2012), MOOCs were originally designed in 2008 based on the ideology of connectivism with the main components of peer-based learning in social networks, self-directed strategies, use of online resources on the web, and reliance on technology-based tools. In the beginning, these so-called cMOOCs (as they were based on theoretical framework of connectivism) were structured as an open online learning community, where educational contents were generated by the community members and were shared in an open manner. However, in a short period of time, xMOOCs were developed which were still open and online but were more based on a traditional, linear learning structure which was centred around an educator or content rather than a community of learners. These learning environments often combined video lectures, assessments, quizzes and discussion forums, and were open to anyone around the world who were interested and had access to the internet (de Barba, 2017).

The growth of these courses around the world has been considerable. They are generally designed and run online through a partnership between universities and online learning providers (e.g. Coursera, EdX and FutureLearn), that have an incredibly large number of students from all over the world. In 2017, over 9400 MOOCs were offered, and their enrolment was upward of 81 million students worldwide (Shah, 2018). MOOCs have many advantages that help students benefit and draw attention to their future promise. They present lower barriers for learners in terms of cost, location, scope and time (Brooker et al., 2018). They are accessible to anyone and enable them to learn about their topic of interest at their own pace, often from the leading institutes in the world.

While MOOCs are characterized by different features, one of the ways to distinguish them is by their curriculum design structure. It perhaps best viewed as a continuum that is linear and very structured at one end, and open and loosely structured at the other end (Farrow, 2019). The majority of MOOCs follow a linear structure that resembles the adaptations of traditional higher education (Farrow, 2019). They are content-focused courses with an instructor and a formal curriculum structure. For example, learners often have a deadline to complete a certain number of activities (e.g., weekly quizzes). In the case of open design, the focus is on learners' control over the content and flexibility of learning pace. For example, in this design, materials are often released at the beginning of the course and learners can organize content and learning activities with respect to their own learning pace (Rubens, 2014). The general design principles of MOOCs include content related resources (e.g., video lectures, slides), assessments (e.g. assignments, and quizzes), and collaboration tools (e.g. discussion forums) (Dembo & Seli, 2012).

Even though there are many advantages of MOOCs, the scale of these environments presents educators with challenges, such as providing the necessary supervision to students and the practical ability of instructors to guide and mentor students through courses. This may well make these learning environments more challenging for students in terms of managing the content as well as organising their time. Thus, not all is perfect in the MOOC landscape. To explore this, researchers have focused on a better understanding of the experiences of successful learners in terms of content management and time organization. MOOCs provides a high volume of learners' data available for this analytic purpose.

Moreover, in this age that technology is an essential part of our life, digital writing has also become more popular due to its accessibility and the opportunities it brings for both learners and educators. It allows educators to provide students with timely feedback on assignments, which then is beneficial for students. Moreover, the use of computers in writing over recent decades has expanded the possibilities of writing research and exploring different aspects of learners' writing that influence their writing outcome. Recent advances in purpose-built software for writing offers researchers to collect all information possible during the writing process. Initially, these initiatives were restricted to laboratories, but recent advances in software development have now released such software naturalistic settings, allowing for researchers to examine the writing process in real educational environments (Trezise et al., 2017).

In these environments, every interaction of learners with digital writing software (i.e. learners' keystroke logs) and MOOCs (i.e. learners' clickstream such as the track of their video watched, forum viewed, comment posted) can be recorded and analyzed at various levels to better understand students' study behaviours. This understanding provides a real opportunity to enhance students' study experiences in these environments.

2.2. What are Learning Analytics and Educational Data Mining

There are two fields devoted to analyzing educational data with the aim of supporting education-related decision making and improving education quality; Educational Data Mining (EDM) and Learning Analytics (LA) (Liñán & Pérez, 2015; Trezise et al., 2017).

EDM involves “developing, researching, and applying computerized methods to detect patterns in large collections of educational data that would otherwise be hard or impossible to analyze due to the enormous volume of data within which they exist” (Romero et al., 2010, p. 35). It mainly concerns the processing of educational data through machine learning algorithms that lead to “the nontrivial extraction of implicit, previously unknown, and potentially useful information from data" (Frawley, Piatetsky-Shapiro, & Matheus, 1992, p. 58). In comparison, LA is defined as the “measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” (Siemens & Gasevic, 2012, p. 1).

Although the overlap between these two fields is considerable, from a general perspective, LA focuses more on applications, whereas EDM deals more with the techniques and methodologies. According to Baker and Yacef (2009), and Vahdat (2017), human judgment is influential in decision making in LA, while in EDM automation tools play a major role. Hence, EDM is known as a bottom-up approach to find new patterns in educational data as well as developing new methodologies and models, whereas LA tends to follow a top-down approach with the focus on assessing educational theories through applying existing methods (Vahdat, 2017).

These two fields have evolved as a result of the used technology in education that has led to increasing access to the underlying data about learning and study environments (Gašević, Dawson, & Siemens, 2015). In many studies, EDM and LA are considered complementary (Papamitsiou & Economides, 2014; Vahdat, 2017), and their differences have become less noticeable over time. They both share the same area of study and the common aim of improving the quality of education through analysis of educational data. Thus, similar to several other studies (Vahdat, 2017), the term EDM and LA are used interchangeably in this thesis.

Several models have been proposed that describe LA/EDM process (Chatti, Dyckhoff, Schroeder, & Thüs, 2012; Clow, 2012; Vahdat, 2017). Figure 2-1 represents a reference model of the LA/EDM process based on (Chatti et al., 2012; Clow, 2012; Vahdat, 2017). In this model, the LA/EDM process is shown as an iterative cycle that starts with learners. First, the research aim and the required data about or by the learners are determined. Then the data is collected from the educational environment. Further, the data will be preprocessed and represented in the required format. EDM/LA methods are applied on this data to get insights about the underlying patterns or models. Then the post-processing step would be applied that mainly deals with the interpretation of patterns and obtaining insight regarding the findings. The results of this iterative process might suggest that there is a need to apply changes in the learning process/environment, or students could be provided with appropriate feedback and intervention, or it might be not conclusive because of several issues such as a small dataset or not selecting the appropriate method, so the study will be designed again and analysis is performed with new hypothesis (Vahdat, 2017).

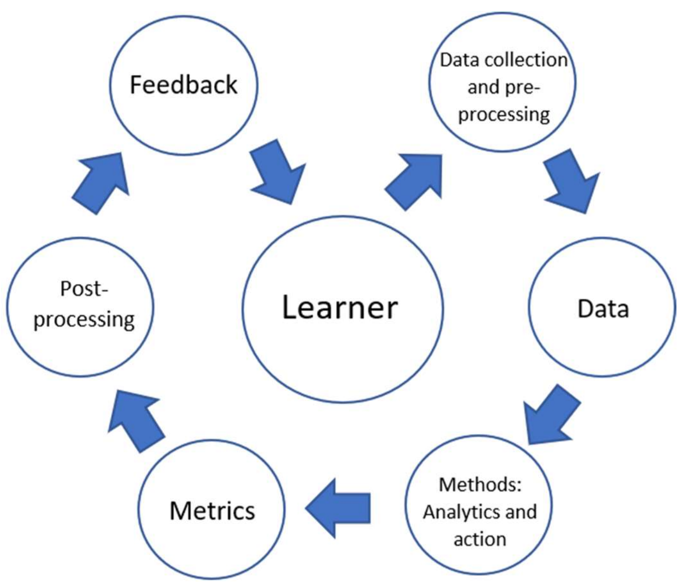


Figure 2-1. A reference model for LA/EDM process according to Chatti et al. (2012), and Clow (2012)

2.2.1. Benefits

The results and findings of LA and EDM can support stakeholders, including educators and administrators with education-related decision making, or they could support learners with feedbacks and interventions to improve their learning. The benefits of LA and EDM have been illustrated in many studies. According to UNESCO policy, the benefits could be explained in three levels of micro (individual user actions), meso (institution-wide), and macro (regional, national, or international) that support three main stakeholder groups including educators, learners and administrators (Buckingham, 2012).

LA and EDM could support educators to become more aware of learners' behaviours, learning processes and learning environments (Baker & Inventado, 2014; Papamitsiou & Economides, 2014; Vahdat, 2017). Such understanding could help educators to monitor students study processes and their study environment, adapt their teaching methods based on learners' needs, discover sources of difficulty in the course material or teaching approaches or provide students with formative feedback about likely areas they can focus on to maximise their chances of success. By increasing the understanding of students' needs and potential areas for success, learners are the other group that could benefit from LA methods. As a result of this understanding, they could be provided with meaningful feedback/intervention related to their behaviour. Lastly, LA and EDM can empower administrators to enhance decision making and budget allowance regarding learning resources and environments (Romero & Ventura, 2007; Long & Siemens, 2011).

Applications of EDM and LA are categorized in many studies. For instance, Romero and Ventura (2013) suggested several areas of the EDM application as; predicting students' performance, scientific inquiry, designing and providing feedback to students and instructors based on their needs, developing student's models which represent their skills and knowledge, developing domain model to define the domain of instruction based on learning items, concepts, skills and their interrelationship, grouping and profiling students' behaviour, designing courseware, planning and scheduling of the course and students, and parameter estimation that derives the parameters of probabilistic models for predicting the probability of an event of interest. Another study by Larusson and White (2014) mentioned that the LA focuses on the following uses; improving student and faculty performance, enhancing student understanding of course material, evaluating struggling learners and attending to their needs, improving marking accuracy, empowering educators to evaluate their own strengths; and investigation of the efficient use of learning resources. EDM/LA applications are further explained with the review of examples in Section 2.6.

2.2.2. Challenges

Even though LA and EDM are beneficial in many ways, there exists challenges and barriers in the application of these methods that need to be tackled by the research. One of the main challenges is making connections between students' low-level data traces with cognition and pedagogical assumptions. This connection is necessary to ensure the improvement of education (Ferguson, 2012). Another factor mentioned in many studies is the ethical obligations (i.e. privacy and anonymity) that is a growing issue to take care of due to the huge amount of data available and new technologies (Blanco et al., 2013; Ferguson, 2012). The other factors are the high cost of applications and techniques, data interoperability and reliability (Vahdat, 2017). The effort here has been made to standardize and enhance the mobility of the educational data (i.e. IEEE standard for learning technology). However, the success in bringing the data together has not been achieved so far since the current state of interoperability is not that effective. With regards to reliability, there are many challenges regarding interpreting and making sense of the educational data through disorganized information (Vahdat, 2017).

2.3. Students' study processes

Learning is an ongoing process that occurs over time (Fautley & Savage, 2008), and in the most simplest of definitions can be considered "as the acquisition of knowledge" (Taylor, 2012). However, it is a complex process and often not observable, thus determining when it has occurred is challenging (Taylor, 2012). Various definitions of learning have been proposed in the research literature. Wittrock (1977) defined it as "the process of acquiring relatively permanent change in understanding, attitude, knowledge, information, ability and skill through experience" (p. 9). In another study, it is defined as "a change in human disposition or capability, which persists over a period of time, and which is not simply ascribable to processes of growth" (Gagne, 1977, p. 3).

For over a century, various theories, models and frameworks have been proposed by researchers to describe and understand how the learning happens and progresses over time. A significant part of this broad body of research is in area of students "achievement motivation" and students learning strategy use (e.g., Pintrich, 2004; Pintrich, Smith, Duncan, & Mckeachie, 1991; Weinstein & Mayer, 1986). This area of research is particularly useful when considering students' study behaviours and processes, which is the focus of the program of research presented in this thesis.

A leading proponent of achievement motivation research, who focussed on students' study behaviours and processes was Paul Pintrich (Pintrich, 2000; Pintrich & Duncan, 1991; Pintrich et al., 1991; Pintrich, Ryan, & Patrick, 1998; Pintrich & DeGroot, 1990). Pintrich, Smith, Duncan, and Mckeachie (1991), outlined several study behaviours and processes that contribute to students' learning. They encapsulated these dimensions in a questionnaire called Motivated Strategies for Learning Questionnaire (MSLQ), which has been used internationally to assess different aspects of the process involved in learning.

This MSLQ is comprised of two sections: motivation and learning strategies. Given the focus of this thesis, the definition of different learning strategies are of most importance (see also Weinstein and Mayer (1986), and Feiz, Hooman, and kooshki (2013). Pintrich and his colleges saw learning strategies section as comprised of *cognitive strategies* (including rehearsal, elaboration, organisation, critical thinking, and self-regulation) and *resource management strategies* (including time and study environment, effort regulation, peer learning, and help seeking) (see Pintrich, Smith, Duncan, & Mckeachie, 1991). A number of research has validated the factor structure of MSLQ and the cognitive and learning strategies that it proposes. Definitions and details of these strategies are provided below.

Rehearsal is defined as a learning strategy which deals with storing information to be learned for learning through repeating, reviewing, or naming the material (Weinstein & Mayer, 1986).

Elaboration is considered as a learning strategy in which students summarise or paraphrase the materials in order to understand them better. This is often applied to integrate and connect prior knowledge to new material to be learned (Weinstein & Mayer, 1986).

Organisation involves learning strategies to select proper information and also make connections between the materials. Examples of this strategy are note-taking, outlining or methods of connecting learning materials to key ideas to be learned (Weinstein & Mayer, 1986).

Critical thinking is a learning strategy that deals with using the previously learned information in new situations (Scriven & Paul, 1987).

Self-regulation involves learners actively planning, monitoring and reflecting on their cognition, motivation, emotion, context and behaviour throughout the learning experiences to achieve their learning goals (Pintrich, 2000). Self-regulated learning (SRL) happens in four cyclic phases which are performed by learners to regulate their learning (Pintrich, 2000): (1) planning (beliefs and thoughts of learners which occur prior to learning about a task), (2) monitoring (learners evaluate the progress of their own learning and evaluate whether it allows them to achieve their learning goals), (3) control (learners take actions as a result of the monitoring phase), and (4) Reflection (learners judge their works and search for reasons regarding their results). A student's success in a task relies heavily on applying several strategies in each phase. For instance, some learners may allocate more time and effort to study specific content that they found difficult in the monitoring phase. Help seeking is another useful strategy in this case. On the other hand, some students may not have enough skills to deal with a certain problem. For instance, they may not know any approach to deal with a difficult task and as a result, give up or drop out the task.

Time and study management involves allocating time in a schedule to dedicate to study and planning. Learners allocate time for studying based on the amount of effort they believe is necessary at that moment and for a specific purpose (Schunk & Zimmerman, 1994). For example, some learners may allocate more time to study certain content than other content prior to an exam. This factor is also associated with learners structuring their environment to conduct their study effectively.

Effort Regulation refers to the amount of effort that learners intent to invest to reach their learning goals, even when there exists difficulties or distractions (Chen, 2002). It has been found to be beneficial for learners in handling the failure and setbacks in their learning process by investing and allocating appropriate effort and resource for better learning in future (Chen, 2002).

Peer Learning is defined as collaborating with peers to understand the learning material. Peers could include classmates, friends, etc. (Jones, Alexander, & Estell, 2010).

Help seeking is defined as an adaptive learning strategy where students manage to seek help and support from resources such as additional books, peers, instructors, and etc. This has been known as one of the important strategies to optimise learning (Zimmerman, 2002). This strategy involves "when, why and from whom to seek help" (Pintrich, 2000).

Students' study or learning processes are typically enacted through a one-way or two-way communication activities where students interact with educators, content and other learners (Gulbahar & Ilgaz, 2014). These interactions have a direct impact on the students' learning outcome and thus play a crucial role in the effectiveness of any learning process (Gulbahar & Ilgaz, 2014).

In traditional education environments, enacting learning process typically relies heavily on what the teacher or educator does (Liñán & Pérez, 2015). In more recent times, and particularly in the last two decades, online learning environments have become pervasive in higher education. While not exclusively the case, in these environments students are provided with access to a range of learning, assessment, and support materials to interact with. However, there is often modest teacher presence and involvement in explicitly guiding students' study or learning processes. Students can often be relatively solitary learners, and their success relies more heavily on their own abilities to master and control their own study and learning processes.

In this program of research, different variables associated with students' learning processes in online learning, including their learning activities and strategies and the subsequent impact on learning outcome are explored and examined.

2.4. Students' writing processes

During recent decades, several models have been proposed to define the writing process. In this regard, two viewpoints have been mainly followed by research. One approach focuses on how writing develops based on the context, whereas the other concentrates on cognition (Graham & Harris, 2013; Graham, Harris, & Santangelo, 2015). The *Contextual view* involves how writers use concrete writing tools such as paper, pencil or word processing software to produce the written text (Graham et al., 2015; Russell, 1997). Whereas the focus of the *Cognitive view* is mainly on the writer and the applied processes involved in developing a written text. For the purpose of this thesis, the cognitive aspect is particularly pertinent due to its association with writing quality (Graham et al., 2015).

Several cognitive models have been used to describe the writing process (for a brief review see Graham et al., 2015). A common approach of writing research in academic settings is to consider three phases: pre-writing, writing, and post-writing (Rohman, 1965).

Prewriting also referred to as planning, has been described as the major element of the internal representation of knowledge used for writing. According to Flower and Hayes (1981), this phase involves three subprocesses as generating ideas, organization of the ideas, and goal setting that goes on throughout composing the essay.

Writing involves composing the ideas and transcribing them onto a paper or screen. According to Indrisano and Squire (2000), writers produce text in sentence parts rather than whole sentences. They suggested that sentence parts are identified by a pause of more than two seconds. These sentence parts are also called bursts in other researches (Xu & Ding, 2014a; M Zhang, Hao, Li, & Deane, 2018).

Post-writing is revising or reviewing the written text or plan with the aim of improving the existing text and fixing the semantic or synthetic problems. According to Grabowski (1996), the revision could be used for making the representation of the text message clearer (semantic) or is concerned with identifying and correcting the text problems such as bad dictation, grammatical errors (synthetic).

Earlier research described writing as a linear series of these three phases, separated in time (Flower & Hayes, 1981; Rohman, 1965). They considered the output of each phase as an input of the other stage. However, the latest writing models and common sense have illustrated that these phases are non-linear and may occur simultaneously (Flower & Hayes, 1981). This means that writers can repeatedly pre-write, compose and revise their written text, and there are great interactions and overlaps between them. For instance, Sommers (1980) found that experienced writers revise their written text constantly while composing the text, rather than postponing all the repair to the end of the writing. This thesis is based on the latter case, which takes into account the non-linear nature of the writing process.

2.5. LA and EDM methods and data features

This section provides an introduction to different types of data in educational settings, various levels that the data can be integrated and analyzed, the most common approaches adopted by LA and EDM, their basic concepts, and relevant literature in various learning context of this thesis.

2.5.1. Data types

The first step of LA/EDM is to collect the data and represent it in a suitable format for the analysis. The data can be gathered from a wide variety of environments such as Learning Management Systems, Personal Learning Environments, Student Information Systems, web-based courses such as MOOCs, word-processing software, and open datasets.

This data can be gathered in different ways based on factors such as the purpose of research and the methodologies selected. For instance, data about student's prior knowledge, demographic factors (such as their age, background), motivation, or interests are usually gathered using self-report items, surveys, questionnaire, or interviews. Another type of important data is the use of specific behaviour or strategy by students toward the accomplishment of their study goal. This data can be gathered using think-aloud protocols (TAP) where students are asked to explain what they are doing while working on a task (Leijten & Van Waes, 2005). This can also be gathered based on retrospective protocols where students are recorded while they are working with the digital environment, and will be asked to explain their behaviour afterwards by looking at the recording (Leijten & Van Waes, 2005). Another important type of data includes some measures from the final study product. For instance, linguistic data from the written essay could be gathered based on the number of words in the essay, vocabulary used, or the number of grammatical errors, which is often the focus of offline writing research.

Another common data type used in most EDM research is students' audit trails which are being recorded by the system. *Audit trails* are students' performed actions within digital environments that usually involve learning time span and the ordered sequence of concepts and interactions (i.e. clickstream, keystrokes). This allows researchers to easily collect, exchange and analyse data from students' study behaviour. This expanded and motivated the development of LA and EDM approaches in educational settings.

The focus of this thesis is on students' audit trail data that can provide valuable insights into students' study behaviours in digital study environments and how these behaviours inform their academic outcome (Park, Denaro, Rodriguez, Smyth, & Warschauer, 2017). This data often has various data levels and contexts that enable different sets of analysis which is the topic of next section.

2.5.2. Data levels, structures and contexts

Depending on the educational environment and the study of interest, different kinds of data exist at various levels and can be gathered and represented with specific structures.

The data provided by educational environments often has various levels of granularity from a relatively coarse level to more fine-grained level (Romero & Ventura, 2013). Romero and Ventura (2013) defined a four-level hierarchy for educational data which included course, students, activities/sessions and events/actions. This hierarchy is represented in Figure 2-2a. Events/actions refer to a single interaction of students within the study environment such as reading a text, pressing a key on the keyboard, or posting comments on discussion forums. Activities/sessions are broader and encompass a series of interactions that have been completed by students. In particular, the term 'session', which is synonymous with 'learning episode', is defined as "a period of time devoted to a cluster or sequence of similar or related activities, which are not interrupted much by other activities" (Tough, 1979, p. 7). The term 'course' refers to a single subject or unit offered by a certain institute, university or online provider. For the analytical purposes and depending on the type of problem, it is often necessary to integrate data from different levels of granularity (Judd & Kennedy, 2011).

In addition to different levels of data, determining the structure and format that represents the data also plays a major role in research that uses educational data, depending on the problem to be solved. Time and sequence are important factors to be considered in this regard (Romero & Ventura, 2013). Time can capture data such as the occurrence of specific study behaviour at a specific time, length of study sessions, time spent on a task or a course, and many others. Sequences could reveal how specific learning items, concepts or activities are ordered in relation to one another.

The context of the data is another important factor that enables interpretation of the results. de Barba (2017) conceptualized the learning context in online learning at three levels of specificity including tasks, sessions and course as presented in Figure 2-2b. The term 'task' is similar to Romero & Ventura's (2013_ 'actions' and refers to an interaction between students and the content or information provided to students in the digital learning environments via an

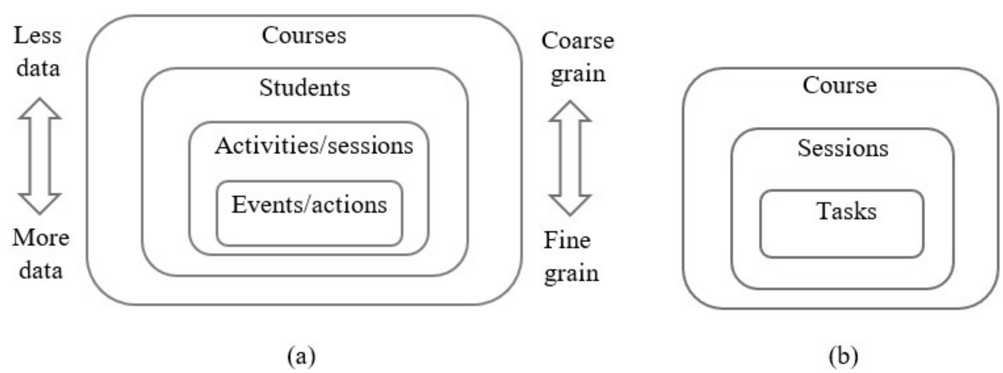


Figure 2-2. a) Different levels of educational data (Romero & Ventura, 2013), b) Different levels of learning context in online learning (de Barba, 2018)

interface (see Richter, 2012). Analysis at this level, involves the use of students' data to investigate the accomplishment of a study task. As explained before, the term session relates to a block of time wherein a series of interactions occurred without interruption (Tough, 1979). Analysis at this level, limits the investigation to students' data to these learning episodes. Finally, the term course refers to a single program of study, such as a MOOC.

Considering these hierarchies, several studies have investigated learners' behavioural data at different levels of granularity (i.e. actions, activities, students and courses) by considering specific units of analysis (tasks, sessions, course). For instance, research studies have explored learners' drop-out behaviour at the course level (Itani, Brisson, & Garlatti, 2018) or quitting behaviour at task level (Ainley, Corrigan, & Richardson, 2005).

Much of LA and EDM research has focused on students' learning interactions and activities in an aggregated format and at task and course level. Even though it is challenging to find studies that examine sessions in online learning, there is a growing recognition of the value of sessions as a unit of analysis (e.g., Hussain, Zhu, Zhang, Abidi, & Ali, 2018; Kovanovic et al., 2015; Riel, Lawless, & Brown, 2018). Moreover, identifying the importance of temporal and sequential structuring of student learning behaviour, compared to aggregated structuring (which is the focus of most literature) has not received much attention. Considering the temporal dimension of students' study behaviour could reveal more information regarding the processes they use during their study and therefore show LA and EDM researchers the importance of these factors, as well as the temporal dimension, in both representing educational data and when investigating learners' study behaviour. These issues are the primary focus of this program of research.

2.5.3. Methods

The EDM and LA apply various methods over the gathered data ranging from descriptive approaches such as clustering and relationship mining (i.e. association rule mining, correlation mining, sequential pattern mining) to predictive techniques (i.e. classification and regression) (Vahdat, 2017). A survey by Liñán and Pérez (2015) described the most popular algorithms in EDM and LA related to prediction, clustering and relationship mining. In this section, a review of EDM/LA methods is provided in two categories of descriptive and predictive with the focus on the used algorithms including classification, clustering and relationship mining.

2.5.3.1. Predictive methods

Much effort has been made by EDM and LA research in applying predictive methods for supporting education and learning (Liñán & Pérez, 2015). In this case, the most common task is classification. Classification concerns with learning a set of rules from a set of observations (input/predictor variables) with known categories (training set) (Duda, Hart, & Stork, 2000). The classifier is then used for identifying (predicting) the category of future observations (test set). Categories are often called classes, targets or labels. Such a predictive model has a considerable role in EDM research with the aim of supporting education. For instance, the success or failure of students in a course can be predicted, and educators can be informed about students who are at risk of having a low learning outcome or not completing the course (Baker & Yacef, 2009; Papamitsiou & Economides, 2014; Vahdat, 2017).

Classification has been used in a wide range of educational research where students' historical behaviour is used to predict their future study behaviour or academic outcome. Many choices exist regarding the representation of students' historical behaviour with different granularities and in different forms and in various contexts. Moreover, numerous decisions could be made regarding the choice of the classifier, which also depends on the application of the model. Some studies have applied various classification methods on the same data and compared the performance of the methods (Romero, Ventura, Espejo, & Martínez, 2008). For instance, Kotsiantis and Pintelas (2005) applied various classifiers to predict students' mark (in a written assignment) based on their demographic characteristics and explore which algorithm is the more accurate one. In the next two subsections, the two types of classifications used in this thesis are explained including artificial neural networks and XGboost.

2.5.3.1.1. Artificial Neural Networks (ANN)

ANNs is a set of computational units connected in a layer-wise structure inspired by the human brain. The units are connected layers by layers, and each connection is associated with a weight that will be adjusted during the learning phase to enable the prediction of the class label. Units in each layer take input from the previous layer, process it, and in turn, pass it to the next layer (Ma, 2018; Zurada, 1992).

Various architectures for ANN have been proposed to develop the model, such as Multi-Layer Perceptron (MLP), Recurrent, Convolutional, etc. The choice of architecture often depends on the application of the model. For instance, Recurrent architecture is appropriate for sequential data, while the convolutional one has found to be effective for image processing applications. Depending on the complexity of the model to be learnt, multiple hidden layers can be defined in the model. The more hidden layers usually enable the model to capture a more complex relationship in the data (Zurada, 1992). However, it also increases the time of training and adjusting weights.

ANNs are popular for learning complex relationships for data in domains such as medicine (Bertolaccini, Solli, Pardolesi, & Pasini, 2017; Shaikhina & Khovanova, 2017), security (Shone, Ngoc, Phai, & Shi, 2018), loan applications (Saha & Waheed, 2017), speech recognition (Zhang, Chan, & Jaitly, 2017), and natural language processing (Goldberg, 2017). There are, thus far, a limited number of studies employing ANNs for tasks in learning analytics. As an example, Okubo, Yamashita, Shimada, and Ogata (2017) confirm the effectiveness of ANNs for predicting students' final grade compared to regression-based models. Here the two ANN architectures used in this thesis are briefly explained.

Multiple Layer Perceptron (MLP): This architecture is a basic three-layer structure ANN also known as feedforward ANN with all layers fully connected; an input layer, hidden layer and output layer (Pal & Mitra, 1992). This is usually considered as a simple proof of concept structure, and more complex structures (e.g. with more layers) are possible. Since all the layers are fully connected, each unit in each layer is connected with every unit of the other layer with a certain weight associated with the connection. Learning is carried out through backpropagation and by updating the weights after processing each data point and based on comparing the output error and expected results. This architecture has a variety of applications such as speech recognition, image recognition and machine translation.

Long-Short Term Memory (LSTM): This architecture is a popular variant of Recurrent ANNs (RNN) (Gers, Schmidhuber, & Cummins, 2000). Unlike a typical feedforward NNs in which the information is processed through a set of operations on the units without a feedback loop (to consider the order of the input), in RNNs the connections between the units are also directed back to themselves (Di, Bhardwaj, & Wei, 2018). In fact, each unit takes the input of the current moment (present) as well as the output of the previous moment (recent past), which then would be combined to determine the new input for the next unit. Thus, the decision of the network at moment t would be affected by the decision it reached at time t-1. The recurrent ANNs are often said to have a memory with the purpose of capturing the information in the sequence that feedforward NNs are unable to capture. Each hidden unit (memory) preserves the sequential information that would be cascaded forward many time steps and would influence the processing of the new data. This way, the correction between the events which are separated by many time steps (called long-term dependencies) can be found. This is because each event depends on (or is a function of) other events in the past.

LSTM is a variant of RNN proposed in the mid-90s (Di et al., 2018). It includes a "gate cell" in which the information can be stored in, written or read from. The gate is similar to NN units. It blocks or passes the information based on their strength that depends on the connection weights. The weights are adjusted through the training process. In LSTM the moment that the data should be entered, left or deleted would be learned through the backpropagation error and weight adjustment. In fact, LSTM is capable of fitting patterns from data across both long- and short-time scales and exhibit the temporal behaviour of data. Unlike many machine learning algorithms that are designed to work with fixed-length data (sequences), it can work with a variable length of the input. A variety of applications have been developed for LSTM using sequential data such as text, spoken word, genomes (Di et al., 2018). In this algorithm, the time and sequential nature of the data are taken into account. Discussion regarding the application of these methods in education field are further explained with the review of examples in Section 2.6.

2.5.3.1.2. eXtreme Gradient Boosting (XGBoost)

XGBoost is a class of *ensemble classifiers* that has been proposed recently dominating the applied machine learning (Chen & Guestrin, 2016). This model is well known as a robust model having consistently good prediction power with time-saving benefits (Chen & Guestrin, 2016). It is a popular implementation of the gradient boosted trees classifier (Ke et al., 2017) in which the estimate of a set of simple models would be combined to predict the target variable accurately. XGBoost uses a set of classification and regression trees for the choice of models. The trees are built sequentially in a way that each tree aims to reduce the error associated with the previous trees. In this algorithm, the structure of the tree and leaf scores will be considered as the weights to be optimised in the objective function. Each tree maps the input data to one of the leaves containing a continuous score. Then the loss function will be optimised based on the difference between the predicted and expected output.

Although ANNs tend to outperform the other algorithms when dealing with unstructured data (i.e. images and text), decision tree-based algorithms are found to perform well on small to medium-sized structured data. In particular, XGBoost algorithm is credited for having a broad range of cutting-edge industry applications since its introduction. However it has rarely been used in education field.

In addition, although complex models such as XGBoost and ANNs generally provide a good prediction power, however, they are usually treated as a black box. Thus, understanding why they made a prediction and what was the contribution of predictor variables seems to be hard due to the complexity of the models (Lundberg & Lee, 2017). That is why many studies prefer to use linear models which are easier to interpret but might have lower prediction power because of not taking the interrelationship between the features into account. However, there is often complex relationships and patterns in students' behavioural data that suggest the necessity of using these complex models when studying students' behaviour.

2.5.3.1.3. Interpreting the predictive models

Providing an interpretation of the complex machine learning models is a challenging task. Several methods are provided that provide interpretation for the used methods and derive the contribution of predictor variables on the prediction. At a very high level, model interpretation means making sense of a certain prediction by the model. Two different levels are defined by studies for model interpretation: Global and local (Lundberg & Lee, 2017).

Global-level interpretation focuses on the entire dataset and answers the question of which features were the most predictive ones (Du, Liu, & Hu, 2018; Lundberg & Lee, 2017). For instance, for interpreting a decision tree-based algorithm, one of the common feature importance measures (to find the contribution of features on prediction) is the number of times a feature was used to split the data. This is an example of global feature importance since the entire data was used for learning.

Local-level interpretation focuses on the individual predictions and identifying which features were most important for each individual prediction (Du et al., 2018; Lundberg & Lee, 2017). Although at the entire dataset, one feature might be the most important one, however, for a specific data point (individual data), it is likely that particular features be more significant in the prediction.

The most common interpretation technique used in the literature and particularly in education field is global-level methods that provide the interpretation for the entire dataset. However having the explanations regarding the individual predictions could provide more insight about the influential features and their interactions.

Below the recently proposed feature interpretation method called SHAP (SHapley Additive exPlanations) algorithm (Lundberg & Lee, 2017) used in this thesis is explained. This algorithm can be used to interpret the result of any machine learning algorithm both at the local and global level.

SHAP algorithm; This method has its theoritical foundation based on the game theory and explains how the predictive model behaved for each data point. In this algorithm, the contribution of each feature for a data point is calculated by comparing the model's predictions with and without that feature. The effect of the feature would be the changes in the prediction of the model which is called the Shaply value. This would be computed by averaging the marginal contribution of a feature value across all coalition, and determines how fairly the prediction is distributed among the feature values of the instance. SHAP explanation is determined as a linear model as provided in Equation 2.1.

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j \, z'_j \qquad \text{(Equation 2.1)}$$

Where g is the explanation model, z' is the coalition vector, $z' \in \{0,1\}^M$, M is the maximum coalition size, and ϕ_j is the Shaply value of feature j which determines the contribution of feature, $\phi_j \in \mathbb{R}$. The values 0 and 1 in the coalition vector detemines the presence and absence of the corresponding feature value respectively. This linear representation of coalitions is a trick for computing the

ϕs. The ϕs would be estimated using KernelSHAP by building the sample coalition, getting the prediction for each coalition, computing the weight for each coalition with the SHAP kernel, fitting the weighed linear model and then returning the Shapely values which are the coefficients of the linear model (for more details see Molnar, 2019).

In this algorithm, every individual is assigned a SHAP value for each feature that determines the feature's contribution to a change in the model's prediction for an individual (local importance). SHAP values are simple value per feature and have the same unit for all features. The global feature importance is computed by averaging the results of absolute SHAP values over all individuals, and is consistent with the local interpretation as the SHAP values are the atomic unit of the global explanations. SHAP algorithm works for both classification and regression.

2.5.3.2. Descriptive methods

Unlike the predictive methods which provide answers to future queries, descriptive methods are used to provide knowledge and understanding regarding the past or recent data. The use of these methods allows us to answer the questions of "what has happened?" and "What is happening?" (Vahdat, 2017). These methods are mainly concerned with discovering interesting patterns or association in the data. Clustering and relationship mining methods are categorized under these methods (Vahdat, 2017), and are used broadly in educational contexts. These methods are explained in the following subsections.

2.5.3.2.1. Cluster analysis

Clustering is a task of grouping objects together such that similar objects are placed in the same group (cluster) and dissimilar objects are in different groups (Jain, Murty, & Flynn, 1999). This task involves three main steps as follow; it begins with feature representation concerning with the selection and extraction of the features from the data. The next step is defining pattern proximity (dissimilarity) measure which usually is selected based on the domain knowledge. This measure would be used to approximate the pairwise similarity between the objects. Finally using the proximity measure, the clustering is performed to discover the groups. The outcome of clustering needs to be validated using an appropriate method (Halkidi, Batistakis, & Vazirgiannis, 2001; Vahdat, 2017). There exists a broad range of clustering algorithms and validation techniques that the result of each could lead to different groupings. In this regard, the human domain knowledge plays a major role for the selection of appropriate algorithm (Bauckhage, Drachen, & Sifa, 2015; Castro, Vellido, & Nebot, 2007; Jain et al., 1999). Clustering algorithms are often classified into four groups of Partitional, Hierarchical, Density-based, and Grid-based clustering (Jain et al., 1999). Here an explanation of these various algorithms is not provided, however, the detail of the algorithms used in this thesis will be explained in the context of the studies.

2.5.3.2.2. Relationship mining

Relationship mining task concerns with discovering the relationship between variables in the data or representing the relations in the format of rules for the subsequent analysis or interpretations (Grigorova, Malysheva, & Bobrovskiy, 2017). Relationships found through this task needs to satisfy two criteria; statistical significance and interestingness (Algarni, 2016). In fact, this task determines the *strength of the relationship* between different variables. Having a large amount of data entails many numbers of variables and thus leads to potentially discovering many relationships between them. The *interestingness* criterion determines the importance of relationships. One of the most common example of this measure is “support” determining how frequently a relationship was observed in the data (Grigorova et al., 2017). Following briefly explains a broad range of relationship mining methods such as association rule mining, sequential pattern mining, and correlation mining.

Association rule mining methods is a procedure of discovering rules from the data that may govern the association between sets of events (Merceron & Yacef, 2010). Its main goal is to find the occurrence of a specific event based on the occurrence of the other events in the data. For instance, it could investigate which errors or mistakes frequently occur together in students' behaviour.

Sequential pattern mining is a specialized task for analyzing sequential data and concerns with understanding and making sense of a time-ordered series of events (Perera, Kay, Koprinska, Yacef, & Zaïane , 2009). More precisely, it takes into account the temporal aspect of the data and aims to explore and find the interesting subsequences from sequential data. For instance, it could be used to recommend students with sequences of resources about a topic to view in order (Cummins, Yacef, & Koprinska, 2006).

Correlation mining explores the underlying dependency between variables (Grigorova et al., 2017). Different patterns from the data could be found, however correlation calculations among the events (rules, sequences, variables) can define the dependency among them. Examples of the methods used for this task are Pearson correlation (Sedgwick, 2012) and Mutual Information (Viola & Wells, 1995) analysis.

These descriptive methods have been widely used in EDM and LA mainly with the aim of determining the relationship in students' behaviours and discovering their behavioural patterns. Each of the methods includes a broad range of algorithms. The details of the related algorithms used in this thesis will be explained further in the context of the studies.

2.6. EDM and LA applications in education

So far, the review of various methods used in LA and EDM fields was provided. The strong point of these approaches has been illustrated by a wide variety of applications in many studies such as those addressed in Section 2.2. Multiple surveys have focused on categorizing applications of EDM based on different properties. For instance, Bakhshinategh, Zaïane, Elatia, and Ipperciel (2017) listed five main categories of EDM applications as student modelling (i.e., representing specific aspects of students such as performance or behaviour), decision support systems (i.e., helping educators in decision making for instance by providing feedback or creating alerts), adaptive systems (i.e., systems which adapt to user's behaviour - personalisation), evaluation (i.e., provide an evaluator of student to help educators, for instance evaluator of student reasoning), scientific inquiry (i.e., testing and developing theories on learning based on the data).

Among the existing applications, "student modelling" has been addressed by most studies and defined as a key area of learning analytics (Bakhshinategh et al., 2017; Graham et al., 2015; Vahdat, 2017). In a comprehensive survey, Bakhshinategh et al. (2017) defined student modelling as the process of representing and analysing characteristics of students such as analyzing their performance or behaviour, representing their misconception, prior knowledge, goals or motivation. They considered two main aims for modelling student as predicting values representing students (i.e. their performance or characteristics) and discovering the structure that describes students. Even though there might be an overlap between these two, the differences in their goals are enough to distinguish them from each other (Bakhshinategh et al., 2017). These two areas are the focus of this thesis.

Another application of EDM and LA is the analysis of discourse data (Bektik, 2019). Learning analytics with the focus on supporting writing have been developed at the intersection of various areas including computational linguistics and automated assessment (Bektik, 2019). An analysis of recent writing research shows that the most addressed approaches are modelling writer behaviours or their written task with the aim of automated assessment grading (prediction) (Balfour, 2013), and describing different aspects of writing behaviour associated with quality of the written text (Xu & Ding, 2014a). This thesis explains the application of EDM in this case based on two categories of prediction and structure discovery as explained previously because of the same goals and properties.

The following subsections start with the general overview of EDM and LA applications including prediction and structure discovery which are the focus of this thesis, accompanied by the relevant work in the field. In each subsection, the review of the closest literature to this thesis is provided based on the two categories of learning and writing analytics.

2.6.1. Prediction

Much effort has been made in this set of applications wherein a variable describing students is estimated. The variables could be student learning, writing outcome, or specific characteristics such as collaboration with other students. It has been a long time that this application attracted researchers from various educational domains. For instance, prediction of high-school drop-out (Sara, Halland, Igel, & Alstrup, 2015), prediction of students' final year result in university based on the first semester result (Tanuar, Heryadi, Lukas, Abbas, & Gaol, 2018), prediction of student writing mark based on think-aloud protocols (Barkaoui, 2010) or based on the linguistic analysis of the written text (Deane, 2013). An analysis of the existing studies in this field shows that the most focus is on creating models of students' behaviour using prediction methods and subsequently using them to predict their learning outcome (Corbett & Anderson, 1994; He, Bailey, Rubinstein, & Zhang, 2015). However, other methods such as feature selection and clustering also have been used. The way that the prediction task is formulated and developed has a huge impact on the results. There exist many choices regarding the representation of historical data and numerous decisions could be made regarding the choice of the prediction model.

Following subsections review the existing studies in this set of applications in two categories of learning and writing analytics.

2.6.1.1. Learning analytics

During the recent decades, the emergence of online courses particularly MOOCs with the related challenges such as low completion rate and low supervision of educators (more details are explained in Section 2.1), attracted many studies on prediction of specific characteristics of students in these environments such as students' final performance or drop-out. Such prediction can be used to enable educators to monitor students' progress and to inform them about the students who are in need of support or are at risk of drop-out. MOOCs provides a high volume of learners' audit trails at a very high granularity for this analytic purpose.

Early research here focused on modelling students based on aggregated measures of their fine-grain interactions (actions/events) within MOOC systems (called participation) (Allen, Davies, Ball, Albrecht, & Bakir, 2019; He et al., 2015), or demographic factors (DeBoer, Stump, Seaton, & Breslow, 2013), or combination of them (Fernandes et al., 2019; Umer, Susnjak, Mathrani, & Suriadi, 2017) to be able to predict their learning outcome or drop-out early in a course. For example, Kovačić and Nz (2010), modelled students' demographic factors in order to predict their drop-out in an online course. Their results suggest that specific demographic groups were at high risk of drop-out. Whereas a study by He et al. (2015) focused on aggregated measures of students' participation in a MOOC (e.g. number of sessions, number of videos watched and access to discussion forums) to predict those who are at risk of drop-out early in the course. In a similar study, Jiang et al. (2014) used various aggregated measures of students' participation such as the number of peer assessments student completed, the number of student social connections (based on the discussion forum interactions) to predict their performance in the first week of a MOOC. On the other study, Qiu et al. (2016) proposed a model with the focus on a combination of students' participation measures (i.e. frequency of forum interactions, watched videos, assignments done, time spent on video and assignments) accompanied with demographic factors (i.e. gender) to predict students' final performance. They found both forum interactions and time spent on videos and assignments as significant predictors of completion rate.

Although there exist numerous ways to formulate learning behaviour, most studies represented behaviour based on simple frequency measures of students' fine-grain interactions with learning environments. This can provide a useful prediction about their learning outcome and can help to shed light on matters of persistence and achievement in online learning. However, one aspect of behaviour that has not been given much attention is the consideration of sequential representation of students' behaviour with a system that could better reveal the effect of the process students used during learning.

Unlike the mentioned category of research with the focus on predicting student final performance or drop out at course level, other research has focused on building machine learning models based on students' historical data with different structure over time in order to predict their learning outcome on exercise tasks (quizzes or assessments) as they complete a MOOC (Minn, Yu, Desmarais, Zhu, & Vie, 2018; Piech et al., 2015). This could be used as a proxy of students' knowledge acquisition during exercise activities and thus be used to suggest students with proactive services such as personalized exercise recommendation (Kuh, Kinzie, Buckley, Ed Bridges, & Ed Hayek, 2007). The main focus of these approaches is on modelling students, based on the concepts that they learned. Among the existing studies, Pardos, Bergner, Seaton, and Pritchard (2013) employed Bayesian algorithms to model students' knowledge over time. They defined acquirable knowledge as atomic pieces, termed knowledge components (KCs), which were determined by subject experts and then used to tag quizzes. A student's past history of quiz responses indicated their level of mastery of the KC. They inferred that mastery had occurred when their model predicted with a high probability that the KC was known by the student. They showed it is possible to predict with good accuracy whether or not each student has acquired the related knowledge during the course. In other work by Yudelson, Hosseini, Vihavainen, and Brusilovsky (2014), a variant additive factors algorithm was used to model students' knowledge. They determined acquirable knowledge using a domain model, which is the vocabulary of skills for the Java programming language. They demonstrated that their model is accurate for both modelling students' knowledge and suggesting the concepts that students can address in their code. In a similar study by Piech et al. (2015), a Recurrent Neural Networks (RNNs) was used to model students' knowledge based on their performance on previously answered quizzes. Each time a student began a quiz task, they predicted whether or not the quiz will be answered correctly. The history of correct and incorrect quiz responses determined the student's gained knowledge. They created a graph of conditional influence between quiz concepts that can be used for curriculum design. Moreover, they demonstrated the effectiveness of RNNs for knowledge tracing applications. Further, the effectiveness of RNNs were confirmed in a very similar study by Minn et al. (2018). They utilised an RNN with a similar procedure as Piech et al. (2015), but for students with different learning abilities separately.

Overall, the focus of these approaches is on modelling students based on the concepts that they learned which provide a great proxy for determining students' knowledge development and mastery on a topic. However, students' success is not just about the content. It also requires appropriate skills to access and work with the most appropriate content. The current focus of educational research in this area is on students' learning strategies to enable effective learning. What remains understudied is evaluating how effective students' behaviours are when it comes to their success.

The other area of research applied and compared various classification models to improve educational prediction. For instance, Al-Shabandar et al. (2017) utilised a range of classifiers to model students based on frequency and aggregated measures of their fine-grain interactions within a MOOC (i.e. the number of videos watched, assessment accessed, posts in discussion forums), accompanied by demographic factors (i.e. gender and date of birth). They compared the prediction power of classifiers namely Logistic Regression (LR), Linear Discriminant Analysis (LDA), Naive Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Neural Network (MLP) and Self Organized Map (SOM) on student learning outcome, and found the random forest as the strongest one. In a similar study, Chen et al. (2018) built various classifiers such as XGBoost, Forward Neural Networks, Random Forest using frequency measures of students interaction within a MOOC (i.e. the number of lecture view, total time spent on the content and number of times specific content posted in the discussion forum). They found XGBoost as the strongest model. The prediction power of a fuzzy neural network was also approved by other studies (Rusli, Ibrahim, & Mohd Janor, 2008).

While comparing and selecting an appropriate algorithm for the prediction task is challenging, switching between algorithms usually results in a minimal change (Baker & Inventado, 2014), whereas the appropriate representation of the data can lead to a substantial change and can be more challenging (Barbour, 2016). In this regard, comparing the prediction power of different algorithms using the same behavioural representation is studied well. However, to the best of our knowledge, comparing the effect of modelling students based on the sequential and aggregated representations of the same behavioural data remained understudied. Such analysis could reveal the importance of considering different structures of data (i.e. temporal dimension) in modelling students' behaviour.

Chapter 4 of this thesis addresses the mentioned challenges in this section by proposing an approach that shows the importance of sequential representation of students' behaviour for predicting their assessment tasks' outcomes compared to aggregated measures. Additionally, this chapter extends the previous work by considering the kinds of study behaviours and interactions students adopt toward a task completion as the foundation of a student model rather than the concepts that have been learned.

2.6.1.2. Writing analytics

Many studies on writing analytics have focused on modelling writers' behaviour or their written text with the aim of predicting the quality of their writing assigned by humans. One of the main application of this task is known as Automated Essay Scoring (AES) which overcomes the challenges of scoring writing by the use of linguistic analysis and machine learning techniques (Lockwood, 2014). An analysis of the literature in this area shows that AES systems were successful in predicting the essay scores given by human in different writing assignments (Lockwood, 2014). Deane (2013) explains the measurable characteristics used by most studies for predicting the essay scores related to the final product of writing, such as average essay length, log of the number of words in an essay, average word length, discourse element length, proportion of grammatical error, scores of other essays with similar vocabulary, or number of least common words. Sophisticated natural language processing such as sentiment and semantic analysis or text summarisation are also used for this purpose (Deane, 2013).

This is clear that the emphasis of the mentioned studies is mainly on using the writing features related to the convolutional print form of the essay (Deane, 2013). However, during the recent decade, the nature of students' writing process has found to have a major influence on their writing outcome rather than only the text and language structure (McCutchen, 2000; Rijlaarsdam, Weijen, & Bergh, 1994). Thus, the emphasis of teaching writing has shifted from focusing on the final product of writing to focusing on the writing process to develop a text (Latif, 2008). Moreover, more insight regarding student writing behaviour can be gained if both information regarding the process and final product are combined (Almond, Deane, Quinlan, Wagner, & Sydorenko, 2012a).

The best way to model the writing process has been defined as the study of a writer in action (Flower & Hayes, 1981). For this purpose, the major elements that make up writing phases and their interactions need to be defined. Early writing research has heavily relied on modelling students' writing process using self-report methods, such as think-aloud protocols (Van den Bergh & Rijlaarsdam, 1996) or interviews (Smith, Campbell, & Brooker, 1999), and examine the related effect (prediction power) on performance measures (scored by a human). However, they are usually considered as inaccurate procedure as what writers say about what they did are likely affected by their notion of what they should have done (Flower & Hayes, 1981).

The use of computers in writing over recent decades has expanded the possibilities of writing research in this filed. A key research area uses keystroke logging (audit trails) in which every keystroke of students is recorded and potentially can reveal information about the students writing processes (Guo et al., 2018; Xu & Ding, 2014a). In this matter, there is a growing body of literature attempted to model students writing processes and use informative prediction methods (i.e. classification models) to find the importance of specific writer characteristics on writing quality (Guo et al., 2018; Zhang et al., 2018). Although it is possible to use these models for improving the AES systems (improving the prediction of essay quality), this goal is not usually of interest in these studies (Guo et al., 2018). In fact, the main focus is on building predictive models that provide actionable and meaningful interpretation regarding the predictions. Among the very few studies that have focused on this matter, a recent study by (Guo et al., 2018) found that patterns of various kinds of pauses between students' keystrokes are a strong predictor of writing quality. They estimated the distribution of the inter-key and intra-word pause durations for each student and found the best fit to be a heavy-tailed probability distribution. They found the estimated parameters of distribution for each student a strong predictor of final score utilizing a linear regression model, then they mapped the related distribution parameters to specific aspects of the writing process (i.e., fluency in transcription process).

Although such analysis based on patterns of pauses provide predictive models characterizing students' writing process, the effect of isolated phases of writing (within planning, transcribing and revising phases) has not been considered and evaluated in these models. In addition, even though the use of simple models of regression makes the interpretation of results easier, the use of more complex models may capture further information about the interrelationships between the extracted characteristics.

Chapter 5 of this thesis addresses the mentioned challenges in this section with utilizing an approach which characterizes students' writing processes connected to isolated writing phases and examine their prediction power on writing quality using a machine learning model which takes into account the interrelationship between features.

2.6.2. Structure discovery

Prediction of students' learning and writing outcome could result in a valuable set of information for academic advisors, who then communicate directly with students. However, these approaches usually do not provide actionable metrics for educators. Thus, in addition to prediction, identifying meaningful factors describing students provides an excellent opportunity to support learners, and to supply early and meaningful feedback to them. Structure Discovery is a second important class of applications in EDM and LA addressing this challenge (Barbour, 2016). The objective of this task is discovering structural properties of learning behaviours based on relational perspectives (Bakhshinategh et al., 2017). This mainly involves profiling students based on a particular set of characteristics and using this profile for different purposes (Bakhshinategh et al., 2017). An analysis of the existing studies in this application shows that the considerable focus has been made on profiling or grouping students based on various factors with the aim of determining the relation of those factors with students' learning or writing outcome using descriptive techniques such as clustering and relationship mining methods.

Following subsections review the existing studies in this set of applications in two categories of learning and writing analytics.

2.6.2.1. Learning analytics

Investigating learners' behaviour throughout a course by way of learning analytics has helped shed light on matters of persistence and achievement in online learning. In such an analysis, it is important to consider the different levels of data available and the levels at which activities are structured (Nguyen, Huptych, & Rienties, 2018). Data can be represented at different levels of granularity such as action, activities, student and course, and can be analysed at various levels of task, session and course (For more detail see Section 2.5.2).

Analysis of the previous LA and EDM research shows that, much effort has been made to model learners' fine-grain interactions within a course or a task, such as quiz taking (Gikandi et al., 2011), lecture viewing (Guo et al., 2014) and sequence of activities (Jovanovic, Gasevic, Dawson, Pardo, & Mirriahi, 2017) in order to inference about various aspects of students learning that have impact on learning outcome. In this case, various measures in online learning environments have been found as important factors that are related with success in learning (Coffrin, Corrin, de Barba, & Kennedy, 2014; Kizilcec, Piech, & Schneider, 2013; Pursel, Zhang, Jablokow, Choi, & Velegol, 2016). For instance frequency measure of video watched and assessment submissions (de Barba, Kennedy, & Ainley, 2016; Pardo, Han, & Ellis, 2016), or lack of participation in discussion forums (Kizilcec et al., 2013), were found to be related to achievement in the course.

These measures of engagement and participation increasingly play an important role in online learning. However, going from initial features of students' participation to more meaningful features taking educational theories into account, requires theoretical understandings of the educational domain. Several educational studies have been proposed that have focused on interactivity of learners (user's task level actions and the underlying purposes) with digital environment and how this can be better understood by taking the students' cognitive and behavioural processes into account when learners are facing with an instructional task (Aldrich, Rogers, & Scaife, 1998; Kennedy, 2004; Rogers & Scaife, 1998). In this regard, Kennedy (2004) has suggested three fundamental elements of interactivity including; design of computer-based instructional task, behavioural strategies that refers to physical activities that students do when facing with instructional task, and cognitive element that refers to the cognitive processes used by learners when facing with instructional task (i.e. specific processes to evaluate and monitor thinking and knowledge). The author provided a model suggesting no direct link between the instructional task and cognitive strategies, however, this relation is considered to be mediated by behavioural strategies (Kennedy, 2004).

Modelling students based on specific interactional behaviour at task level with reference to known learning theory and examing their effect on students' learning outcome has been widely studied using learning analytics approaches (Diana, Eagle, Stamper, & Koe dinger, 2016; You, 2016). For instance, the direct association of re-reading pages and video watching in response to an incorrect assessment submission as a measure of revision strategy (Diana et al., 2016), or the positive effect of regular study and negative impact of the number of late submissions (You, 2016) as a measure of time organisation were found on learning outcome using correlation analysis methods. Even though it is challenging to map these measures to a specific learning strategy, there is a growing recognition of these relationships in recent studies. Studies tried to further augment these relations with utilising self-report questionnaire inquiring about students' use of various learning behaviours and strategies. Among the existing studies, (Pardo et al., 2016) illustrated the significant association between online activities and self-report SRL strategies by students. They identified a significant relationship between several SRL strategies and learners' final performance. Similarly, de Barba et al. (2016) explored the correlation of self-reported motivational constructs with learners' online interactions and final performance in a MOOC. Using correlation analysis, they found that several motivational constructs such as value beliefs influenced performance indirectly through participation, while the learners' interest relates directly to the participation. Although these studies consider pedagogical assumptions using self-reported strategies as the base for their analysis, there exists concerns regarding whether self-reported data properly reflects the actual learning behaviors and strategies.

The empirical investigations of students' interactional behaviour in online learning environments have mostly focused on predefined aggregated factors and evaluating their effect on learners' achievement. However, utilising a data-driven approach might reveal novel behaviours which were not identified by domain experts. Additionally, considering the sequential representation of students' interactional behaviour might reveal the effect of the process students used during learning. Use of sequential pattern mining approaches can be of great advantage here. Among the studies with this orientation, Perera, Kay, Koprinska, Yacef, and Zaïane (2009) utilized sequential pattern mining method in an online collaboration environment to be able to distinguish successful and unsuccessful collaborative groups based on their interactional behaviours. Their result suggested the importance of leadership and interaction within a group. Similarly, Brinton et al. (2016) mined patterns from students' sequences of video-watching behaviour. They identified several video-watching characteristics correlated with a high chance of success in a quiz, such as repeatedly playing and pausing the video (reflection behaviour), and repeatedly revising the previously watched video (revision behaviour). Sequential pattern mining can also be applied through other methods such as hidden Markov model. As an example of this case, Coffrin et al. (2014) investigated students' transitions between course materials in a MOOC and were able to visualize patterns of students' engagement based on their achievement. In these studies, mining patterns, instead of predefining them, led to the identification of previously unknown solutions to defined problems.

As discussed, investigating learners' fine-grain behaviour at task level within a course has been addressed extensively by studies. However, one aspect of behaviour that has not been given much attention is "study sessions" that could reveal how learners organize their time when studying online.

In traditional face-to-face educational settings, learning study sessions are generally experienced by learners as in-class sequences of tasks which are highly structured and teacher-led. Outside of the classroom, learners are typically only required to self-manage the division of learning tasks into sessions when they complete homework. These self-management skills develop over a long period of time, usually starting with short and highly scaffolded sessions in the early years of schooling, and culminating into semester-long projects at university. In contrast, online learning – particularly when asynchronous – places the responsibility with learners to self-manage both their learning sessions and their homework. In doing so, learners need to determine how much learning content or how many activities they should include in a session, and when and where to complete them (Vavoula & Sharples, 2002; Whipp & Chiarelli, 2004).

Recent research has found sessions are still the approach learners take to organise their learning in online environments. For example, when Whipp and Chiarelli (2004) investigated how online learners use SRL skills, participants cited the importance of creating a "psychological space" to be able to dedicate time to their studies. They mentioned these sessions could be very structured, such as setting a specific time every week in a home office, or unstructured, such as deciding to quickly read a text while commuting. In another example, Vavoula and Sharples (2002) conducted a phenomenological study to examine adults' lifelong learning practices. They found that participants organised their learning experiences in three operational levels: (1) activities or tasks, which would be grouped into (2) sessions or learning episodes, which in turn are grouped into (3) learning projects.

Even though it is challenging to find studies that examine sessions in online learning, there is a growing recognition of the value of sessions as a unit of analysis. This is particularly found in studies in the fields of LA and EDM, where researchers have been modelling (predicting or observing) learner behaviour using their audit trail data (e.g., Fincham, Gasevic, Jovanovic, & Pardo, 2018; Gasevic, Jovanovic, Pardo, & Dawson, 2017; Hussain, Zhu, Zhang, Abidi, & Ali, 2018; Jovanovic et al., 2017; Kovanovic et al., 2015; Riel, Lawless, & Brown, 2018). For example, recent studies have used learning analytics to classify and identify learners' use of specific learning strategies within sessions (Fincham et al., 2018; Gasevic et al., 2017; Jovanovic et al., 2017; Kovanovic et al., 2015). These researchers generally define a session as an uninterrupted sequence of tasks where any two tasks must occur within a set number of minutes. The use of learning analytics allows us to examine sessions in the online learning context on a large scale. This is possible as most learning management systems record learners' interactions as audit trails.

Chapter 3 and Chapter 4 of this thesis deal with the specific challenges mentioned in this section. In particular, they study students' behaviour at different levels and with various structures which were understudied in LA literature. In this regard, Chapter 3 focuses on students' study behaviour at session-level. It examines students' behaviour in terms of organising their study sessions across a MOOC through a temporal approach. In addition, it investigates how students prioritise interactions within study sessions, and how their patterns are related to achievement.

Chapter 4, investigates the sequential structure of students' behaviour at task level. As mentioned in Section 2.6.1.1, this chapter also evaluates the importance of the sequential structure of students' behaviour on predicting learning outcome (assessment task outcome). Additionally, it extends the analysis by using a sequential pattern mining approach to investigate which sequences of students' interactional behaviour in a MOOC were used by students to be prepared for assessment tasks and whether these sequences differed between high or low level of achievement in assessments.

2.6.2.2. Writing analytics

The recognized significance of the underlying writing process has led to the emergence of a wide variety of research exploring the process of students' academic writing. A key research area uses keystroke logging to better understand the phases that comprise the writing process (Guo et al., 2018; Zhang et al., 2018). In this matter, an analysis of the studies shows that the main extracted characteristics from keystroke logs to profile students' writing processes are based on estimating patterns of pauses between keystrokes, modelling transcription behaviour, and summarizing the revision behaviour. This profile can be used to gain insight regarding the process of writing and hopefully improving the end product of writing.

Much effort here has focused on characterizing the *transcribing phase* of the writing process which is more related to keyboarding skills and writing fluency.

To characterize this aspect of writing from keystrokes, several studies focused on identifying students' bursts of writing. A burst is defined as the sequence of fast text production and can be identified based on the production of text between two pauses. Thus, the definition of a pause could determine the definition of a burst (Zhang et al., 2018). In these studies, a burst depends on the length of a pause that is considered long enough to count as a burst boundary (Guo et al., 2018; Wengelin, 2006; Zhang et al., 2018). A majority of works have defined pauses as 2 seconds of inactivity in this context (Rosenqvist, 2015). Reasons for this choice are that this threshold is twice the mean typing rate and also using a common threshold allow researchers to compare their results easily with other studies (Rosenqvist, 2015). However, some studies considered this threshold problematic and developed an alternative pause criterion (Zhang et al., 2018). Evaluation of these approaches is often based on examining the correlation of bursts summaries (i.e., mean burst length) with essay score. For instance, a study with this approach has indicated that experienced writers write in longer bursts than others (Matsuhashi, 1981).

Earlier research has also focused on analyzing pauses in keystroke logs and associating them with different phases of the writing process, such as *planning and revising* (Xu & Ding, 2014b). This is often accomplished by exploring the distribution of pauses and then mapping the related parameters to specific writing processes such as planning and revising (Xu & Ding, 2014b). It has long been indicated in literature that long pauses are more likely indicators of planning the text, investigating the source material, revising the written text and deliberation (Baaijen, Galbraith, & de Glopper, 2012; Chukharev-Hudilainen, 2014; Deane, 2014; Xu & Ding, 2014b). Whereas, short pauses are usually assigned to be reflective of the fluency of writing and keyboarding skills (e.g., Alves, Castro, De Sousa, & Strömqvist, 2007). Among the existing studies in this case, Chukharev-Hudilainen (2014) found the exponentially modified Gaussian distribution a good fit for inter-key pauses and mapped specific pauses to the planning process of writing in case they were identified to be long enough. In other work by Almond, Deane, Quinlan, Wagner, and Sydorenko (2012b), a mixture of lognormal distribution was used to describe pause patterns in students' inter-key and intra-word pauses. The parameters of the distribution were found to be associated with writing score based on correlation analysis. They suggested that the smaller pauses were related to the transcription process and longer ones were more involved with the cognitive process such as planning and word selection.

Another direction of research focused on using keystroke logs to more comprehensively characterize various phases comprising the writing process (Guo et al., 2018; Zhang et al., 2018). In a study by Baaijen et al. (2012) on modelling students' writing process, some measures were defined to analyze pauses, bursts and revisions. The authors a) applied burst analysis using the 2 seconds pause limit as the boundary, b) modelled pause durations based on parameters of a mixture model and c) defined revision based on the frequency of text deletions and modifications. They suggested that long pauses may reflect planning, as the writers are more likely to have short but well-formed bursts of writing afterwards.

Overall, extracted characteristics from keystroke logs in terms of burst, revision summaries and the pattern of pauses, have been identified as providing important information regarding the underlying writing process. Association of each extracted characteristic with writing quality (i.e. score by humans) has been mainly considered as the metric for evaluating the usefulness of the features. However, although some keystroke log features were found as a marker of a successful writing process (i.e., high writing quality), there is not always sufficient evidence for the relationship. This may be due to decisions made during data processing and analysis. Statistics and features collected by most studies examining the relationship between keystroke logs and the writing process mainly rely on aggregated or distribution-based representations of the overall keystroke logs (i.e. at the granularity of actions) (Guo et al., 2018). Even when the data has been summarized to a high standard, neglecting the time dimension may hide the effect of specific behaviours at particular moments of the writing.

Temporal analysis has been suggested as a better way to uncover the writing process and its related stages (Van den Bergh & Rijlaarsdam, 1996). It may be that writing quality is related to when writers engage with certain writing phases over time (Van den Bergh & Rijlaarsdam, 1996). An approach to temporal data processing has been the aggregation of data in multiple consecutive episodes, so all participants have a similar number of observations (Van den Bergh & Rijlaarsdam, 1996). A study by Van den Bergh and Rijlaarsdam (1996) has suggested that the relationship between specific aspects of the process with writing quality varies as students' progress in their writing. For instance, they reported the correlation of structuring activities and writing quality is highest at the start of the writing and is lower toward the end. The importance of the temporal analysis of the writing process has been further emphasized by other studies. As another example, Tillema et al. (2011) emphasized that the quality of writing depends on when the specific writing behaviour is performed, rather than how frequently it is applied. They used a think-aloud procedure for data gathering, and then compared the temporal distribution of several activities in students' behaviour such as planning, text production, revising for two groups of students; planner and revisor. They found the temporal distribution of several activities different between the two groups. Even though they found the temporal dimension of students' writing behaviour as a significant factor distinguishing different writers, the data representation procedure mainly relied on think-aloud procedures and offline measurements that have been criticized as being imprecise representation of the writing process (Tillema et al., 2011).

Although the importance of considering the moment(s) at which specific aspects of the writing process occur has been emphasized, this is not a dominant view in keystroke logging analysis, which mainly relies on characterizing students' writing processes based on aggregated and distribution-based measures extracted from overall keystrokes.

As mentioned in Section 2.6.1.2, Chapter 5 of this thesis characterizes students' writing process (connected to isolated writing phases) based on their keystroke logs to evaluate its prediction power on writing quality. This chapter also extends the analysis to show the importance of specific writing characteristics on writing outcome and investigates whether the significance of specific writing characteristics (that predict students' writing quality) varies as students' progress in their writing task (considering the temporal dimension into account).

2.7. Chapter summary

In this chapter, a background on the fundamental concepts of this thesis is provided. The chapter started with a background in digital educational environments. Then LA and EDM, their similarities and differences, as well as their challenges and benefits were described. The students' learning and writing process were introduced afterwards which comprise the educational contexts of this thesis and enables interpretation of the results. Next, the important analytical methods of LA and EDM were explained, followed by their key applications in education including prediction and structure discovery, to determine the common use of the LA and EDM methods in the field. For each application of EDM and LA, the explanation was started from a generic point of view, then the kinds of literature which were closest to the studies in this thesis were reviewed as well as their underlying challenges addressed in this program of research.

The three following chapters employ the studies in this thesis in which many aspects of the reviewed literature are implemented as the basis of them. In Chapter 3, the concept of structure discovery and study sessions are used to uncover students' time organization behaviour. In Chapter 4, the concept of prediction in online learning is adapted to determine the impact of the sequential nature of students' task level behaviour for improving prediction of learning outcome in online learning. In addition, the concept of structure discovery is used to discover sequences of behaviours that distinguish successful and unsuccessful assessment preparation behaviour. Finally, Chapter 5 uses the concept of prediction and structure discovery, to predict students' writing outcome and gain insight regarding the importance of temporal representation of their behaviour in this regard. In addition, this study examines the need to account for the models that consider complex relationships when studying students' behavioural data.

Chapter 3

Investigation of Students’ Session Behavioural Data using Temporal Analysis

In the previous chapter, an overview of learning analytics research in online learning was provided. This chapter defines the research questions emerged from the gaps in the literature and pursued them in the current program of research. In particular, the focus of this chapter is on relating students’ behaviour in terms of their “study sessions” to their engagement and learning outcome over time in a MOOC. This is carried out while taking into consideration the course-level learning design in a weekly format using a temporal approach. The content of this chapter is adapted from the following paper:

Paula G. de Barba[1], Donia Malekian., Eduardo A. Oliveira, James Bailey, Tracii Ryan, & Gregor Kennedy. The importance and meaning of session behaviour in a massive open online course, *Computers & Education,* 2019. DOI: 10.1016/j.compedu.2019.103772

---

[1] First authorship is shared between Paula G. de Barba and Donia Malekian. Donia Malekian contributed in framing the motivation of the paper, writing and performing all the analytics except for the self-regulated related part that is not provided in this program of research.

3.1. Introduction

As explained in Section 2.1, since the integration of modern technology in educational settings and high availability of educational data, much efforts have been made to promote learners' learning experiences and academic outcomes (Chatti et al., 2012). As a result, relatively new fields of research and practice, known as Learning Analytics (LA) and Educational Data Mining (EDM), have emerged as introduced in Section 2.2. Essentially, LA and EDM leverage digital trace data from online systems to gain insights into learners' behaviour and their measure of success (see Section 2.6 for more details on LA and EDM applications in education).

In this view, the way that the students' success is understood is mainly bounded up with intended learning outcomes and the context in which it occurs (Gibson & Lang, 2018). The main focus of this study is on Massive Open Online Courses (MOOCs) which are generally free online courses with the huge number of students enrolled from all over the world without time and location limitation (for more details see Section 2.1). During recent decades, much efforts have been made to better understand the experiences of successful learners in these environments. The current study follows this line of research.

In understanding learners' behaviour, it is important to consider the different levels of data available and the levels at which their activities are structured (Nguyen, Huptych, & Rienties, 2018). Data can be represented at different levels of granularity such as action, activities, student and course, and can be analysed at various levels like task, session and course (for more details see Section 2.5.2). Analysis of the previous research shows that much effort in this matter has been made by way of LA and EDM to examine learners' fine-grain learning interactions within online learning environments, such as quiz taking (Gikandi et al., 2011), lecture viewing (Guo et al., 2014) and sequence of activities (Jovanovic et al., 2017). Such an analysis can provide useful information about students' learning experiences and behaviours and helps shed light on matters of persistence and achievement in online learning. However, one aspect of behaviour that has not been given much attention is the consideration of behaviour at session level. The few studies that have focused on this matter found that online learners typically divide their interactions with the course into discrete study sessions (Vavoula & Sharples, 2002; Whipp & Chiarelli, 2004); that is, dedicated blocks of time in which learners complete single or multiple contiguous learning tasks (e.g., watching a lecture, completing a quiz) without interruption (Tough, 1979) (for more details see Section 2.6.2.1). Creating and managing sessions when learning online has been associated with an important factor that is associated with success in online learning (Whipp & Chiarelli, 2004).

According to de Barba (2017), the occurrence of a session is associated with students' time management. This important aspect of students study behaviour is reflected in Section 2.3 in the area of *time and study management* strategies. With these strategies learners allocate time for studying based on the amount of effort they believe is necessary at that moment and for a specific purpose. In online learning, previous research has identified time management as one of the most important skills related to achievement (Broadbent & Poon, 2015). This is because, when education moves entirely online, learners are under more pressure to manage their own learning experiences, tending to receive less guidance and support than in traditional learning contexts (Dabbagh & Kitsantas, 2004; Kizilcec et al., 2017; Wang, Shannon, & Ross, 2013). This is particularly true in MOOCs, where learners have high levels of autonomy over how to organise their studies. Although the concept of "session" has been previously considered in research as a unit of analysis (e.g., Mccardle & Hadwin, 2015; Morgan, 1987; Winne & Hadwin, 1998), it rarely has been explicitly examined in online learning (exceptions include Kizilcec et al., 2017; Trezise et al., 2017).

This study proposes a way to aggregate learning analytics data from a micro level (i.e., tasks) into a session level while taking into consideration the course level learning design in a weekly format using a temporal approach. This is addressed with an exploratory temporal analysis using clustering and group comparison tests as methods.

First, cluster analysis was applied to explore the patterns of how learners organise their time and learning experiences into sessions, including how they distribute sessions across the course in relation to their length and frequency, and which types of activities they prioritise within these sessions. Then the patterns of sessions over time (course weeks) were investigated in relation to learners' achievement using a group comparison test method. A better understanding of how learners organize their sessions can assist learning or instructional designers in optimising the design of online courses and activities to enhance learning experiences and outcomes. It can also inform educational researchers about a meaningful unit of analysis to better understand online learning. Overall, the research questions investigated in this study were:

Research question 1. How do learners organize their time in terms of sessions across the course? In particular, how do they distribute their time in sessions.

Research question 2. How do learners organize and prioritise activities within sessions across the course?

Research question 3. How do these patterns of sessions distributions and sessions activities relate to engagement and achievement groups?

The results provide insight into varied ways that students organize and prioritise their time in terms of sessions and activity when learning in a MOOC. In addition, the findings indicate how some session-based behavioural patterns of students reflect specific time management strategies, which may be more effective in promoting learning compared to others.

3.2. Approach

3.2.1. Study context and participants

The context for this study was an introductory MOOC in Animal Behaviour. Anyone who had access to a computer with an internet connection could enrol in this course. The registration process consisted of only providing a name and email address. Registration was open while the course was running. There were no fees for registration unless students wanted to be on the accreditation track. All the course materials were not available from day one. Instead, the MOOC was delivered over eight weeks and used a traditional xMOOC design (Yuan & Powell, 2013). That is, the course contents and assessments were released weekly, with a similar structure to traditional higher education courses. Therefore, in this MOOC, the related materials including short video lectures, practise quizzes (week 1), or graded quizzes (weeks 2, 3, 4, 5, 6 and 8) were released at the beginning of each week. Additionally, students were asked to complete two essays in this course, where the grading was done using peer assessment. The final grade of students in the course was calculated based on students' score on the graded quizzes (worth 60% of final grade), together with the peer assessment grades of the two essays (each worth 20% of final grade).

The course offered several resources such as discussion forums, weekly announcements and weekly study guides. Weekly announcements were provided to briefly explain the week content to be covered, released materials, and highlighted activities in discussion forums in the previous week. Weekly study guides summarised the week's aim, key concepts being studied, link to discussion forums and social media used in the course as well as the link to the tasks available in that week (i.e. video or quizzes). Each week of a course was dedicated a discussion forum exclusively to administrative topic related to that week. Additionally, students could create their own threads to post comments, reply to other peer's comments and vote the comments. Figure 3-1 presents a summary of the weekly course design of Animal Behaviour de Barba (2017).

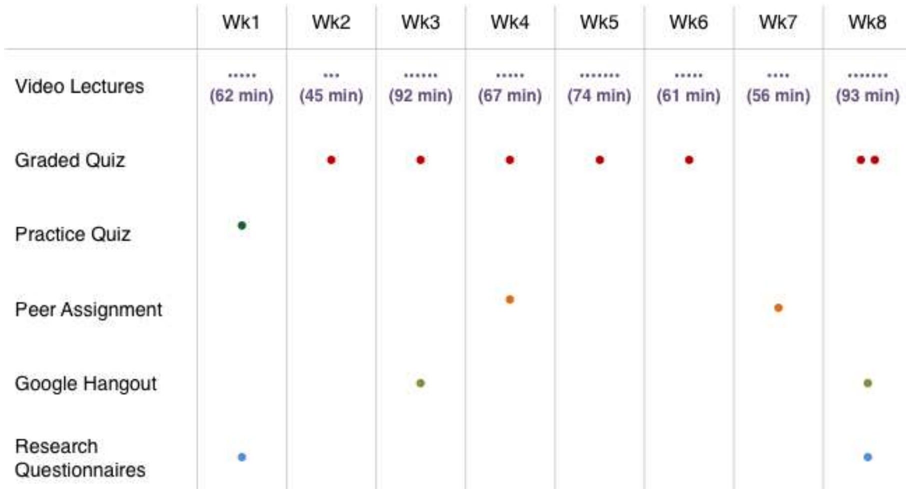


Figure 3-1. Learning design of the Animal Behaviour MOOC (de Barba, 2017)

Participants consisted of learners enrolled in the Animal Behaviour MOOC, offered by the University of Melbourne in partnership with Coursera. Table 3-1 presents global statistics about students and their progression through the course. From the 18,761 learners who had expressed an interest in the course, 9,272 learners engaged with the MOOC (had at least one activity), which were considered as the focus of the session behaviour analysis. Ethics for the use of learning analytics for research purpose was covered by Coursera's Terms and Conditions.

Table 3-1. Global statistics for the Animal Behaviour MOOC

Start date	June 1, 2015
Registered	18,761
Engaged with the MOOC	9,727
Completed the course	403
End Date	July 27, 2015

3.2.2. Measures

This study measured learners' behavioural data, achievement and engagement.

Behavioural data consisted of learners' learning analytics data captured throughout the course and was the basis for analysing session behaviour. Audit logs included a URL which allowed the identification of the type of learning resource accessed. The resource types were: lecture (e.g., watching video lectures including all the related interactions such as play/pause), assessment (e.g., doing practice or graded quizzes, submitting peer assignments), discussion forum (accessing or posting to the forum), weekly guides (accessing the weekly guides), and miscellaneous (accessing additional resources, the dashboard, weekly announcements, landing pages from email notifications, email notification management, signature track pages, honour code page and conducting content search). Audit logs also registered access to these resources, which included a timestamp for each record. These audit logs were then processed to define session behaviour for each learner across the course (more details on the next section "Defining a session").

Achievement was measured by the learners' final grades in the MOOC. A grade ranged from zero to 100 and was an average of the seven graded quizzes (60% of the final grade) and the two peer assignments (40% of the final grade). Based on students' final grade and the type of **engagement** with course, three groups were created: Auditors, Failed and Passed. Auditors were learners who had a final grade of zero, due to not attempting any graded quizzes, but did interact with other resources of the course. Auditors were included as this study was also interested on examining session behaviour of learners lacking achievement, or not oriented towards achievement. Failed were active learners who had a final grade between 0.1 and 64.9. Passed were active learners who had a final grade above 65. From the 9,272 learners who engaged with the course, 14% (1,257) failed the course, 4% (403) passed, and a high percentage of 82% (7,612) were auditors.

Table 3-2 gives a summary of the sample size in each engagement and achievement groups across weeks of the course that shows the number of students in each group at the level of the week (i.e. if a student dropped out a course at week 3, their data is not considered for the following weeks). As can be seen, for each of Auditors and Failed groups there exist differences in the number of students across weeks illustrating those who dropped out the course in the previous week.

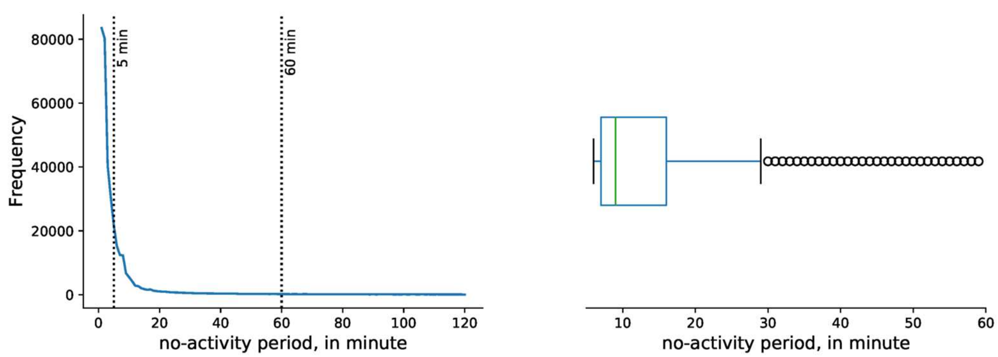
Table 3-2. Summary of students' achievement and engagement groups along the course weeks

	W1	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**
Auditors	7,612	7,272	5,238	4,056	3,124	2,471	1,895	1,336
Failed	1,257	1,257	1,257	1,162	978	826	675	568
Passed	403	403	403	403	403	403	403	403

3.2.3. Defining a session

The definition of session implies that each session has a clear beginning and an end. However, sessions do not come in a well-defined format from audit logs. Audit logs with timestamps were available as a raw distribution of accesses over a period of time. It is up to researchers to determine when a session begins and when it ends. Defining a session using audit logs mainly consists of identifying inter-session times, that is, periods of no activity between two sessions (Mao, Kamar, & Horvitz, 2013). Previous research has used different approaches to choose the inter-session time, such as data-driven (e.g., distribution analysis of inter-session times; Mao et al., 2013) and learning design-based (e.g., no learning resource expected to last longer than 30 minutes; Khan & Pardo, 2016). The current study combines a data-driven approach taking learning design into consideration.

To determine the definition of session, the inter-activity periods of learners' interactions within the system were used. Sessions were defined based on the length of the inter-activity period (IAP) to be long enough to be considered as a session boundary. To be more clear, IAP is the length of time between the start times of learner's two successive interactions with the system. The distribution of all IAPs for all learners are plotted in Figure 3-2a. Since the frequency of IAPs tends to zero as they get larger, this visualization has been limited to the IAPs with up to 120 minutes length. As can be seen, the majority of periods are short and are probably related to the learners' quick subsequent sets of activities, however larger IAPs are not necessarily reflecting that the learner is inactive. In fact, the learner might be active but without any interaction with the system (i.e. watching the lecture or reflecting on the material). Since the IAPs outside the interval of 5 to 60 minutes are too small or too large to be considered as a session boundary, this study limits the further investigation into the IAPs that fall within the range 5 to 60 minutes. The boxplot of the IAPs distribution within this interval is shown in Figure 3-2b. The distribution indicates the first, second and third quartiles as 7, 9 and 16 minutes respectively.



(a) (b)

Figure 3-2. a) Distribution of the inter-activity periods, in minutes, b) Boxplot of the inter-activity periods' distribution within the interval 5 to 60 minutes

The session boundary is defined based on determining a threshold (within 5 to 60 minutes interval), where above that period, the IAPs deviate from the overall patterns in the distribution (i.e. learners are unlikely to do any activity). To determine this threshold, a commonly used Tukey method for outlier detection is used that determines a point as an outlier if it is more than 1.5 interquartile range above the third quartile (Tukey, 1977). The IAPs that are determined as outliers in the distribution are indicated by dots in Figure 3-2b. Applying this method, the session boundary is defined as 30 minutes as can be seen in the figure.

The MOOC learning design indicated that the maximum period of time that a learner could be engaged in an activity without requiring an interaction would be 23min, as this was the maximum length of a video. Other learning resources either required constant interaction (e.g., clicking while completing an online quiz) or did not justify long periods of interaction (e.g., viewing the dashboard or the weekly announcements). Therefore, from a learning design point of view, the minimum inter-session time to be considered while defining a session would be 23 minutes.

Combining both distribution analysis and the MOOC learning design, 30min was decided to be kept as the required period between successive activities for them to be considered as part of the same session. This definition of session was used on all subsequent analysis.

3.2.4. Data processing and analysis

The data processing and analysis performed in this study are summarised in Figure 3-3 and explained below.

1- Defining a session: In data processing part, sessions were identified in the learners' trace logs based on the threshold of 30min. That means, any two activities which are within 30 minutes of one another are considered as a part of the same session (explained in Section 3.2.3).

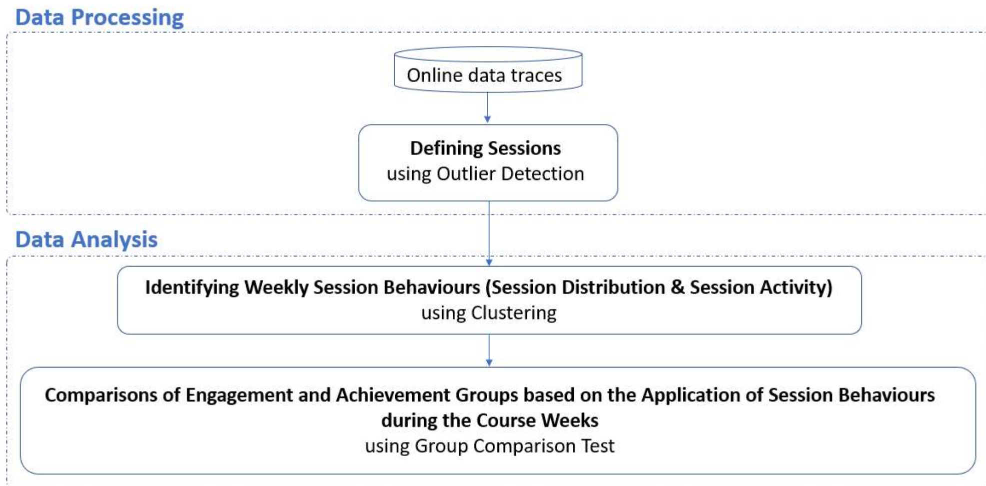


Figure 3-3. A high-level representation of the data processing and analysis

2- Identifying weekly session behaviours (Session Distribution & Session Activity): After defining a session, the analysis was built at the level of weeks of the course. Based on the frequency of the content and learners' engagement in the course, different granularities such as every day, every week, or every two weeks could be examined. This study examined *weekly* session behaviour of learners. This included exploring patterns of session distribution (frequency and length) and session activities (percentage of time spent on each learning resource) over the weeks. **Session distribution** and **session activities** were defined using a **clustering method** described below:

***Session Distribution**:* Session length varied considerably between learners and also over the duration of the course. To account for this variation, the session lengths were discretised using the three intervals in-between the first quartile and third quartile of the distribution and obtained three different sizes of sessions: small, medium and large. However, there was a particularly large number of sessions with total length smaller than 3 minutes, which hindered the estimate of session lengths variations. Therefore, this study did not include sessions with a total length smaller than 3 minutes as part of the discretisation process and added them later as small sessions. A session was determined as small if its length was less than or equal to 10 minutes, as medium if its length was an interval more than 10 and less than 49 minutes, and as large if its length was more than or equal to 49 minutes. Based on these three sizes of sessions, eight weekly behaviours were obtained for each learner, each of which corresponds to the session distribution for them in a specific week. Then clustering task was performed on the overall weekly behaviours of all learners represented by three attributes: frequency of small-size sessions, frequency of medium-size sessions, and frequency of large-size sessions per week. The result of such clustering could determine patterns of weekly sessions, including the frequency and size of sessions across the course for all learners. In fact, for all students, all weekly behaviours were obtained across all weeks. Then these weekly behaviours were clustered (independent of students), to determine the overall patterns of learners' weekly behaviours.

Session Activity*:* session activity was defined based on the percentage of time learners dedicated to specific resources during their weekly sessions. This includes assessment actions, lecture access, discussion forum, weekly guides, and miscellaneous. As the duration of learners' access to these learning resources was not available in the trace logs, the duration of each activity was defined to be the period between the current access and their next access in the same session. The duration of the last access was determined based on the average time spent by learners on that type of learning resource inside the sessions. A clustering task was performed on this data to find overall patterns of weekly sessions in terms of time dedication to specific resources across the course for all learners represented by five attributes: percentage of time on lectures, percentage of time on forums, percentage of time on assessments, percentage of time on guides, and percentage of time on miscellaneous per week.

In order to define both *Session Distribution* and *Session Activity* behaviours, the same ***clustering procedure*** was used as described below. Clustering is a task of grouping objects together such that similar objects are placed in the same group (cluster) and dissimilar objects are in different groups (Jain et al., 1999). An analysis of recent learning analytics research shows that most studies relied on clustering behaviour of learners' participation in MOOCs to identify different learner profiles (Poellhuber & Roy, 2019; Rizvi, Rienties, Rogaten, & Kizilcec, 2019; Van den Beemt, Buijs, & Van der Aalst, 2018). K-means is by far the most commonly used clustering algorithm used in scientific and industrial applications because of scalability, ease of implementation, convergence speed and adaptability to sparse data (Arthur & Vassilvitskii, 2007; Oyelade, Oladipupo, & Obagbuwa, 2010). This study utilised K-means with Euclidean distance metric for the clustering tasks. In this analysis, the number of clusters was determined using the Elbow method (Kodinariya & Makwana, 2013) as well as visualising each cluster members to make sure the clusters were reasonably well separated. Initial centroids were selected randomly with specific probabilities to speed up convergence (for more explanation see algorithm provided in Arthur and Vassilvitskii, 2007). K-means is a localised optimisation method with sensitivity to starting centroids. To account for this, it is common to run the algorithm for multiple starting centroids and select the best output of the consecutive runs (Agrawal & Gupta, 2013). In this study, K-means was performed 1000 times with a different choice of centroids. Additionally, the Silhouette cluster validation test (Rousseeuw, 1987) was performed to verify whether or not a reasonable structure had been found for the resulting clusters. Silhouette scores range from -1 to 1, with a higher value illustrating a stronger structure (i.e. the maximum score of 1 means all learners in each cluster are exactly similar; Rousseeuw, Hubert, & Struyf, 1997).

3- Comparisons of engagement and achievement groups based on the application of session behaviours during the course weeks: After identifying weekly session distribution and activities, the differences in session behaviour were investigated for the three engagement and achievement groups: Auditors, Failed and Passed across the weeks. For this purpose, each of the session distribution and session activity clusters was subject to the three-step analysis. Step one consisted of calculating the frequency of each session behaviour cluster for each engagement and achievement group in each week of the course. Step two used the frequency data to examine the association between groups and *overall session behaviour* in each week of the course applying Fisher's Exact test (Fisher, 1922). This included detecting the triplet and pairwise group differences. In the third step, Fisher's Exact test was applied to examine pairwise group comparisons *for each cluster* in each week of the course. Bonferroni correction (Bland & Altman, 1995) was applied for all pairwise comparison tests to avoid Type I error.

3.3. Results

3.3.1. How do learners organize their time in terms of sessions across the course?

Weekly session distributions were determined based on the frequency of small, medium and large sessions for each learner per week. K-means clustering with Euclidean distance metric was conducted to find the weekly session distribution (SD) patterns. Weekly session behaviours with zero activity were removed from the clustering analysis and added later as a distinct cluster. The optimal number of clusters was found to be four using the Elbow method as well as visualising the members of each cluster to make sure the clusters were reasonably well separated (SD0-SD3). The Elbow plot is represented in Figure 3-4. This method visualises the value of cost function based on different k. The value of k at which the distortion (Sum of the squared Euclidean distances between each member of a cluster and its centroid) declined the most is called elbow and is used as the optimum number of clusters. For the cost function, we used the sum of squared distances of each point to its assigned centroid called distortion.

The algorithm converged to a solution in each run and the best output was selected as final clusters. The Silhouette coefficient for the resulting clusters (k=4) was calculated as 0.51 which indicated a reasonable structure (Struyf et al., 1996).

The weekly session behaviours with zero activity were then added as two separate clusters including zero activity but not left the course (SD4) and zero activity and left the course (SD5). The centroids of the obtained clusters are provided in Figure 3-5, and a description of each cluster is presented in Table 3-3. The cluster size in this table indicates the total number of weekly behaviours across all weeks in specific SD groups.

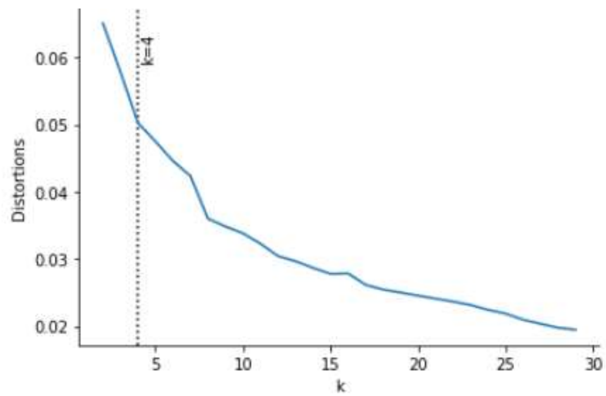


Figure 3-4. Elbow plot for k-means clustering of weekly Session Distribution behaviour

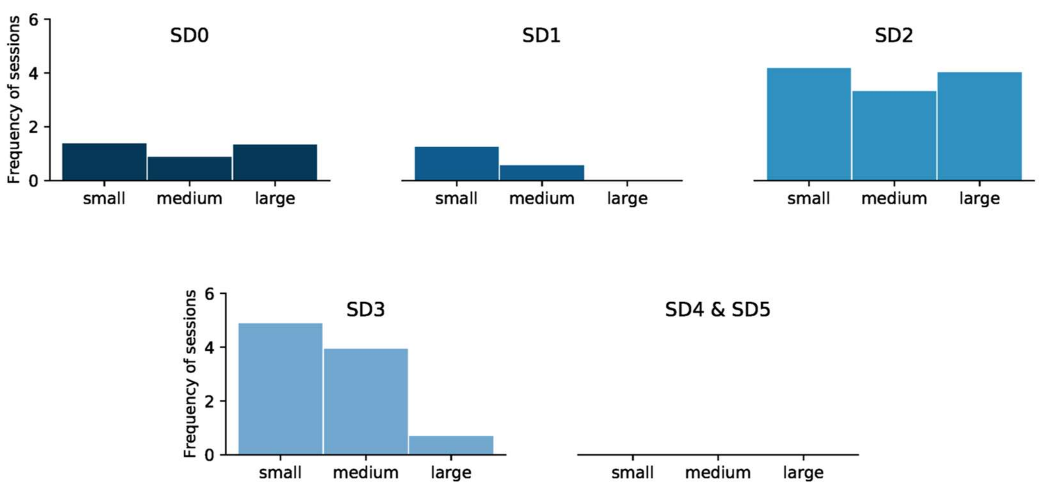


Figure 3-5. Cluster centroids for weekly session distribution (SD cluster centroids)

Table 3-3. Description of SD clusters

Cluster ID	Cluster name	Description	Cluster size
SD0	Low frequency	Low frequency of small (1.4), medium (0.9) and large (1.3) sessions across the week.	5,002
SD1	Low frequency, no large session	Low frequency of small (1.1) and medium (0.6) sessions across the week, but no large sessions.	13,186
SD2	High frequency	High frequency of small (4.2), medium (3.3) and large (4.0) sessions across the week.	919
SD3	High frequency, low large sessions	High frequency of small (4.9) and medium (3.9) sessions and low frequency of large (0.7) sessions across the week.	2,157
SD4	No activity	No activity across the week but did not leave the course on that week.	15,231
SD5	Drop out	No activity and left the course that week.	8,010

3.3.2. How do learners organize and prioritise activities within sessions across the course?

K-means clustering was applied to find patterns of weekly sessions in terms of time dedicated to specific resources across the course for all learners. The optimal number of clusters was found to be as five using the Elbow method and visualising the members of clusters (SA0-SA4). The Elbow plot is represented in Figure 3-6.

Initial centroids were selected randomly with specific probabilities to speed up convergence (Arthur & Vassilvitskii, 2007)). K-means was performed for 1000 times with a different choice of centroids to account for its sensitivity to starting points. The algorithm converged to a solution in each run and the best output was selected as final clusters. The Silhouette coefficient for the resulting clusters (k=5) was calculated as 0.51 which indicated a reasonable structure (Struyf et al., 1996).

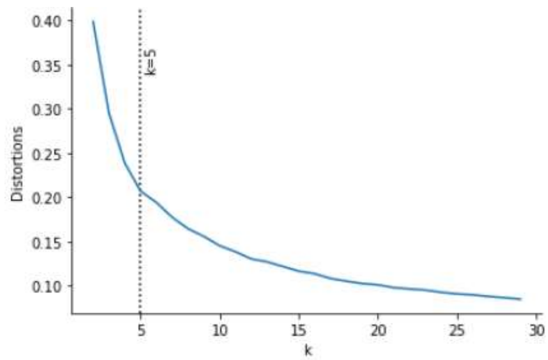


Figure 3-6. Elbow plot for k-means clustering of weekly Session Activity

Similar to the session distribution analysis, weekly session behaviours with zero activity were removed from the clustering analysis and then added as two separate clusters including zero activity but not left the course (SA5) and zero activity and left the course (SA6). The centroids of the obtained clusters are provided in Figure 3-7 and a description of each cluster is presented in Table 3-4.

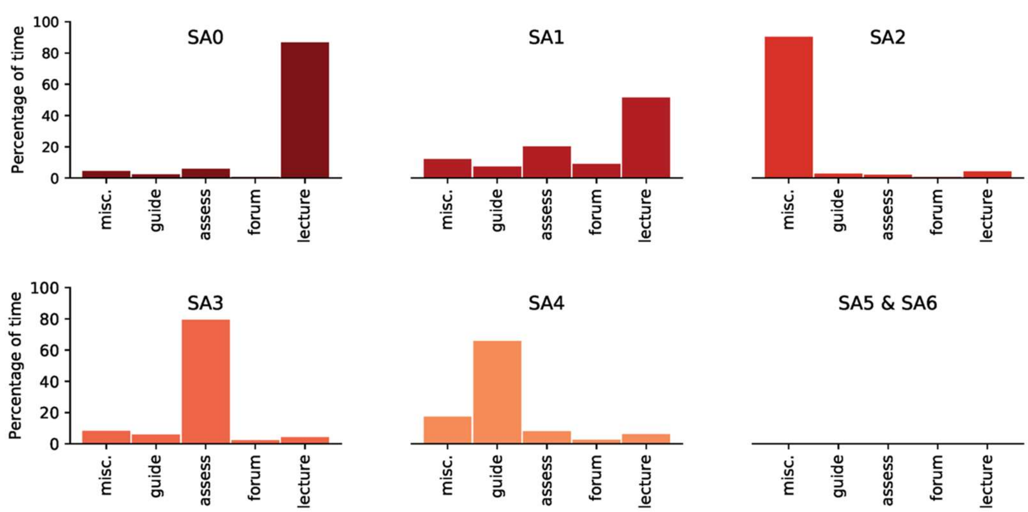


Figure 3-7. Cluster centroids for weekly session activity (SA cluster centroids)

Table 3-4. Description of SA clusters

Cluster ID	Cluster name	Description	Cluster size
SA0	Lecture	High proportion of the session time dedicated to accessing lectures (89%) in relation to other learning resources combined.	8,197
SA1	Lecture mix	Great proportion of the session time dedicated to accessing lectures (51%), but also considerable time dedicated to accessing assessment (20%), miscellaneous (12%), forum (10%) and guide (7%).	4,277
SA2	Miscellaneous	High proportion of the session time dedicated to accessing miscellaneous (90%) in relation to other learning resources combined.	3,530
SA3	Assessment	High proportion of the session time dedicated to accessing assessment (79%) in relation to other learning resources combined.	3,260
SA4	Guide	High proportion of the session time dedicated to accessing the weekly guide (66%) in relation to other learning resources combined.	2,000
SA5	No activity	No activity across the week but did not leave the course on that week.	15,231
SA6	Drop out	No activity and left the course that week.	8,010

3.3.3. How do these patterns of sessions distributions and sessions activities relate to engagement and achievement groups?

3.3.3.1. Session distribution and engagement and achievement groups

Three visualisations presented in Figure 3-8 represent the percentage of session distribution clusters for each engagement and achievement group across the weeks. Learners who left the subject in a specific week are not considered in the analysis of the later weeks.

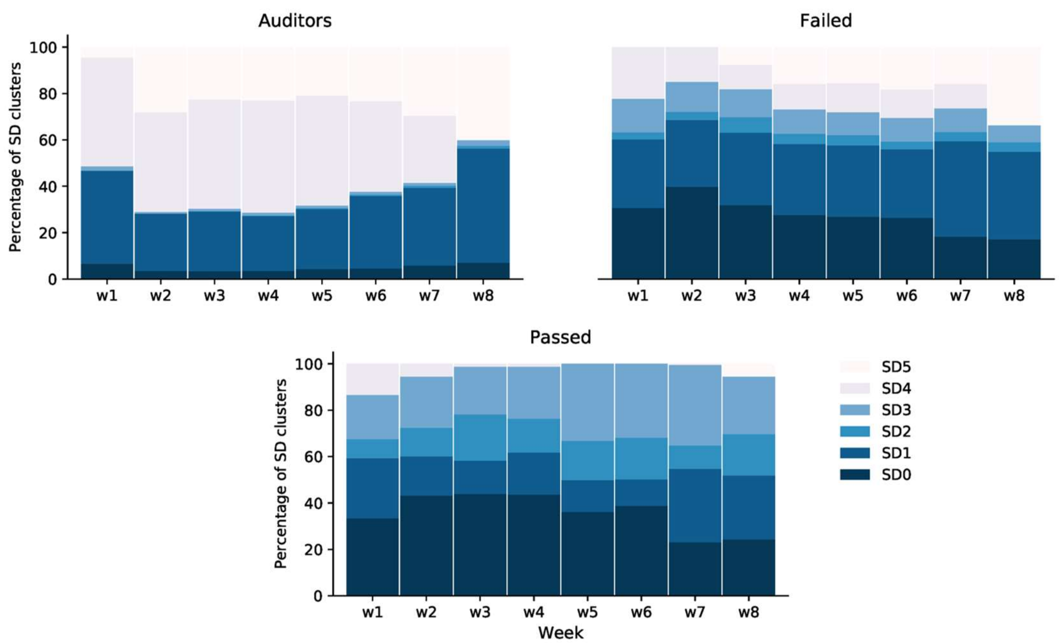


Figure 3-8. Percentage of SD clusters over the weeks for each engagement and achievement group

In this phase, the frequency of each SD cluster for each engagement and achievement group in each week was calculated. Then the frequency data was used to examine the association between groups and all SD behaviours in each week of the course using the Fisher's exact test. This included detecting the triplet and pairwise group differences. In fact, Fisher's Exact test was applied on the 3*6 contingency table (3 engagement groups versus 6 SD clusters) for each week, that indicated an overall association between the three engagement and achievement groups and the session distribution behaviours over all weeks as presented in Table 3-5.

Table 3-5. Association between engagement and achievement groups and SD clusters

	Session Distribution Clusters							
	W1	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**
All groups	**	**	**	**	**	**	**	**
Auditors vs. Passed	**	**	**	**	**	**	**	**
Auditors vs. Failed	**	**	**	**	**	**	**	**
Passed vs. Failed	**	**	**	**	**	**	**	**

Note. ** $p < .001$.

Further analysis using Fisher's Exact test examined pairwise group differences related to the proportion of session distribution clusters applied across the weeks.

In this analysis, the frequency of existence and non-existence of each SD cluster for each engagement and achievement group in each week was calculated. Then this frequency data was used to examine the association between groups and each SD behaviour in each week of the course using the Fisher's exact test. This included detecting the pairwise group differences. In fact, Fisher test was applied based on the 2*2 contingency table (2 rows for pairwise engagement groups versus 2 columns for SD exists/ SD not exists). The adjusted alpha level of .003 was used for determining the significance of the difference between groups in each week. These results are presented in Table 3-6.

Table 3-6. Group Differences in the Proportion of SD Clusters per Week

	Session Distribution Clusters							
	W1	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**
Low frequency (SD0)								
Auditors	$.07^{a}$	$.04^{a}$	$.03^{a}$	$.04^{a}$	$.04^{a}$	$.05^{a}$	$.06^{a}$	$.07^{a}$
Failed	$.31^{b}$	$.40^{b}$	$.32^{b}$	$.28^{b}$	$.27^{b}$	$.26^{b}$	$.18^{b}$	$.17^{b}$
Passed	$.33^{b}$	$.43^{b}$	$.44^{c}$	$.44^{c}$	$.36^{c}$	$.39^{c}$	$.23^{b}$	$.24^{c}$
Low frequency, no large session (SD1)								
Auditors	$.40^{a}$	$.25^{a}$	$.26^{a}$	$.24^{a}$	$.26^{a}$	$.31^{a}$	$.34^{a}$	$.49^{a}$
Failed	$.30^{b}$	$.29^{b}$	$.31^{b}$	$.31^{b}$	$.31^{b}$	$.30^{a}$	$.41^{b}$	$.39^{b}$
Passed	$.26^{b}$	$.17^{c}$	$.14^{c}$	$.18^{a}$	$.14^{c}$	$.11^{b}$	$.32^{a}$	$.28^{c}$
High frequency (SD2)								
Auditors	$.00^{a}$	$.00^{a}$	$.00^{a}$	$.00^{a}$	$.00^{a}$	$.01^{a}$	$.01^{a}$	$.01^{a}$
Failed	$.03^{b}$	$.04^{b}$	$.07^{b}$	$.04^{b}$	$.04^{b}$	$.03^{b}$	$.04^{b}$	$.04^{b}$
Passed	$.08^{c}$	$.12^{c}$	$.20^{c}$	$.15^{c}$	$.17^{c}$	$.18^{c}$	$.10^{c}$	$.18^{c}$
High frequency, low large sessions (SD3)								
Auditors	$.02^{a}$	$.01^{a}$	$.01^{a}$	$.01^{a}$	$.01^{a}$	$.01^{a}$	$.01^{a}$	$.02^{a}$
Failed	$.15^{b}$	$.13^{b}$	$.12^{b}$	$.10^{b}$	$.10^{b}$	$.10^{b}$	$.10^{b}$	$.07^{b}$
Passed	$.19^{b}$	$.22^{c}$	$.21^{c}$	$.22^{c}$	$.33^{c}$	$.32^{c}$	$.35^{c}$	$.25^{c}$
No activity (SD4)								
Auditors	$.47^{a}$	$.43^{a}$	$.47^{a}$	$.48^{a}$	$.47^{a}$	$.39^{a}$	$.29^{a}$	.00
Failed	$.22^{b}$	$.15^{b}$	$.11^{b}$	$.11^{b}$	$.12^{b}$	$.12^{b}$	$.11^{b}$	.00
Passed	$.13^{c}$	$.06^{c}$	$.01^{c}$	$.01^{c}$	$.00^{c}$	$.00^{c}$	$.01^{c}$	.00
Drop out (SD5)								
Auditors	$.04^{a}$	$.28^{a}$	$.23^{a}$	$.23^{a}$	$.21^{a}$	$.23^{a}$	$.29^{a}$	$.40^{a}$
Failed	$.00^{b}$	$.00^{b}$	$.08^{b}$	$.16^{b}$	$.16^{b}$	$.18^{b}$	$.16^{b}$	$.34^{a}$
Passed	$.00^{b}$	$.00^{b}$	$.00^{c}$	$.00^{c}$	$.00^{c}$	$.00^{c}$	$.00^{c}$	$.06^{b}$

Note. Different superscripts indicate significant group differences depending upon occurrence of using SD behaviours across weeks (columns). Statistical significance is indicated by $p<.003$, with p-values adjusted using Bonferroni correction

Different superscripts in the table indicate the significant differences (p-value <.003, Fisher's exact) between learners' application of Session Distribution clusters in relation to their engagement and achievement groups across each week. For instance, application of SD0 cluster in week 1, was significantly different for Auditors and both Passed and Failed groups (different superscripts), while not significantly different between Passed and Failed groups (same superscripts). For week 3 and based on the application of SD0 cluster, all groups were significantly different from each other (different superscripts).

Auditors had a large proportion of No activity (SD4) across the course, followed by Low frequency, no large session (SD1) and Drop out (SD5). These proportions were generally stable across the course. Other clusters were present, but in very small proportions across the course.

Failed had a large proportion of both Low frequency (SD0) and Low frequency, no large session (SD1) across the course. At the outset of the course, Low frequency (SD0) had a higher proportion than Low frequency, no large session (SD1), but that changed at week four and persisted until the end of the course. Clusters No activity (SD4), Drop out (SD5) and High frequency, low large sessions (SD3) were moderately present throughout the course. The High frequency (SD2) cluster was present in small proportion across the course.

Passed had a large proportion of both Low frequency (SD0) and High frequency, low large sessions (SD3) across the course. Low frequency (SD0) had a higher proportion than High frequency, low large session (SD1) until week six, then High frequency, low large session (SD1) had the largest proportion for the last two weeks of the course. Low frequency, no large session (SD1) and High frequency (SD2) were moderately present across the course, with the first one having a higher prevalence in the first week and the final two weeks of the course. No activity (SD4) and Drop out (SD5) clusters were present in a very small proportion at the outset and end of the course.

Comparisons between engagement and achievement groups revealed interesting and significant group differences on session distribution across the course. The Low frequency, no large session (SD1) was often applied significantly different across groups and was more often in the following order across the course: Auditors > Failed > Passed. Conversely, the Low frequency (SD0), High frequency (SD2) and High frequency, low large sessions (SD3) were more often in the following order of engagement and achievement groups across the course: Passed > Failed > Auditors. Taken together, these findings indicate a trend of successful learners applying session distributions clusters that included at least a few weekly large sessions across the course, rather than organising their time to have mainly weekly small and medium sessions.

3.3.3.2. Session activity and engagement and achievement groups

Three visualisations presented in Figure 3-9 represent the percentage of session activity clusters for each engagement and achievement group across the weeks. Learners who left the subject in a specific week were not considered in the analysis of the later weeks.

In this phase, the frequency of each SA cluster for each engagement and achievement group in each week was calculated. Then the frequency data was used to examine the association between groups and all SA behaviours in each week of the course using the Fisher's exact test. This included detecting the triplet and pairwise group differences. In fact, Fisher's Exact test was applied on the 3*7 contingency table (3 engagement groups versus 7 SA clusters) for each week, that indicated an overall association between the three groups and the session activity behaviours over all weeks as presented in Table 3-7.

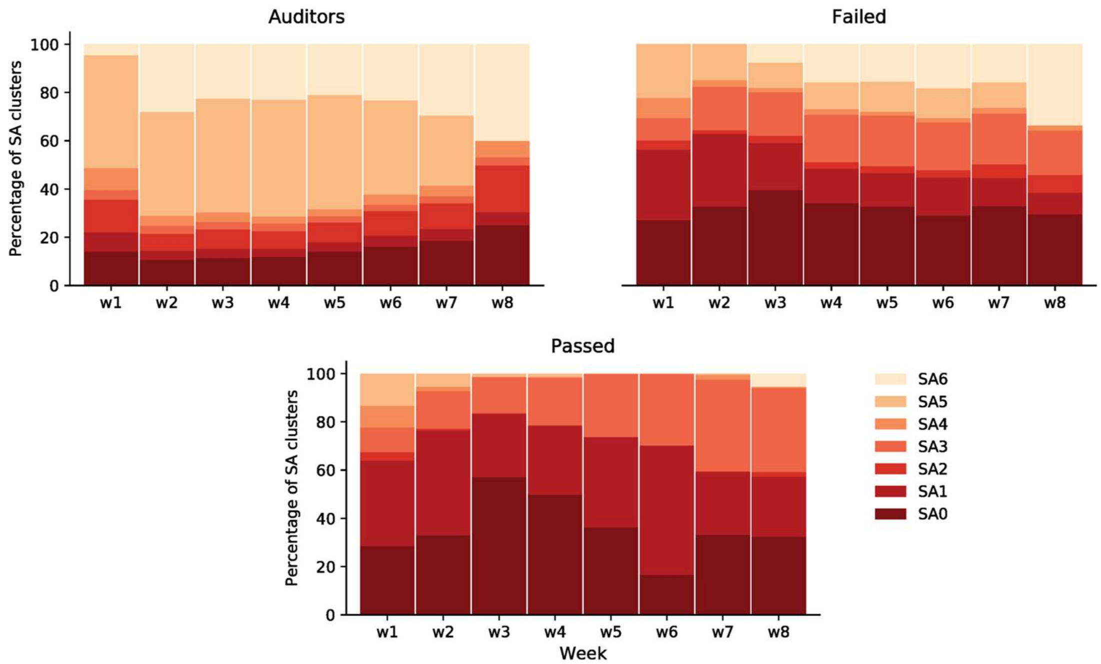


Figure 3-9. Percentage of SA clusters over weeks for each engagement and achievement group

Table 3-7. Association between engagement and achievement groups and SA Clusters

	Session Activity Clusters							
	W1	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**
All groups	**	**	**	**	**	**	**	**
Auditors vs. Passed	**	**	**	**	**	**	**	**
Auditors vs. Failed	**	**	**	**	**	**	**	**
Passed vs. Failed	*	**	**	**	**	**	**	**

Notes. * $p < .05$, ** $p < .001$.

Further analysis using Fisher's Exact test examined pairwise group differences related to the proportion of session activity clusters applied across the weeks. In this analysis, the frequency of existence and non-existence of each SA cluster for each engagement and achievement group in each week was calculated. Then this frequency data was used to examine the association between groups and each SA behaviour in each week of the course using the Fisher's exact test. This included detecting the pairwise group differences. In fact, Fisher test was applied based on the 2*2 contingency table (2 rows for pairwise engagement groups versus 2 columns for SA exists/ SA not exists). The adjusted alpha level of .002 was used for determining significant differences between groups in each week. These results are presented in Table 3-8.

Different superscripts in the table indicate the significant difference (p-value <.002, Fisher's exact) between learners' application of Session Activity clusters in relation to their engagement and achievement groups in each week of the course. For instance, application of SA1 cluster in week 1 was significantly different between Auditors and both Passed and Failed groups (different superscripts), while not significantly different between Passed and Failed groups (same superscripts). For week 1 and based on the application of SA4 cluster, groups were not different (same superscripts).

Auditors had a large proportion of No activity (SA5) and Drop out (SA6) clusters across the course. No activity (SA5) was more prevalent than Drop out (SA6) until week 7, with Drop out (SA6) being more prevalent in week 8. The two largest clusters with some activity across the course were Lecture (SA0) and Miscellaneous (SA2). Other clusters were present, but in very small proportions across the course.

Table 3-8. Group differences in the proportion of SA clusters per week

	Session Activity Clusters							
	W1	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**
Lecture (SA0)								
Auditors	$.14^{a}$	$.11^{a}$	$.11^{a}$	$.12^{a}$	$.14^{a}$	$.16^{a}$	$.19^{a}$	$.25^{a}$
Failed	$.27^{b}$	$.33^{b}$	$.40^{b}$	$.34^{b}$	$.33^{b}$	$.29^{b}$	$.33^{b}$	$.30^{a}$
Passed	$.29^{b}$	$.33^{b}$	$.57^{c}$	$.50^{c}$	$.36^{b}$	$.17^{a}$	$.33^{b}$	$.33^{a}$
Lecture mix (SA1)								
Auditors	$.08^{a}$	$.04^{a}$	$.04^{a}$	$.03^{a}$	$.04^{a}$	$.05^{a}$	$.05^{a}$	$.05^{a}$
Failed	$.29^{b}$	$.30^{b}$	$.20^{b}$	$.14^{b}$	$.14^{b}$	$.16^{b}$	$.12^{b}$	$.09^{a}$
Passed	$.35^{b}$	$.43^{c}$	$.26^{b}$	$.29^{c}$	$.37^{c}$	$.54^{c}$	$.26^{c}$	$.25^{b}$
Miscellaneous (SA2)								
Auditors	$.14^{a}$	$.07^{a}$	$.08^{a}$	$.07^{a}$	$.08^{a}$	$.10^{a}$	$.11^{a}$	$.19^{a}$
Failed	$.04^{b}$	$.02^{b}$	$.03^{b}$	$.03^{b}$	$.03^{b}$	$.03^{b}$	$.06^{b}$	$.07^{b}$
Passed	$.03^{b}$	$.01^{b}$	$.00^{c}$	$.00^{c}$	$.00^{c}$	$.00^{c}$	$.00^{c}$	$.02^{c}$
Assessment (SA3)								
Auditors	$.04^{a}$	$.03^{a}$	$.03^{a}$	$.03^{a}$	$.03^{a}$	$.03^{a}$	$.03^{a}$	$.03^{a}$
Failed	$.09^{b}$	$.18^{b}$	$.18^{b}$	$.20^{b}$	$.21^{b}$	$.20^{b}$	$.21^{b}$	$.18^{b}$
Passed	$.10^{b}$	$.16^{b}$	$.15^{b}$	$.20^{b}$	$.26^{b}$	$.30^{c}$	$.38^{c}$	$.35^{c}$
Guide (SA4)								
Auditors	$.09^{a}$	$.04^{a}$	$.04^{a}$	$.03^{a}$	$.03^{a}$	$.04^{a}$	$.04^{a}$	$.07^{a}$
Failed	$.08^{a}$	$.03^{a}$	$.02^{b}$	$.02^{a}$	$.01^{a}$	$.02^{b}$	$.02^{a}$	$.02^{b}$
Passed	$.09^{a}$	$.02^{a}$	$.00^{b}$	$.01^{a}$	$.00^{a}$	$.00^{b}$	$.02^{a}$	$.01^{b}$
No activity (SA5)								
Auditors	$.47^{a}$	$.43^{a}$	$.47^{a}$	$.48^{a}$	$.47^{a}$	$.39^{a}$	$.29^{a}$	.00
Failed	$.22^{b}$	$.15^{b}$	$.11^{b}$	$.11^{b}$	$.12^{b}$	$.12^{b}$	$.11^{b}$	.00
Passed	$.13^{c}$	$.05^{c}$	$.01^{c}$	$.01^{c}$	$.00^{c}$	$.00^{c}$	$.01^{c}$	.00
Drop out (SA6)								
Auditors	$.04^{a}$	$.28^{a}$	$.23^{a}$	$.23^{a}$	$.21^{a}$	$.23^{a}$	$.29^{a}$	$.40^{a}$
Failed	$.00^{b}$	$.00^{b}$	$.08^{b}$	$.16^{b}$	$.16^{b}$	$.18^{b}$	$.16^{b}$	$.34^{b}$
Passed	$.00^{b}$	$.00^{b}$	$.00^{c}$	$.00^{c}$	$.00^{c}$	$.00^{c}$	$.00^{c}$	$.05^{c}$

Note. Different superscripts indicate significant group differences depending upon occurrence of using SA behaviours in each week of the course (columns). Statistical significance is indicated by $p<.002$, with p-values adjusted using Bonferroni correction

Failed had four main clusters across the course: Lecture (SA0), Lecture mix (SA1), Assessment (SA3), No activity (SA5) and Drop out (SA6). The prevalence of each cluster fluctuated across the course. Overall, the Lecture (SA0) cluster was the largest cluster across the course, with the exception of week 1. Lecture mix (SA1) had a larger proportion than Assessment (SA3), but that changed in week 4, persisting until the end of the course. No activity (SA5) was more prevalent than Drop out (SA6) until week 4, with Drop out (SA6) being more prevalent until the end of the course. Both Miscellaneous (SA2) and Guide (SA4) were present in very small proportions across the course.

Passed had three main clusters across the course: Lecture (SA0), Lecture mix (SA1), and Assessment (SA3). The prevalence of each cluster fluctuated across the course. Lecture mix (SA1) was the most prevalent in weeks 1 and 2, changing to Lecture (SA0) in weeks 3 and 4, then changing back to Lecture mix (SA1) in weeks 5 and 6. Assessment (SA3) was then the most prevalent cluster in weeks 7 and 8. Other clusters were present, but in very small proportions across the course.

Comparisons between engagement and achievement groups revealed significant and interesting group differences on session activities across the course. The groups applied Lecture (SA0) and Lecture mix (SA1) in different ways across the course. While Auditors mainly applied Lecture (SA0) rather than Lecture mix (SA1), Failed applied a higher proportion of Lecture mix (SA1) than Lecture (SA0), particularly at the beginning of the course. Passed, on the other hand, alternated the application between these two clusters throughout the course; at times it applied more Lecture (SA0), and other times more Lecture mix (SA1). Regardless of these differences, Passed applied more Lecture mix (SA1) than both Failed and Auditors for most of the course. Moreover, Passed applied a higher proportion of their session time to Assessment (SA3) than Failed towards the end of the course. These findings indicate that successful learners change the focus of their activities within sessions across the course, while still dedicating time to weekly sessions that include a variety of activities rather than focusing on one main activity.

3.4. Discussion

This study investigated how persisting learners organize their workload into sessions across a MOOC, considering both the macro (RQ1) and micro (RQ2) aspects of session behaviour, and its relationship with engagement and achievement groups (RQ3). Audit trails data was used to examine the research questions.

For the first research question, meaningful patterns were identified for session distribution across the course for all learners, as expected. The session distribution had six clusters with varying frequency and length of sessions: Low frequency (SD0), Low frequency, no large sessions (SD1), High frequency (SD2), High frequency, low large sessions (SD3), No activity (SD4), and Drop out (SD5). The analysis for the second research question also revealed meaningful patterns for session activity. Across the course, session activities had seven clusters with most of the session time dedicated to one learning resource: Lecture (SA0), Lecture mix (SA1), Miscellaneous (SA2), Assessment (SA3), Guide (SA4), No activity (SA5), and Drop out (SA6). These results indicate that people have varied time and study management strategies (see Section 2.3) to organise their time in terms of sessions while learning in a MOOC, which is not a surprising finding. However, the analysis for the third research question provided insights into how some patterns of sessions (reflecting various time management strategies) may be more effective in promoting learning than others.

In responding to the third research question, the patterns of session distribution and session activity were shown to be related to learners' group of engagement and achievement in the course. The session distribution analyses showed that successful learners had a higher number of sessions, regardless of their size, compared to failed learners and auditors. This indicates that, overall, having more weekly sessions was related to high engagement and performance. This supports the importance of general high engagement with the course as an important indicator of performance in online learning (Gardner & Brooks, 2017). If learners are not regularly showing up from the outset of the course, they are at a disadvantage in relation to course achievement. Successful learners also typically included at least a few large weekly sessions across the course, rather than organising their time to have mainly small and medium weekly sessions. Small and medium weekly sessions was the main pattern of behaviour associated failed learners and auditors. Previous research suggests that this behaviour may have contributed to higher performance as longer sessions can allow learners to more fully engage with challenging tasks, positively impacting their learning, which is less common in smaller sessions (Son & Metcalfe, 2000).

Session activity was also clearly related to engagement and achievement. During their sessions, successful learners dedicated more time to access a mixture of learning resources, represented by the cluster Lecture mix (SA1). Based on previous research, this session behaviour with a varied activity focus could reflect students using a range of different learning strategies within a session (e.g., Fincham et al., 2018; Gasevic et al., 2017; Jovanovic et al., 2017). That is, if students engage in a greater number and varied types of activities during their sessions, they are more likely to employ a variety of cognitive and resource management strategies (see Pintrich, Smith, Duncan, & Mckeachie, 1991 and Section 2.3). In a previous study, for example, researchers found that a session with a mixture of activities, mainly viewing lectures and taking quizzes, also included metacognitive actions, such as accessing the dashboard which contained real-time feedback on access to content and assessment and the page with the learning schedule and objectives (Jovanovic et al., 2017).

The findings also suggest that successful learners changed the priority of activities within sessions across the course. This suggests they strategically divided their time to focus on whatever activity they perceived as most effective to contribute to their course achievement at a particular moment (Jovanovic et al., 2017; Metcalfe & Kornell, 2005). For example, in the current study successful learners dedicated most of their session time towards the end of the course to completing assessments. This is particularly interesting considering that this behaviour was also found in week 7, when there were no graded quizzes, which suggests that students were revisiting previous assessments in that week.

Taken together, the findings suggest that session distribution and session activity are important behaviours related to engagement and performance in a MOOC, and also confirm the importance of varied students' study and time management strategies (see Section 2.3) in online learning. Two major applications of the findings were identified. First, educators and designers could apply this approach to identify patterns of session behaviours in their courses. This would allow them to identify which session behaviours require support – and (most importantly) when – so they can select relevant interventions to include in their courses. Second, educators and designers could implement interventions to support learners' time management skills and expect a positive impact on learners' session behaviour. This is particularly applied to session distribution. Based on previous research, such interventions could include providing study plans and reflections (Davis, Chen, van der Zee, Hauff, & Houben, 2016) and tools for time management (Nawrot & Doucet, 2014; Tabuenca, Kalz, Drachsler, & Specht, 2015). For example, use of the LearnTracker app (Tabuenca et al., 2015), which assists learners to track and review time dedicated studying and planning ahead, has resulted in improvement of time management skills measures.

This study has three main limitations commonly found in MOOC research. First, it was conducted in only one MOOC. Replication considering other MOOCs, particularly with different foci and learning designs (e.g., Brooker et al., 2018; Coffrin et al., 2014), would contribute to a better understanding of the different contextual influences on session behaviours, and their relationship with performance. Second, there may be different approaches to examining session behaviour. These include, for example, considering non-persisting learners in the analyses and running cluster analysis considering both session distribution and activity. Another approach could be focusing only on the constant number of students who did not drop out of the course. In addition, a qualitative investigation of learners' contribution to the discussion forum and social media could provide further insight into learners' skills, such as content analysis or natural language processing. Finally, some learners' personal factors related to behaviour and performance in MOOCs were not examined as they were beyond the scope of this study. These include, for example, learners' individual differences, such as age, level of education and course goals (e.g., Kizilcec et al., 2017), previous online learning experience (e.g., (Astani, Ready, & Duplaga, 2010)), and their motivation prior and throughout the course (e.g., de Barba et al., 2016).

3.5. Chapter summary

This chapter provided a better understanding of the importance and meaningfulness of session behaviour in online learning, particularly in MOOCs. This investigated how learners organize their workload into sessions across a MOOC, and its relationship with engagement and achievement groups. Based on the results, session distribution and session activity were important behaviours related to engagement and achievement in a MOOC. Successful learners had more frequent and longer sessions across the course, mixed up activities within sessions, and changed the focus of activities within sessions across the course. This research contributes to the field of educational technology research with the use of a temporal approach combining learning analytics and learning design as a way to provide a meaningful interpretation of learners' behaviour. Implications for both practice and research are related to using session behaviours identified in successful learners as a benchmark for potential interventions to underperforming learners. Due to the temporal aspect of the analyses, these findings can inform both types of possible effective interventions (e.g., time management) and when the intervention would be most useful (e.g., throughout a course or only at the outset of a course). Another aspect of students' behaviour that has not been given much attention in the literature, is the importance of sequential representation of students' behaviour. This could better reveal the effect of the process students used during learning, and thus, is the topic of next chapter (Chapter 4).

Chapter 4

Investigation of Students' Sequential Behaviour using Temporal Analysis

This chapter utilises a range of learning analytics techniques to evaluate the importance of the sequential nature of students' behaviour on their learning outcome. In particular, it focuses on relating students' task-level behaviour to their assessment task outcome rather than the final achievement. This gives more insight regarding students' progress over time. The content of this chapter is adapted from the following paper:

Donia Malekian, James Bailey, Gregor Kennedy. Predicting Students' Assessment Readiness in Online Learning Environments: The Sequence Matters, To appear in *Proceedings of the 10th International Conference on Learning Analytics and Knowledge (LAK'20),* March 23-27, Frankfurt, Germany, 2020.

4.1. Introduction

As explained in Section 2.1, Massive Open Online Courses (MOOCs) have made a significant contribution to the landscape of online learning (Liyanagunawardena et al., 2013). They are fully online courses usually offered through a partnership between universities and online learning platforms to anyone in the world with Internet access. As discussed in Section 2.6, during recent decades, many efforts have been made to understand and promote experiences of learners in these environments. The current study follows this line of research.

As discussed in Section 2.6, two directions of research with this purpose focus on modelling students' behaviour and relating it to their learning outcome; The first approach focuses on creating models of students' behaviour using prediction methods and subsequently using them to predict their learning outcome (Corbett & Anderson, 1994; He et al., 2015), and the focus of the second approach is on profiling or grouping students based on various factors with the aim of determining the relation of those factors with students' learning outcome. In modelling learners' behaviour, it is important to consider the different levels of data available and the levels at which their activities are structured (Nguyen, Huptych, & Rienties, 2018). Previous analytics-based research in online learning has mostly focused on predefined aggregated factors of learners' fine-grain behaviours (learning actions) and evaluating their effect on their learning outcome. However, a sequential representation of students' learning interactions might be more useful in showing how students' learning approaches and behaviours over time impacts on their learning processes and outcomes.

Among the prediction-based studies, few considered the sequence of learners' behaviour into account. These studies often focused on modelling students based on the concepts that they learned, and the result of prediction was considered as an indication of students' knowledge development and mastery on a topic (for more detail see Section 2.6.1). However, success is not just about the content. It also requires appropriate skills to access and work with the most relevant content. In addition, numerous decisions could be made regarding the choice of algorithms. While comparing and selecting an appropriate algorithm is a challenging task, switching between algorithms usually results in a minimal change (Baker & Inventado, 2014), whereas a proper representation of data can lead to a considerable difference and can be more challenging. To the best of knowledge, comparing the effect of modelling students based on the sequential and aggregated representations of the same task-level activities in their behaviour within MOOCs has not yet been investigated. Such analyses could raise the awareness of EDM and LA research regarding the need to account for the sequential nature of data when studying students' behaviour.

As discussed in Section 2.6, there exist many studies with the aim of predicting students' task/course outcome in online learning environments. However, to the best of our knowledge, comparing the effect of modelling students based on the sequential and aggregated representations of the same activities in their behaviour within MOOCs has not yet been investigated. Such analyses could reveal the importance of considering the sequential nature of students' behaviour for the application of student modelling. In addition, investigating students' preparation behaviour before submitting assessments and examining the most effective behaviour to be considered (more historic versus most recent activities prior to submission) and its effect on student task performance is under-studied in the literature. This analysis could reveal the importance of considering the overall process used by students throughout the course rather than the most recent activities prior to task submission.

In addition, among the studies that have focussed on the identifying the relation between various factors of learners' behaviour and learning outcomes, most focus is on pre-defined aggregated factors (see more details in Section 2.6.2). However, utilising a data-driven approach for extracting the distinct patterns of students' learning behaviours (measures of them) might reveal emergent behaviours which were not identified by domain experts. Additionally, considering the sequential representation of students' behaviour might better disclose the effect of the process students used during learning (for more detail about students' learning processes see Section 2.3).

Considering the mentioned challenges, the main contribution of this study is to relate students' sequence of learning behaviour to their learning outcome using

LA and EDM methods. This study considers assessment task outcomes as a proxy for learning outcomes rather than students' final achievement in the course examined in most studies. This gives more insight regarding students' progress over time. Assessment tasks often form a critical component of digital learning environments, as they allow students to reflect on their own level of understanding and provide a mechanism to determine whether or not students have mastery of the course materials (Gibbs & Simpson, 2005). Educational research has shown that when students are unsuccessful in assessment, this can have a negative impact on their confidence in their ability to succeed, which in turn can lead to disengagement (Dembo & Seli, 2012). Thus, having insight regarding students' performance on assessment tasks (i.e. prediction, monitoring) could provide an opportunity for developing an adaptive learning environment aware of students' progress and readiness for online assessment tasks, and potentially delay or postpone tasks in cases where students are seen as not ready. Moreover, a deeper understanding of students' behaviour before completing an assessment and knowing whether these behaviours differ for students with a high or low level of performance, could be used to tailor support and interventions that could maximise the likelihood of students' success.

Thus, the first goal of this study was to utilise learning analytic techniques to build a students' profile (prediction model) – in two forms of aggregate and sequential – to predict their performance in assessment tasks (as a proxy of readiness). Monitoring of students' knowledge based on the concepts they have learned – as determined by their performance in assessments – is well-studied in the literature. This study extends this work by considering the kinds of study behaviours students adopt in an online course as the foundation of a student model. A range of sequential and non-sequential machine learning algorithms was employed to determine whether a particular profile would be a stronger representative of students' outcome in assessment tasks.

The second goal was to explore whether the most recent activities prior to submissions are more influential on assessment outcomes compared to incorporating the more historic activities applied by students. This analysis could reveal the importance of considering the overall process used by students throughout the course rather than the most recent activities prior to assessment submissions.

The final goal was to explore students' assessment preparation behaviours using sequential pattern mining. The aim was to explore and expose underlying factors that might influence students' assessment results. These factors could not only illustrate the different ways in which students prepare for assessment tasks but could also identify behaviours that need support to improve students' future preparation. Overall, in this chapter the following research questions are explored:

Research question 1. Can useful prediction models be developed for forecasting students' readiness for assessment tasks (i.e. prediction of their success in each assessment submission)? Does it matter if the sequential nature of students' task-level activities is considered in the model rather than aggregated measures?

Research question 2. What is the impact of considering the most recent activities prior to submission in the models compared to incorporating more historic activities applied by students?

Research question 3. Are there differences in assessment preparation behaviours that lead to high or low performance?

The results show it was possible to successfully predict student readiness for assessment tasks, particularly if the sequential aspects of students' behaviour are represented in the machine learning model. Moreover, the findings confirm the importance of the overall process students applied in the course as a significant indicator of learning outcome rather than only the most recent activities before submissions. Additionally, the results show how some behavioural patterns reflecting specific learning strategies may be more effective in promoting or undermining learning than others.

4.2. Approach

4.2.1. Study context and participants

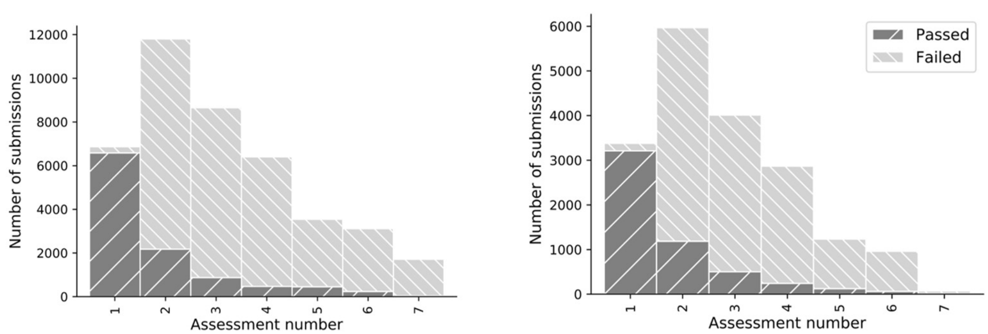
This study focuses on two offerings of a MOOC called Discrete Optimization (Disc Opt.) provided by The University of Melbourne in 2013 and 2014. Anyone who had access to a computer with an internet connection could enrol in this course. The registration process consisted of only providing a name and email address. Registration was open while the course was running. There were no fees for registration unless students wanted to be on the accreditation track. All the course materials were available from day one. This MOOC was a graduate-level course consisting of nine weeks of material presented in an open curriculum structure. It included seven summative assessments, each of which contained a bank of questions. Students needed to earn a passing grade (0.7/1) in each assessment to complete the course. The student's final outcome for an assessment was based on their best submission score for that assessment. The content of assessments was slightly different across both offerings. A summary of the two offerings and students' participation in assessment tasks are presented in Table 4-1.

Table 4-1. Summary of the two offerings of Discrete Optimisation

	1st offering	**2nd offering**
Number of students enrolled	51,306	33,975
Number of assessments	7	7
Number of students engaged in assessments	6,664	3,258
Number of students dropping-out after assessment failure	2610	332
Number of assessment submissions	42,094	18,495
Number of failed assessment submissions	31,285	13,159
Number of students completed	795	322

In the first offering of this course, of the 51,306 students who initially enrolled, around 12% (6,664) engaged in the assessments. From those, 39% (2610) dropped out of the course right after a failure in an assessment. No certain reason(s) for this abrupt disengagement can be provided, but one possible explanation is that failure experiences influenced some students to doubt their abilities to succeed, which then resulted in their disengagement (Dembo & Seli, 2012).

In this course, students were permitted to make an unlimited number of submissions for each assessment. This study has restricted the analysis to the first ten student's submissions per assessment. Moreover, for the successful submissions, this has been limited to those that finalised the assessments' outcome. This means that if a student had multiple successful submissions on an assessment, only the first successful one was considered. Statistics summarising the total number of passed and failed submissions in each assessment of the two MOOCs are shown in Figure 4-1. As can be seen, in both offerings, most submissions were passed for the first assessment (as it was considered to be an easy task). In contrast, for all the other assessments, less than half of the submissions were successful. Ethics for the use of learning analytics for research purpose was covered by Coursera's Terms and Conditions.



(a) (b)

Figure 4-1. Number of passed and failed submissions in the a) 1st offering, and b) 2nd offering of Discrete Optimisation

4.2.2. Data processing and measures

A standard MOOC usually contains three different instructional events for students: content-related resources (e.g., video lectures, slides), assessments (e.g. assignments, and quizzes), and collaboration tools (e.g. discussion forums) (Brinton et al., 2015). The focus of the analysis in this chapter is on students' fine-grain learning interactions in response to these instructional events, particularly those that help students achieve course outcomes and reflect particular kinds of approaches to online study. This involves accessing and downloading lecture, re-accessing and re-downloading the previously accessed lecture, accessing the discussion forums, and submitting assessments. A sample representation of a student's list of ordered activities is shown in Figure 4-2. This figure has been referred repeatedly throughout the chapter as the sequential analysis are developed. For simplicity, the activity names are mapped to IDs as provided in the figure caption.

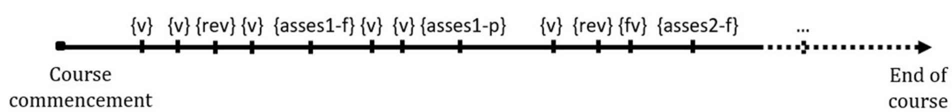


Figure 4-2. A sample representation of a student's ordered list of learning activities in a MOOC. v= lecture viewed or downloaded, rev= lecture reviewed or redownloaded, fv= forum viewed, asses#-p= passed submission on a specific assessment#, asses#-f= failed submission on a specific assessment#

Following subsections describes how learners' interactions and learning outcome are represented to assist on answering the research questions. This includes; learners' learning activities before assessment *submission* in aggregated form, and in sequential form, *submission* outcome, learners' learning activities before *completing* assessments in sequential form, and *final assessment* outcome.

4.2.2.1. Aggregated representation of learning activities before an assessment submission

In this study, learner's learning interactions before each assessment submission s are represented by a d-dimensional vector of features ($s \in \mathbb{R}^d$) as presented in Table 4-2, started from the beginning of the course and ended before that submission. This study aimed to build the analysis irrespective of the specific assessment. Thus, for each submission, the previous submissions are divided into two groups of; submissions on the same assessment, and on the other assessments with their respective results as represented in Table 4-2.

Table 4-2. Aggregated profile for each student's submission

Aggregated profile
Number of lectures viewed or downloaded
Number of lectures reviewed or redownloaded
Number of forums viewed
Number of failed submissions on the same assessment
Number of passed submissions on the other assessments
Number of failed submissions on the other assessments

4.2.2.2. Sequential representation of learning activities before an assessment submission

In this study, learner's learning activities before each assessment submission are represented by a variable sequence of t ordered activities $(X_1, \ldots, X_{t-1}, X_t)$ conducted by student started from the beginning of the course and ended before that submission. ($X_t \in$ Learning activities). Table 4-3 presents the extracted sequences for the example provided in Figure 4-2 for illustration purpose.

Table 4-3. Sequences before each *submission* and their outcome (class label) based on the example provided in Figure 4-2. Assess= assessment, Sub= submission, casses-f= failed submission on the same assessment, oasses-f= failed submission on the other assessments, oasses-p= passed submission on the other assessments

Assess #	**Sub #**	**Sequence before each sub (sub outcome)**	**Finalised Sequence**	**Sub outcome**
1	1	v v rev v (asses1-f)	v v rev v	failed
1	2	v v rev v asses1-f v v (asses1-p)	v v rev v casses-f v v	passed
2	1	v v rev v asses1-f v v asses1-p v rev fv (asses2-f)	v v rev v oasses-f v v oasses-p v rev fv	failed

As can be seen, each submission is assigned a sequence that starts from the beginning of the course and ends before that submission. Similar to the aggregated profile, a distinction has been made between submissions on the "same assessment" and "other assessments". As can be seen in Table 4-3, they are mapped to "casses" and "oasses" respectively in the "Finalised Sequence" column.

4.2.2.3. Submission outcome

This corresponds to the result of each submission on assessment (typically passed or failed) for a student. As mentioned before, for each assessment, students needed a mark of at least 70% in one of the submissions to pass. Therefore, in the analysis, the pass mark for each submission on an assessment was considered to be 70% of the total assessment score.

4.2.2.4. Student Readiness

This corresponds to the prediction of the assessment submission outcome by the prediction model. In other words, if the model predicts a submission to be successful, then the student is considered to be ready for the assessment.

4.2.2.5. Sequential representation of learning activities before completing an assessment

This study also aims to explore students’ behaviour before *completing* assessments. This behaviour is defined as student’s learning interactions before the submission that *finalised* the outcome for an assessment task (which is the last failed submission or the first passed submission). In fact, each student that engaged with an assessment (with one or more submissions) was assigned one sequence for that assessment. The beginning of the sequence was considered to be before the first submission for the same assessment and started with the latest submission for another assessment. If it was the first assessment attempted through all the course, the sequence started from the beginning of the course. Table 4-4 presents the extracted sequences for the example provided in Figure 4-2 for illustration purpose.

Table 4-4. Sequences for each *assessment* and their respective outcome based on the example provided in Figure 4-2. Assess= assessment, casses-f= failed submission on the same assessment

Assess #	Sequence before completing each assess (assess outcome)	Finalised Sequence	Assess Outcome
1	v v rev v asses1-f v v (asses1-p)	v v rev v casses-f v v	passed
2	v rev fv (asses2-f)	v rev fv	failed

As can be seen, a sequence started with the first activity after the other assessment submission and ended before the submission that finalised the outcome for the assessment. Since the aim is to gain insight about the preparation behaviour irrespective of the specific assessment, for each sequence, the submissions on the “same assessment” and “other assessments” are mapped to “casses” and “oasses” respectively as shown in the “Finalised Sequence” column of Table 4-4.

4.2.2.6. Final Assessment outcome

This corresponds to the *best result of submissions* on an assessment (typically passed or failed) for a student. Similarly, the pass mark for each assessment was considered to be 70% of the total assessment score.

4.2.3. Data analysis

The data processing and analysis performed in this study are summarised in Figure 4-3. Following subsections describe the three-step analysis undertaken in this study to assist on answering the research questions.

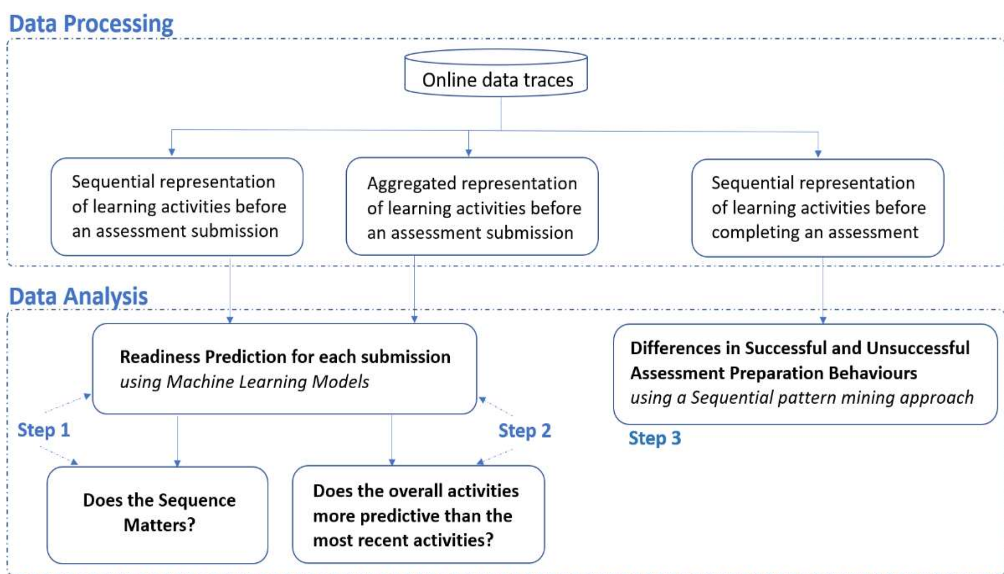


Figure 4-3. A high-level representation of data processing and analysis in this study

4.2.3.1. Readiness prediction for each submission: does the sequence matter

Using predictive models, a student profile was built based on their past history of learning activities, to determine which students were not ready for their assessment submission. Once developed, this profile was then used to predict students' success for each *submission* they make on an assessment task. This problem was cast as a binary classification task, where a classifier was developed based on a student's past learning activities before submitting an assessment in two forms of aggregated (Section 4.2.2.1) and sequential (Section 4.2.2.2), and the class label was student's submission outcome (Section 4.2.2.3).

This study first evaluated the effectiveness of the aggregated representation of students' learning activities using a non-sequential machine learning classifier. Then the prediction power of a sequential machine learning model was examined to investigate the impact of modelling the sequential nature of students' behaviour. The first offering of the MOOC was used to train the model. Using this model, at each student's submission in the second offering of the MOOC, the probability of success was predicted to determine the effectiveness of the models.

4.2.3.2. Algorithms

Neural Network (NN) models were utilised for the evaluation purpose. NNs are popular for learning complex relationships from data in domains such as medicine (Bertolaccini et al., 2017; Shaikhina & Khovanova, 2017), security (Shone et al., 2018), loan applications (Saha & Waheed, 2017), speech recognition (Zhang et al., 2017), and natural language processing (Goldberg, 2017). There are, thus far, a limited number of studies employing NNs for tasks in learning analytics. As an example, Okubo et al. (Okubo et al., 2017), confirm the effectiveness of NNs for predicting students' final grade compared to regression-based models (for more details about NNs see Section 2.5.3.1.1). This study explored the utility of the aggregated profile for assessment readiness prediction, employing a basic three-layer structure NN with all layers fully connected - a Multiple Layer Perceptron (MLP). This is a simple proof of concept structure, and more complex structures (e.g. with more layers) are possible. the output was a prediction model learned for all the submissions. In order to investigate the impact of the sequential nature of students' activities on the prediction task, a temporal prediction model was used based on a popular variant of Recurrent Neural Networks - Long Short-Term Memory (LSTM). An LSTM is capable of fitting patterns from data across both long- and short time scales. This model used the sequential form of data as the input. The output was a prediction model learned for all the submissions (for more details about MLP and LSTM see Section 2.5.3.1.1).

For this analysis, N was the number of assessments in a course, and there were n submissions for each assessment $s_{i,}$ $1 \leq i \leq N.$ Each assessment i was defined by a set of submissions, $s_{i,} = \{s_{i1}, s_{i2}, \dots, s_{in}\}$. For each submission s_{ij}, a class label was defined as passed (y = 1) or failed (y = -1). Thus each assessment was given by$(s_i, y_i) = [(s_{i1}, y_{i1}), (s_{i2}, y_{i2}), \dots, (s_{in}, y_{in})]$. The binary classification algorithm generated a probability estimate for each class by fitting a single model on all assessments' submissions, $\{(s_{11}, y_{11}), (s_{12}, y_{12}), \dots, (s_{21}, y_{21}), (s_{22}, y_{22}), \dots\}$.

4.2.3.3. Experimental configurations

For both MLP and LSTM, binary cross-entropy was used as a loss function and models were trained with mini-batch stochastic gradient descent. For determining the best parameters, grid search hyperparameter optimization (Bergstra & Bengio, 2012) was used. In this technique, a set of parameter values for mini-batch size, the number of hidden dimensions (for MLP) and memory units (for LSTM) was defined. Then the model was trained using all combinations of those parameters and the solution that minimized the cost function was selected (Bergstra & Bengio, 2012). As there existed relationships between submissions, this study considered the temporal order in which submissions were made when performing the cross-validation in parameter optimisation. For this purpose, the walk-forward cross-validation approach was used (López de Prado, 2018), in which multiple split points were selected in the ordered list of observations for determining the train and test data. Additionally, early stopping was used on the number of epochs to avoid overfitting and unnecessary computations. For doing so, 20% of the training data at each fold of parameter optimisation was used as a validation set, and then the training in each fold stopped when the validation error started to increase. Both networks were implemented using the Keras framework with TensorFlow™ back-end.

4.2.3.4. The most recent activities prior to submission and the submission outcome

In this study, the models (MLP and LSTM) were built based on the variable length of learning interactions in aggregated (Section 4.2.2.1) and sequential (Section 4.2.2.2) forms before submissions. Having these models, then their prediction performance was compared to examine whether the most recent activities prior to submission were more predictive of submission outcome (Section 4.2.2.3) in each of the algorithms rather than including the more historical activities.

4.2.3.5. Differences in successful and unsuccessful assessment preparation behaviours

This study investigated students' behaviour and their approach to preparing for assessment tasks for each of the MOOC offerings. It assessed, through a sequence mining approach, whether students exhibited systematic behavioural patterns before accomplishing assessment tasks (before and in between submissions), whether these patterns differed for students with a high and low level of task outcome, and whether they can be mapped on to interpretable learning behaviours. Overall, the aim was to relate students' assessment preparation behaviour to their final assessment outcomes. Final assessment outcome and preparation behaviour were defined in Section 4.2.2.6 and Section 4.2.2.5 respectively.

In this study, learners' sequences of behaviour were categorised into two groups based on their final assessment outcome: 1) sequences that ended with a passed submission (Passed group), and 2) sequences that ended with a failed submission (Failed group).

Then, a sequential pattern extraction method (Baker & Yacef, 2009) was employed to identify the distinct behavioural patterns for the sequences. Sequential pattern mining is an analytical technique to identify all common subsequences of n-item in a sequence dataset that the proportion of their occurrence to the total number of sequences (called support) is more than a threshold. In other words, for each subsequence of n-item, support is generated, which determines the proportion of sequences that have enacted that subsequence at least once.

In summary, for each of the course offerings, all the subsequence of n-item (are called n-item in this study for the simplicity) were extracted from every sequence. For instance the subsequences of 3-item from the finalized sequences presented in Table 4-4, are {<v v rev> <v rev v> <rev v casses-f> <v casses-f v> <casses-f v v> <v rev fv>}. Each unique n-item was then subject to the following analysis:

- The support of each n-item within each group of sequences (Passed/Failed) was computed. This allows comparison across groups as the groups might have different numbers of sequences.
- The most frequent n-items in at least one of the groups were retained as patterns for further investigation.
- The statistical significance of each frequent pattern was computed, using a chi-square test (William, 1952) on the 2*2 contingency table (exists/not exists versus Passed/Failed). The chi-square test provides a measure of the statistical significance of the association between each pattern and the assessment outcome. A Bonferroni correction for multiple testing was applied to the result of the chi-square test to avoid Type I error.
- The patterns that were highly associated with outcome were further explored to determine the direction of association with the assessment outcome. This was performed based on comparing the fraction of each pattern in the Failed and Passed groups of sequences. In this way, the top patterns associated with the successful outcome were considered to be those that were more frequent (higher support) in the Passed group compared to the Failed group, ordered by their significance of association with the assessment outcome (based on smallest p-values).
- The top patterns in each group that were highly associated with assessment outcome were then explored to determine whether they can be related to interpretable learning behaviours.

Comparison of the two MOOCs in terms of their influential patterns; This study also compared the influential patterns between the two MOOCs based on their significant association with the assessment outcome. This comparison could reveal if the two MOOCs were similar in terms of their prominent patterns. For this purpose, the Spearman Rank-order correlation (Daniel, 1990) was utilised that shows the association between the ranked value of two variables. In the case of this study, the correlation between the patterns' rank in the two MOOCs was examined.

4.3. Results

The analytic approach described above was used with two offerings of a MOOC and is presented in three parts in this section.

4.3.1. Readiness prediction for each submission: does the sequence matter

The first set of analyses report on the ability of the models to predict assessment readiness. The results are reported using the two metrics of accuracy and area under the ROC curve (AUC), a widely used metric in machine learning which is informative for the scenario of imbalance between classes (Ling, Huang, & Zhang, 2003). The models were trained based on the first offering of the MOOC (with three-fold cross-validation for parameter optimisation) and evaluated based on the second offering of the course.

The outcome of the predictive performance of MLP and LSTM is shown in Table 4-5. In this study, the utility of the aggregated profile is also investigated based on a baseline model of Majority Classifier (MC) and Logistic Regression (LR) to make sure the result obtained using MLP was not misspecified. The AUC and accuracy results of both MLP and LSTM algorithms demonstrate a high success rate in predicting student readiness for the second offering (AUC score > 85.0% and accuracy>83.7%). A comparison of the two models illustrates that LSTM outperformed MLP in both AUC and accuracy. This suggests that taking the sequential nature of data into account was more beneficial.

Table 4-5. Comparison of MC, LR, MLP, and LSTM performance based on AUC and accuracy. The models were trained based on the first offering of the MOOC and evaluated based on the second offering.

	MC	LR	MLP	LSTM
Accuracy	71.1	83.3	83.7	**87.1**
AUC	50.0	84.1	85.0	**86.4**

4.3.2. The most recent activities prior to submissions and the submission outcome

This study examined whether building the models based on the various numbers of recent activities before submission would make any difference in their prediction power. This analysis was limited to at most 77 recent activities. This number was the median number of interact done before submissions. Figure 4-4 presents the result of this analysis based on AUC. In this figure, the x-axis corresponds to the upper limit of the number of recent activities prior to the submissions that the models were built on. As can be seen, all algorithms worked better when considering more activities. Furthermore, LSTM worked better than MLP in almost all cases. This could mean that the overall processes students applied through the course had more impact on the submission outcomes rather than only the recent activities prior to submissions.

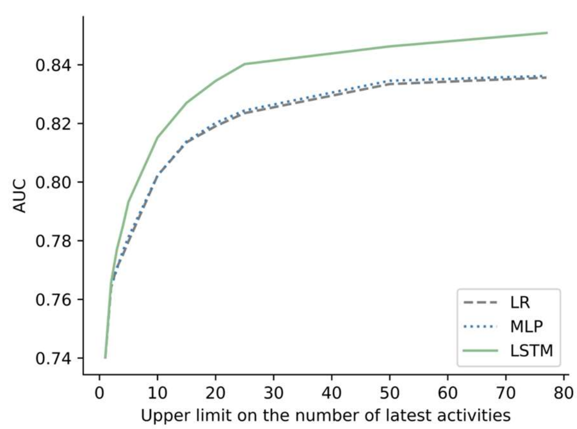


Figure 4-4. Comparison of LR, MLP, and LSTM performance based on AUC for the variable length of activities before submissions

4.3.3. Differences in successful and unsuccessful assessment preparation behaviours

It was possible to collect 15,748 and 7,544 student sequences for all the assessments of the first and the second offerings of the course respectively. Using sequential pattern extraction method, a large number of subsequences (n-items) of various length ($n \in \{1, 2, .., 10\}$) and support were extracted for each group of sequences in each of the offerings. Then for each offering, only the subsequences having more than 20% support in at least one of the groups were retained as potential patterns in order to contrast them across groups with different levels of the assessment outcome. The maximum subsequence length (n) was considered as 10 because there was no unique 10-item subsequence that had a support value of at least 20% in each of the two groups.

From all the subsequences, 95 and 84 number of n-items were found as frequent patterns from the first and second offerings of the MOOC respectively. Figure 4-5 reports the top-10 patterns with the most significant association (p-value <0.0005) with each of the passed and failed assessment outcome for each of the MOOCs, utilising chi-squares. The patterns are represented in the form of state transition diagrams to be able to present as much pattern as possible. The general representation of a pattern is illustrated in Figure 4-5a. S and E indicate where a pattern started and ended. Each state illustrates an activity. The notation above each self-transition illustrates the [lower limit - upper limit] on the number of times that activity can be repeated. Thus, the general pattern starts with activity A, that can be repeated for n to m times; then there is activity B, which can be repeated for k to l times. For instance if the values of [n,m,k,l] are considered to be [1,2,1,2], the pattern set would be equal to {(AB), (AAB), (ABB), (AABB)}.

Results revealed that sequences with failed outcome in both MOOCs had exactly the same top-10 patterns as presented in Figure 4-5b. This includes multiple failed attempts in an assessment without doing any activity in between as well as exploring the discussion forum before multiple failed assessment submissions. Whereas, sequences with passed outcome in both offerings had the

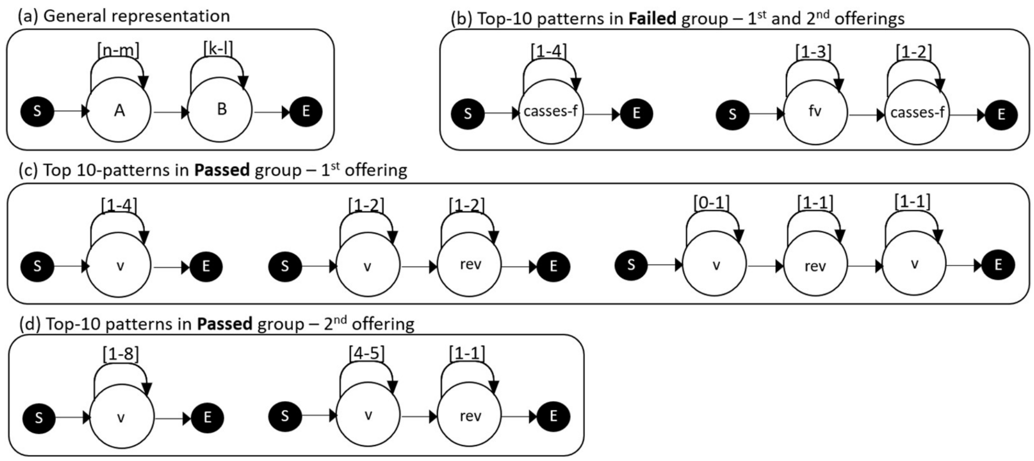


Figure 4-5. The top-10 significant patterns identified across each of the Failed and Passed groups of sequences for the 1st and 2nd offerings of Discrete Optimisation. S and E indicate where a pattern starts and ends. Each state illustrates an activity. The notation above each self-transition illustrate the [lower limit – upper limit] on the number of times that activity can be repeated. v= lecture viewed or downloaded, rev= lecture reviewed or redownloaded, casses-f= failed submission on the same assessment, fv= forum viewed

most sequences related to accessing the lecture material for one or multiple times and reviewing the already accessed content (Figure 4-5c&d).

In this group of sequences (Passed group), for both offerings, there is a pattern showing that numerous lecture materials observed before submissions (The leftmost pattern for both offerings). Furthermore, there exists a pattern in which already viewed lecture content was reviewed after viewing the lectures multiple times (the second pattern from the left for both offerings). In the first offering, there exists a pattern in which some contents were accessed after reviewing the material (The rightmost pattern in the first offering). It is worth noting that this analysis is independent of the material contents and there were multiple lectures that a student could get access to before attempting an assessment. For instance, the student could access lecture1 (part-a) and then review the same lecture, then might access lecture 1 (part-b). So it is possible to have a "view" action after a "reviewing" activity.

Comparison of the two MOOCs in terms of their influential patterns; Ranking in each MOOC was obtained by assigning a rank of 1 to the n-item with the highest association with the assessment outcome, 2 to the next highest and so on. The ranking was calculated based on all the n-items rather than the frequent ones (patterns). The reason for considering all n-items for the ranking was that there were few patterns in one of the offerings which were not frequent in the other. By ranking all the n-items, for every existing pattern, a rank was obtained, even if it was not prevalent in one of the offerings. The result of this analysis revealed a strong positive correlation (correlation coefficient of 0.83) between the rank of significant patterns of the two MOOCs. This means the two MOOCs shared similar influential patterns.

4.4. Discussion

In this study, a range of learning analytics techniques was used to examine the impact of the sequential representation of students' behaviour on their assessment outcome.

For the first research question, it was possible to predict student's outcome for assessment task submissions based on modelling their fine-grain behaviour with a MOOC in two forms of aggregated and sequential. The models were built and evaluated based on the same course but different cohorts. Based on the results, students' sequence of task preparation interactions was a more powerful predictor of submission outcome than the overall aggregated behaviour. This suggests that taking the sequential nature of students' interactions into account was more beneficial than the frequency measures. Monitoring and predicting students' performance on future tasks could provide an opportunity to support students in their learning path. Such analyses can be used to monitor students' progress and readiness over time, to delay or postpone tasks that students are not yet ready for. The result of the sequential analysis revealed the importance of taking the Sequential nature of behaviour into account to have better models and understandings of student behaviours.

The second research question suggested that considering the more historical activities prior to submission led to the higher prediction power for the models. This supports the importance of the overall process students applied in the subject as a significant indicator of learning outcome rather than only the most recent activities before submissions.

For the third research question, sequence mining was used to provide insight into significant differences in the study behaviour for each of low and high-performance groups in assessments. The results suggested viewing and reviewing the lecture content was an important factor for success, which is not surprising. Reviewing behaviour could be indicative of a range of productive learning strategies used by students include rehearsal, organisation, and self-regulation, particularly reflection (for more detail see Section 2.3). This result is also consistent with educational findings that reviewing learning material is beneficial in students' learning (Rohrer, 2015). As the study by Rohrer (2015) contends, reviewing strategies can improve memory retention and lead to an improvement in learning. Other studies have also found that reviewing content is effective in algebra word problem solving (Christianson, Mestre, & Luke, 2012), and advantageous in students' memory retention of reading content (Roediger & Karpicke, 2006).

However, an interesting insight is that a considerable number of students tried to improve their performance by attempting the assessments multiple times without doing any activity in between their attempts, that resulted in low performance in all their submissions. There was a significant negative association between these patterns and learning outcome. This pattern could reflect lack of knowledge or effort regulation strategies (see Section 2.3). Effort regulation has found to be beneficial for learners in handling the failure and setbacks in their learning process by investing and allocating appropriate effort and resource for better learning in future. This may imply that simply making multiple submissions without applying any effort and clear strategy does not lead to success in online learning. This result also confirms educational findings that too many consecutive errors undermine students' learning performance (Rodrigo & Beck, 2014). According to Ashcraft and Kirk (Ashcraft & Kirk, 2001), when an error is related to lack of knowledge and skill, it can lead to anxiety and negatively affect a student's learning.

Another group of patterns with an indirect association with learning outcomes, were students who accessed discussion forums immediately before one or multiple consecutive failed attempts. Accessing discussion forums could kind of reflect help seeking behaviour (discussed in Section 2.3) which is often emphasised as an influential factor for promoting learning (Pintrich, 2000). However as it happened prior to assessment submission, it may be more related to accessing the content of discussion forums rather than seeking help. It could also be a sign that students were not confident before they attempted assessments. In addition, in the discovered patterns, accessing discussion forums were observed before failure, and subsequently students attempted multiple failed submissions. Again these patterns could indicate making multiple submissions without applying appropriate strategy could not lead to success.

These patterns appear to be a valuable source of information to better understand students' behaviour before completing assessment tasks and could be used to inform educators about the behaviours in need of support. Moreover, adaptive learning environments could be designed to steer students away from such behaviour, for instance, by placing a time constraint between successive attempts or limiting the possible number of attempts on an assessment.

Overall, a clear implication of the research presented in this study is that the sequential dimension in modelling students' behaviour is important, and there may be a need to account for it in learning analytics research. In addition, the findings of pattern extraction analysis provide insights into how a framework for both educators and adaptive environments could be developed to identify which behaviours require support so that relevant interventions could be provided. However, this study has limitations commonly found in most MOOC studies. The first limitation relates to the generalizability of the analysis and interpretations of the patterns found. This study was based on different cohorts of one MOOC. Considering other MOOCs with different learning designs could provide better insight into students' assessment preparation behaviour and its impact on their learning outcomes. Additionally, the MOOC that was the focus of this study was an optimization course, and the assessments required applied activities that expected students to put efforts in to achieve a good outcome. Our approach might not be as effective for other assessment types, such as when students are able to play with the system or guess the answers without effort. Another limitation is not examining the learners' personal factors such as skills, prior knowledge, demographic factors, goal and motivation throughout the course, as it was beyond the scope of this study.

4.5. Chapter summary

In this chapter, a range of LA and EDM techniques was used to evaluate the importance of the sequential nature of students' behaviour on their assessment outcome. In this regard, a range of machine learning techniques was utilized to predict students' learning outcome for assessment tasks based on the aggregated and sequential representations of their behaviour. The results show the effectiveness of the prediction models, especially if they were formulated using sequential models and focusing on the overall behaviour of students before submission (rather than the most recent activities). This analysis could be useful in monitoring students' progress and postponing tasks that students are not yet ready for. The analysis also extended to explore how students approached their assessments and what strategies they applied to be prepared for this task using a sequential pattern mining approach. The discovered patterns suggest there were several behaviours which impacted on students' learning outcome, like reviewing the learned content, and multiple consecutive failed attempts without mastering the topic. These results align with previous educational research.

This chapter, together with chapter 3, focused on studying students' learning in MOOCs and suggested the importance of considering different levels of data and structure in their behaviour. LA also targets academic writing as it plays a critical role for learners in terms of their academic achievement. The temporal structure of the data has also been neglected in the literature in this context, and thus, is the topic of next chapter (Chapter 5).

Chapter 5

Investigation of Students' Temporal Writing Behaviour using Temporal Analytics

This chapter utilises a range of learning analytics techniques to evaluate the importance of the temporal nature of students' writing behaviour on their writing outcome. The analysis is carried out by characterizing students' writing processes connected to isolated writing phases and examining whether the significance of specific writing characteristics (that is associated with students' writing outcome) varies at specific moments of writing using a temporal approach. The content of this chapter is adapted from the following paper:

Donia Malekian, James Bailey, Gregor Kennedy, Paula G. de Barba, & Sadia Nawaz, Characterising Students' Writing Processes Using Temporal Keystroke Analysis, *Proceedings of The 12th International Conference on Educational Data Mining (EDM'19),* Montreal, Canada, July 2-5, 2019.

5.1. Introduction

Academic writing plays a critical role for students in terms of their academic achievement (Hyland, 2013). During recent decades the emphasis of teaching writing has shifted from focusing on the final product of writing to focusing on the process of developing a written text (Latif, 2008). The recognized significance of the underlying writing process toward completing a writing task has led to the emergence of a wide variety of research exploring the process of students' academic writing. The writing process, broadly including planning, writing and revising phases, is a non-linear process (Hayes & Gradwohl Nash, 1996). These phases often occur simultaneously, making it challenging for researchers to examine features related to specific writing phase (for more details regarding models of writing process see Section 2.4).

As discussed in Section 2.6.1.2 and 2.6.2.2, a key research area uses keystroke logging to better understand students' behaviours that comprise the writing process (e.g., Zhang et al., 2018) or predict writing quality (e.g., Guo et al., 2018). In modelling students' behaviour, it is important to consider the various levels of data available and the levels at which their activities are structured (see Section 2.5.2). Previous research mainly relied on aggregated or distribution-based representations of the overall keystroke logs in writing tasks. For instance, a recent work by Guo et al. (2018) estimated the distribution of the inter-key and intra-word pause durations for each student between their keystroke logs and found the best fit to be a heavy-tailed probability distribution. They found the estimated parameters of the distribution for each student to be a strong predictor of final score utilizing a linear regression model (for more examples see Section 2.6.1.2).

Although these studies provide informative (predictive) models characterizing students' writing process, the effect of isolated phases of writing process has not been considered and evaluated in these models. In addition, even though the use of simple models of regression and correlation analysis makes the interpretation of results easier, the use of more complex models may capture further information about the interrelationships between the extracted characteristics.

Another direction of research focused on using keystroke logs to more comprehensively characterize the phases comprising the writing process (see Section 2.6.2.2). For instance, in a study by Baaijen et al. (2012) on modelling students' writing process, some measures were defined to analyze pauses, bursts (sequences of fast text production) and revisions. They suggest that long pauses may reflect planning, as the writers are more likely to have short but well-formed bursts of writing afterwards.

Although some keystroke log features can be a marker of a successful writing process (i.e., high writing quality), there is not always sufficient evidence for the relationship. This may be due to decisions made during data processing and analysis. Even when the data has been summarized to a high standard, neglecting the time dimension may hide the effect of specific behaviours at particular moments of the writing. Temporal analysis has been suggested as a better way to uncover the writing process and its related stages (Van den Bergh & Rijlaarsdam, 1996). It may be that writing quality is related to when writers engage with certain writing phases over time (Van den Bergh & Rijlaarsdam, 1996). However, there are only a few studies that combine writing process and temporal analysis, and these are mostly focused on think-aloud procedures and offline measurements. These have been criticized as being inaccurate representations of the underlying writing process (Tillema et al., 2011).

Considering the mentioned challenges, the first aim of this study was to characterize students' writing processes in terms of a set of features demonstrating the isolated phases. This is more explicit than earlier work by Guo et al. (2018) in which the process was modelled irrespective of the isolated phases. This analysis was enabled by the unique dataset in which students were provided with dedicated sections with a software program for taking notes, writing the main body of the essay, and organizing references (each separately). Utilizing a machine learning model, features associated with isolated writing phases were modelled to examine whether they were more predictive of writing quality compared to the baseline model (Guo et al., 2018). To evaluate how influential each feature was for each student's writing quality, an additive feature attribution method was adopted that described the importance of each feature on the model's prediction for each individual student.

The second aim of the paper was to understand whether the significance of specific writing characteristics (that was associated with students' writing outcome) varied at specific moments of writing using a temporal approach. This problem was addressed in two steps: a) by breaking down students' keystrokes into several writing episodes, in each of which the prediction power of extracted characteristics was evaluated, b) by comparing the influence of each feature on writing quality across writing episodes.

This study contributes to the field by providing a new way to examine the writing process. The temporal analysis of keystroke logs related to specific phases of the writing process provides an interpretation of how each of those phases and keystroke log features affects students' writing quality over time. This can support educators to make judgments regarding students' writing processes and change the ways they teach writing skills. Overall, in this study the following research questions are explored:

Research question 1. Do models that characterize the overall writing process miss informative features associating with particular phases of writing? Can new characteristics (features) of the writing process be defined from students' keystrokes to improve these models?

Research question 2. How predictive of writing quality are the extracted features at different times? Is there any evidence that the importance of specific features varies with time?

The results highlight that characterizing students' keystrokes separately while writing, taking notes and organizing references is more predictive of their writing quality. Additionally, based on the findings, the importance of several features on writing quality changes or even disappears if not taking the time dimension into account. Investigation of the influential features for individual students clarified the non-linear relationship between some features and writing quality (i.e. the influence of each feature on the writing quality is likely determined in combination with other features). This is consistent with a theory of writing which suggests there exist considerable overlap and interaction between writing phases (Biggs, 1988). This also reveals the necessity of using models that capture the interrelationship between features rather than simple correlation or linear regression analysis.

5.2. Approach

5.2.1. Study context

The study involves 107 students from the University of Melbourne who enrolled in a business undergraduate course. Students were asked to use a specific online word processing software called Cadmus to write a 1000-word essay that worth 10% of their final mark in a course. Students had to choose between two topics (1) "The business of business is to make profits", and (2) "What is the business case for Corporate Social Responsibility?". Participants had 19 days to complete the essay. The aim of the task was to evaluate a student's understanding of the material and the performance was marked by teaching staff using a score between 0 to 100.

Cadmus has similar properties to other word processing software tools such as a section to write the main body of the essay (body section), editing, highlighting, and additional features such as dedicated sections to take notes (note-taking section), and to organize the reference materials (reference section) as well as a single paste restriction of 90 words. Cadmus records every keystroke in each section via the keyboard while students work on their assignment. Ethics for the use of learning analytics for research purpose was covered by Cadmus's Terms and Conditions.

5.2.2. Data processing

In this section, the procedure undertaken to characterize students' writing processes by engineering a set of features from keystroke logs are described.

First, the procedure introduced by Guo et al. (2018) to represent patterns of pauses in students' keystrokes is defined. This study applied the same procedure on the pauses of each writing section separately. To be able to capture more aspects of the process, more features than previously identified in the literature were considered in this study including bursts and revision summaries. Moreover, a new set of features was introduced regarding how students divided their overall time on the task to specific aspects of writing including revision, note-taking, transcribing, and reference organization. In addition, two concepts of writing quality and writing episode were processed to assist on answering the research questions.

(A preliminary analysis revealed 4 students who did not complete the writing task and had less than 600 words in their essay. They were removed from further analysis.)

5.2.2.1. Pauses

In this study, the focus is on inter-key pauses (the duration between successive keystrokes) that are more likely to be related to the processes such as text planning and reviewing the written text (Guo et al., 2018). As a study by Guo et al. (2018) suggested, there is a tendency for inter-key pause durations to follow a stable distribution. They fitted this distribution to pause duration data to obtain an informative estimation of the related parameters for each student.

In this study a similar procedure as Guo et al. (2018), was followed; however, pauses in each writing section of the dataset (body, note-taking, reference) were summarised separately. This section-specific summarisation could reveal more explicit information regarding which processes were more likely to be engaged during the pause. For instance, pauses in the reference section could be more related to preparing the required material or organizing references. Patterns of pauses in the note-taking section could be more connected with the planning behaviour (Hayes, 2012).

The aim is to represent students' writing process while they were working on their essay, thus the pauses between writing sessions and pauses more than 2 hours inside a session that meant the student had left the session were decided to be removed from the analysis.

In the mentioned dataset, the exploration of the distribution of pause durations for each student in each writing section revealed that there was a tendency for pause duration in each section to follow a heavy tail distribution. Figure 5-1 illustrates the pause duration for a sample student in each section of the essay separately. The figure shows that in each section, the majority of the pauses were short but there exist a few long pauses. Believing a stable distribution to be a plausible hypothesis, this distribution was fitted to each student's pause data to obtain an estimate of the related parameters in each section.

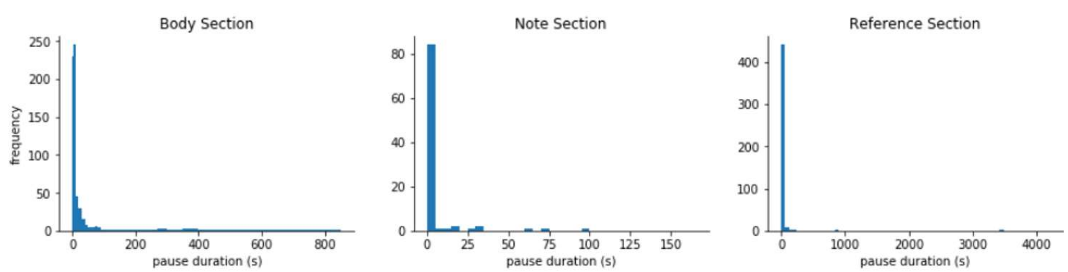


Figure 5-1. Histogram of inter key pause durations in each section of body, note-taking and reference for a sample student

The definition of this distribution is explained briefly in this section. In probability theory, a quantity x obeys a stable distribution if for any n≥2, there exist constants $a_n > 0$ and real number b_n such that

$$\sum_{i=1}^{n} X_i \stackrel{d}{=} \alpha_n X + b_n$$

As there exists no closed form for this distribution, its characteristic is given by the following function:

$$\Phi_x(u) = \begin{cases} \exp\left\{-|\gamma u|^{\alpha}\left[1 + i\beta \operatorname{sgn}(u)\tan\left(\frac{\pi\alpha}{2}\right)(|\gamma u|^{1-\alpha} - 1)\right] + i\delta u\right\}, & \alpha \in (0,1) \cup (1,2) \\ \exp\left\{-|\gamma u|\left[1 + i\beta \operatorname{sgn}(u)\frac{2}{\pi}\ln|u|\right] + i\delta u\right\}, & \alpha = 1 \\ \exp\left\{i\gamma u - \frac{1}{2}\sigma^2 u^2\right\}, & \alpha = 2. \end{cases}$$

that needs four parameters (α, β, γ, δ) for the complete description.

- The parameter alpha $\alpha \in (0, 2]$, which is called the index of stability or tail index. This parameter gives information about the height of the tails.
- The parameter beta $\beta \in [-1, 1]$, called the skewness parameter. The distribution is skewed to the right when $\beta > 0$. It is symmetric when $\beta = 0$, and it is skewed to the left in case of $\beta < 0$.
- The parameter delta $\delta \in R$, called the location parameter is equal to the median. Depending on how heavy the tail is, some extreme part of the data (i.e. the first and last quartiles) may need to be discarded to have a good estimate of this value. For normal distribution, the mean can be used as an estimate for δ.
- The parameter gamma $\gamma > 0$, called scale parameter is a measure of dispersion. For symmetric distribution it is equal to half of the variance.

Parameters alpha and beta specify the distribution's shape, while parameters gamma and delta define the scale and location of it. For parameter alpha, the smaller value (a heavier tail of the distribution) indicates a student having larger pauses. Positive beta means a student has relatively extreme large pauses (skewed behaviour in pauses). Smaller delta indicates the peaks are relatively smaller pauses. Smaller gamma indicates relatively smaller variation in pause durations.

For each student, in each section of the dataset these four estimated parameters were obtained. This can indicate specific phases of the writing process more explicitly compared to the pattern of pauses extracted from the overall summary of keystrokes. These parameters for the overall keystrokes of each student (irrespective of the specific section) as suggested by Guo et al. (2018) were also estimated and considered as a baseline for the purpose of evaluation.

5.2.2.2. Bursts

For modelling the transcribing phase of the process, previous research has tried to identify bursts in writing. A burst is defined as a sequence of fast text production that depends on a defined threshold for the length of a pause, determining the burst boundary. Burst summaries (i.e. mean length of the bursts) can reveal students' fluency in the transcribing phase of the writing process (Zhang et al., 2018).

In keeping with the majority of the literature (Rosenqvist, 2015), this study identified bursts by breaking up keystrokes at every pause that have longer than 2 seconds of keystroke inactivity. This procedure was applied in the body section of the dataset, where students wrote the main part of their essay. Considering the bursts in which at least one word is typed, burst length was summarised by two features based on the number of words in a burst: The mean and the standard deviation of burst length for each participant. Additionally, burst duration was summarised by two features of the *mean* and *standard deviation of burst duration.*

5.2.2.3. Revision

In this step, the aim was to isolate the revision phase of the writing process from the transcribing phase. Authors of (Chanquoy, 2009) associated single backspaces to spelling correction which reflect self-monitoring, and multiple backspaces to editing in which longer revision in the written text occurred. Similarly, this study summarized revision at small and large scale with a slight change in the definition; revisions were identified based on the number of word deletions in a writing burst rather than in isolation. A burst was labelled with single deletion as small, whereas bursts with multiple deletions as large-scale revision.

Two features were extracted to provide measurements related to the revision phase of the writing process: The *frequency of both small scale* and *large-scale revision bursts*.

5.2.2.4. Time percentage on each writing aspect

To summarize the extent of each specific aspect of writing (i.e., revision, planning, transcribing) that was used by each student, a new set of features were introduced, including the percentage of the total writing time dedicated to: note taking (total time in note-taking section), small- and large-scale revisions (bursts of writing with small or large deletion), transcribing (total time of bursts in the body section), and reference organization (total time in reference section). Overall time spent on the essay is also another important feature that may indicate a student's effort and motivation (Guo et al., 2018).

5.2.2.5. Writing quality

Writing quality corresponds to the students' final grades on the essay (Guo et al., 2018; Zhang et al., 2018). Although final grade can be used for evaluating the effectiveness of the extracted features, previous research has found that students' final grade may not be a reliable measure of success, due to variations in grading essay writing by raters (Kayapinar, 2014). To decrease the effect of raters' variability in marking, this study mapped the students' writing quality to two categories - high and low level instead of the exact grade. In the dataset, the distribution of students' grades lied within the range 60 to 95 as presented in Figure 5-2. The median value was adopted as a threshold for this mapping. In this dataset, the median of score is 80. Based on this value 40 students had high quality writing (with grade more than 80), and 67 students had low quality writing.

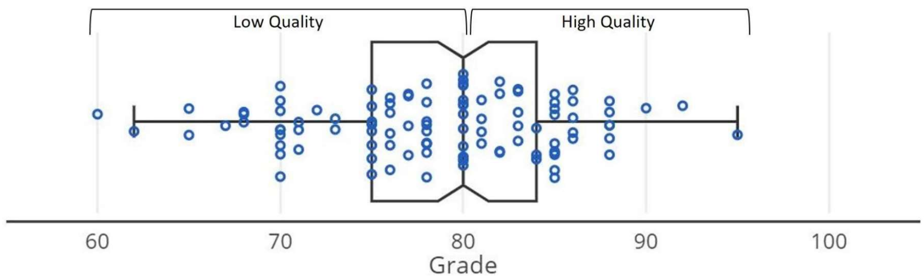


Figure 5-2. Distribution of students' grade

5.2.2.6. Writing episode definition

One approach to temporal data processing is the analysis of data in multiple consecutive episodes to ensure all participants have a similar number of observations (Van den Bergh & Rijlaarsdam, 1996). There exist two criteria to define the episodes (Van den Bergh & Rijlaarsdam, 1996): the first is based on equal time units, the second is based on the equal number of observations. From a statistical point of view, the second is preferred (Van den Bergh & Rijlaarsdam, 1996).

In this study, students were asked to write a 1000-word essay that provides a good criterion for defining the fixed observations. The episodes were defined based on the keystrokes used to complete the fixed number of words in the essay. This way, meaningful episodes were obtained recording a specific draft of writing irrespective of the differences in the time students spent on the writing task.

For this purpose, students' writing data were split into N writing episodes, N = {N1 + N2 + … + Nn}, each of which records students' keystrokes used to complete n*k words of the essay. Five writing episodes were defined, each of which involved all the keystrokes from the start of writing to the completion of n*200 words. There is no theory for defining the number of episodes and the results may differ based on the choice of this number. Figure 5-3 illustrates the writing episodes of this study.

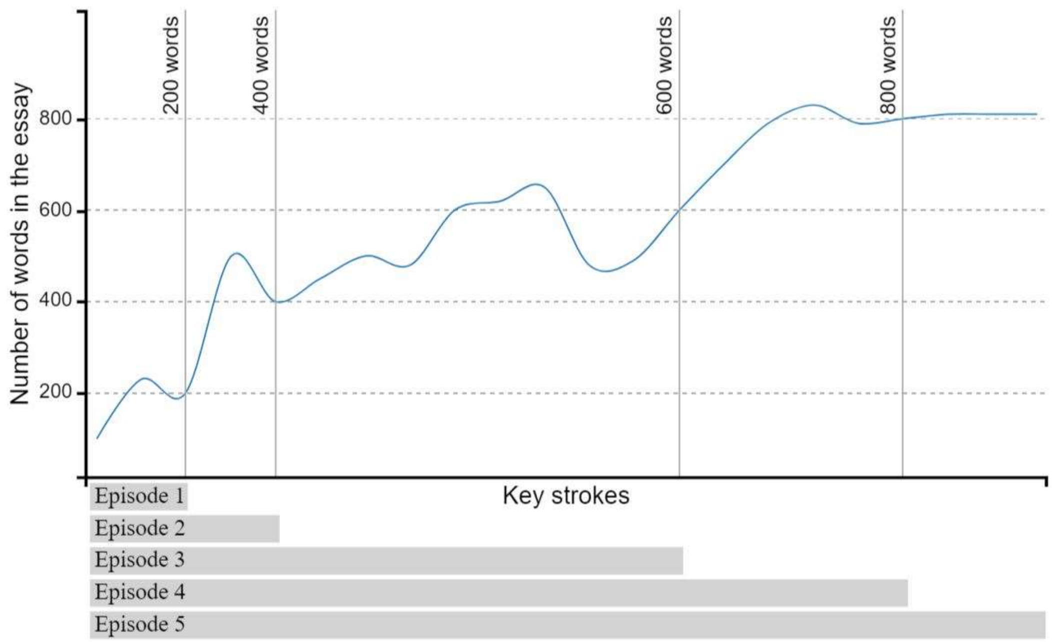


Figure 5-3. Representation of writing episodes in this study

As can be seen, the first episode involved all the keystrokes for the development of 0 to 200 words, the second episode involved all the keystrokes used to complete 0 to 400 words and the last episode (episode 5) recorded the total number of keystrokes.

5.2.3. Data analysis

This study conducted two sets of analyses: one for each research question. Each research question applied features defined in the previous subsection.

5.2.3.1. Research question 1. Do models that characterize the overall writing process miss informative features associated with certain phases of writing? Can new characteristics (features) of the writing process be defined from students' keystrokes to improve these models?

To address this question, the performance of a machine learning model trained on the pause related features extracted from overall keystrokes (by the procedure suggested by Guo et al. (2018) explained in Section 5.2.2.1) were compared against the performance of a model trained on pause features of each writing section separately. These feature sets are called as a baseline and section-specific respectively. Work by Guo et al. (2018) also identified the total time on task, along with the pause related features as a strong predictor of writing quality. Thus, this feature was included in the baseline and section-specific feature sets of this study. These sets are presented in Table 5-1. The evaluation is based on the prediction performance of writing quality. Existing research has mainly approached the evaluation task by correlation or linear regression analysis (Guo et al., 2018; Yang et al., 2018). Although these simple models are easy to interpret, more complex machine learning models may capture extra information related to the interrelationship between features. To test this, this study used a complex model of machine learning as the prediction model.

In the next step, the model was extended by adding further aspects of the writing process, in terms of burst and revision summaries, reflecting transcribing and revising phases of the writing process (explained in Section 5.2.2.2 and Section 5.2.2.3). In addition, the features representing the extent to which specific aspects of the writing process were used during writing (Section 5.2.2.4) were also included. These are represented as the combination set in Table 5-1. Then the performance of the model was evaluated to examine whether the introduced features were informative (predictive) of writing quality.

Table 5-1. Brief overview of feature sets

Feature Set #	Features
Baseline	Parameters of stable distribution estimated based on the overall keystrokes, total time on task.
Section-specific	Parameter of stable distribution for each section of body, note-taking and reference, total time on task
Combination	Section-specific set, burst and revision summaries, time percentage on each writing phase.

In order to evaluate how meaningful the extracted features are in terms of being associated with writing quality, the contribution of each feature to the prediction of model for each individual student was computed, using a consistent and accurate additive feature attribution method (Lundberg & Lee, 2017). This facilitates a better interpretation of the machine learning model.

Algorithm. A binary prediction model was utilised based on each feature set where the class label was students' writing quality. An XGBoost classifier (Chen & Guestrin, 2016) was utilised for this purpose. This model is known as a robust model having good prediction power. Even though the XGboost classifier generally provides a good prediction power based on the features, understanding why it made a prediction and what was the contribution of each feature seems to be hard due to the complexity of the model (For more details about XGBoost see Section 2.5.3.1.2). That is why many studies prefer to use linear models which are easier to interpret but might have lower prediction power because of not taking the interrelationship between the features into account.

To evaluate and derive the influence of each feature on writing quality, this study used a recently proposed approach of the SHAP (SHapley Additive exPlanations) algorithm which can be used to explain the output of any machine learning model (Lundberg & Lee, 2017). In this algorithm, the contribution of a feature is calculated by comparing the model's predictions with and without that feature. Every individual is assigned a SHAP value for each feature that determines the feature's contribution to a change in the model's prediction (for more details see Section 2.5.3.1.3).

5.2.3.2. How predictive of writing quality are the extracted features at different times? Is there any evidence that the importance of specific features varies with time?

The essay is developed gradually and the influence of some aspects of the writing process (characterized by extracted features in this study) may vary over time. To examine to what extent each extracted feature was informative of writing quality over time, and how their contribution may vary, the following steps were proceeded:

- Students' keystroke logs were broken down into n writing episodes (Section 5.2.2.6),
- For each episode, the predictive features (combination feature set) were extracted. For the pause parameters, the underlying distribution for each episode was evaluated and stable distribution was found to be a relatively good fit.
- For each episode, the predictive power of the extracted features on writing quality was evaluated using the XGBoost model.
- To identify the contributing features to the model's prediction in each episode, the influence of each feature was derived on the prediction using the SHAP algorithm.

5.3. Results

This section presents summary statistics of the extracted features, then the research questions are addressed.

5.3.1. Statistical summaries of the features

Table 5-2a gives a summary statistic of the pauses in each writing section estimated based on fitting a parametric stable distribution. These parameters are also estimated for the overall pauses in the combination of sections. The statistics are presented in terms of mean and standard deviation. As can be observed, for each section (Body, Reference, Note Taking) as well as the combination of them (All sections), on average alphas(α) ranged from 0.42 to 0.62, suggesting that pauses were heavy-tailed. Estimated alphas were on average smaller in the reference and note-taking sections compared to the body which implies students had relatively larger pauses while taking notes or organizing references. Betas ranged from 0.69 to 1, indicating data were highly skewed to the right.

On average gamma(γ) ranged from 17.58 to 67.2, indicating high variation in pause duration. The higher value in the body section indicates on average a higher variation of pause length compared to the other sections.

delta(δ) ranged from 70.1 to 400.6, indicating the peaks. Peaks were larger pauses in the reference and note-taking sections compared to the body, which might imply students' writing tempo was slower in these two sections.

One interesting observation is that on average the estimated parameters in the body section were close to the combination of all sections. This implies that most dominant pauses were related to the body section of writing and the pattern of pauses in the note-taking and reference sections were not reflected in the overall keystrokes.

Other statistics regarding bursts and revision summaries, as well as the dedication of time to each writing phase are presented in Table 5-2b. As can be seen, on average the students spent 14.5 hours to complete the essay. Of that, on average 50 percent was dedicated to transcribing (bursts of writing without deletion) and only 5 to 18 percent to each of the other activities of revision, note taking and reference organization. For simplicity of reporting results, the feature names are mapped to IDs as provided in Table 5-2b.

Table 5-2. Summaries of the extracted features: a) estimated parameters of a stable distribution fitted for each student's pause data in each of body, note taking and reference section, as well as the combination of all sections (Overall). b) Revision and burst summaries as well as the percentage of time spent on each phase of writing

a)	Body section - Pause				Note Taking section - Pause				Reference section - Pause				All sections - Pause			
Feature	α	β	γ	δ	α	β	γ	δ	α	β	γ	δ	α	β	γ	δ
Feature ID	B-Pα	B-Pβ	B-Pγ	B-Pδ	N-Pα	N-Pβ	N-Pγ	N-Pδ	R-Pα	R-Pβ	R-Pγ	R-Pδ	O-Pα	O-Pβ	O-Pγ	O-Pδ
Mean (std)	0.62 (0.14)	1.0 (0.01)	67.23 (54.5)	70.12 (61.3)	0.42 (0.31)	0.69 (0.46)	17.58 (27.3)	109.51 (792.2)	0.52 (0.12)	0.98 (0.14)	32.47 (64.6)	400.6 (1566.4)	0.63 (0.13)	1.0 (0.01)	61.11 (37.7)	60.35 (45.03)

b)	Body section - Burst summaries				Percentage of time on					Body section – Revision summaries		Total time on task (hr)
							Body section					
Feature	Mean burst-length	Std burst-length	Mean burst-time (sec)	Std burst-time (sec)	Note taking section	Reference section	Transcribing	Small-revision	Large-revision	Freq small-revision	Freq large-revision	
Feature ID	B-MBrsL	B-SBrsL	B-MBrsT	B-SBrsT	N-T	R-T	B-T	B-TSRev	B-TLRev	B-FSRev	B-FLRev	TT
Mean (std)	4.1 (2.1)	5.66 (4.98)	68.7 (421.1)	743.5 (4500.7)	0.13 (0.22)	0.05 (0.08)	0.51 (0.25)	0.11 (0.14)	0.18 (0.17)	125.7 (74.9)	160.03 (108.93)	14.5 (29.3)

5.3.2. Results for Research question 1

The results answering the first research question are reported in two sections. First, the prediction power of the introduced feature sets is compared and reported on writing quality. Then the contribution of each feature on prediction is derived and discussed.

5.3.2.1. Examining the prediction power of feature sets on writing quality

For each set of features (Table 5-1) a model was trained using leave one out cross-validation (with 5-fold nested cross-validation for parameter optimization). The performance of each model on the prediction of writing quality is reported in Table 5-3 based on the metrics of accuracy and the area under the ROC curve (AUC) (Ling et al., 2003).

The accuracy and AUC were slightly higher for the section-specific feature set compared to the baseline set. This could indicate that characterizing the overall process irrespective of the isolated phases may miss specific moments where the features associated with certain writing phase become more important. Adding burst and revision summaries, as well as the time dedication features (combination feature set), to the model, improved the performance of the prediction significantly. The "combination feature set" obtained the highest prediction power implying that the introduced features were a better representation of the students' writing quality.

Table 5-3. Prediction power of each feature set on writing quality based on accuracy and AUC

Features set	Accuracy	AUC
Baseline	70.09	70.63
Section-specific	71.03	71.75
Combination	**81.31**	**82.72**

5.3.2.2. Examining the contribution of features on prediction

Focusing on the most predictive feature set, the next step was to evaluate the influence of each feature on predictions. For this purpose, an XGBoost model was built on the combination feature set (using 5-fold cross-validation for parameter optimization). The model was then passed to the SHAP algorithm to explain the influence of each of the features on the model's prediction for each student.

Since individualized explanations of every feature for every student (based on the SHAP values) were obtained, the distribution of the importance of each feature on the model's prediction could be plotted, as is presented in Figure 5-4. The features are sorted by the mean of absolute SHAP values over all students to gain a global insight into the most influential features across all students. As can be observed, the most influential features (globally) were total time on task (TT), frequency of small revisions (B-FSRev), and estimated gamma parameter from pauses in body section (B-Pγ) respectively, while some of the pause parameters estimated in the note-taking and reference sections (i.e. N-Pα, R-Pβ) had the lowest contribution. For the convenience of viewing, only the top-20 features that were globally influential are presented in the figure. In this figure, each row corresponds to a feature and every student has a dot on each row of the plot. The x position for each dot determines the contribution of that feature on the prediction for the student (SHAP value), and the colour of the dot indicates the value of that feature for the student (red high, blue low). This reveals, for example, that a high percentage of time on the reference section (R-T) increases the chance of having high quality writing for a subset of students (red dots in the R-T row, and on the right side of the plot).

Figure 5-5 illustrates SHAP decision plot determining how the model made prediction (arrived at the decision) for each instance and how each feature influenced the overall prediction. Moving from the bottom of the plot to the top, SHAP values of each feature are added to the base value of the model.

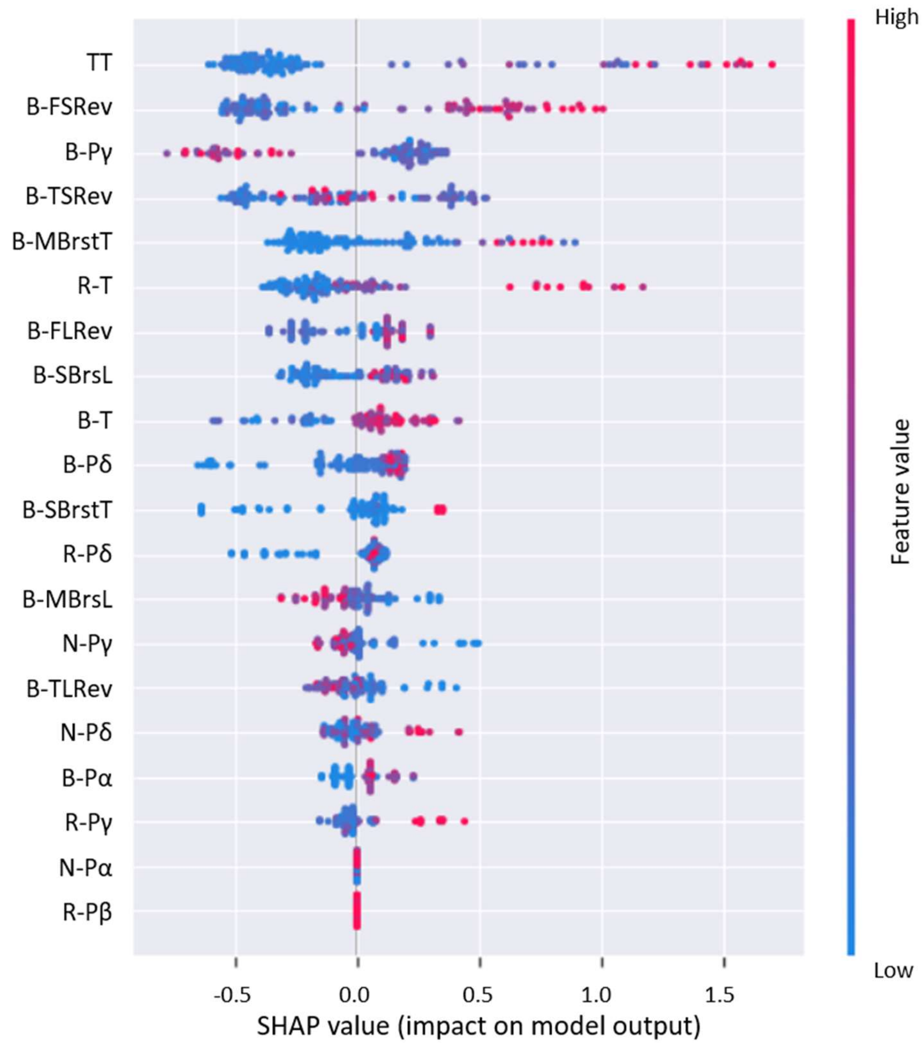


Figure 5-4. SHAP values for each feature indicating the influence of that feature on writing quality of each student. Every student is associated a dot on each row. The x position for each dot determines the impact of that feature on the prediction for the student, and the color represents the value of that feature for the student.

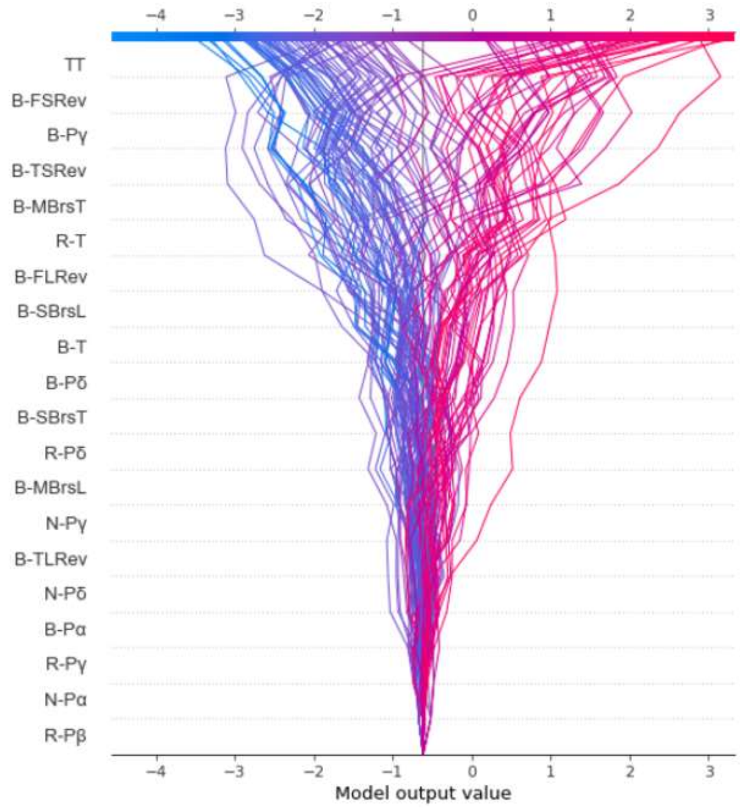


Figure 5-5. SHAP decision plot determining how complex the model made predictions for each instance and how each feature influenced the overall prediction. Moving from the bottom of plot to the top, SHAP values of each feature are added to the base value of the model.

Below is the interpretation of some of the features detected as important by the prediction model. It is worth mentioning that this interpretation is intended for high-level model interpretation, and the model's decision making was more complex and took the interaction between features into account. It is worth mentioning that, the explanations provided by SHAP are SHAP's own views of the world and not the actual causal representations. In other words, the accuracy of SHAP's explanative models might also need to be assessed and validated in the future.

Total time on task (TT) was found to be the strongest predictor of writing quality. A subset of students (red dots on the related row of figure) who spent a lot more time than others on the essay were more likely to produce high-quality writing. This supports previous research (Guo et al., 2018; Zhang, Hao, Li, & Deane, 2016) and could be an indication of student's motivation in completing the writing task (Hayes & Gradwohl, 1996). However, there are also students with lower time spent on the task, but a higher chance of producing high-quality writing (blue dots on the right side of the plot in the related row). This could indicate this feature might be influential on the prediction mainly in combination with other features.

The colouring of the second most important predictor - the percentage of time on small revision (B-FSRev) shows how a higher frequency of small revision means a higher the chance of producing high-quality writing for most students. This pattern also holds true for the frequency of large revision (B-FLRev). It could imply that writing quality improves if revisions are performed more frequently. This is in agreement with previous research (Stallard, 2019) suggesting that good writers stop their writing more often to perform revision and to correct their spelling and grammatical errors as they write compared to weak writers.

Percentage of time on reference section (R-T) is not the most important feature globally, however it is one of the most important features for a subset of students. This could indicate students' effort in gathering and organizing the required materials for the writing have a positive effect on writing quality. The estimated gamma parameter from pauses in this section (R-Pγ) also shows a positive influence on writing quality for most of the students which indicate students with higher variation in pause durations were more successful. Again, it can be observed that this argument does not hold true for several students (blue dots on the right side of the plot), indicating that the influence of each aspect is likely determined in combination with other aspects.

Percentage of time on transcribing in the body section (B-T) shows most students who allocated more portion of the time on transcribing were more likely to have a high-quality writing. Although there might be numerous reason for this observation, one alternative justification could be found in a study by Green and Wason (1982). They found that writers who spend most time on other aspects of writing such as planning tend to dislike writing, and this may lessen the quality of the text.

The estimated gamma and alpha parameters from pauses inside the body section (B-Pγ, B-Pα) shows a (negative, positive) association with the chance of producing a high-quality writing for most of the students. Together these parameters mean lower variation of pause durations without having very large pauses. This could be an indication of steadiness and fluency in writing and its direct relation with writing quality which is in agreement with previous research (Guo et al., 2018).

Estimated delta and gamma parameters from pauses in the note-taking section (N-Pδ, N-Pγ) shows a (positive, negative) relation with the chance of producing a high-quality writing for most students. Together these parameters mean more large pauses and lower variation in pause duration. This means that steadily taking large pauses while taking notes leads to a higher chance of high-quality writing. Based on a work by Baaijen et al. (2012), large pauses may reflect the planning process. Since this pattern is observed in the note-taking section of this dataset, these steady long pauses could be more certainty be connected to thinking periods on planning and its direct association with writing quality could be inferred.

An interesting observation is the negative association of the mean burst length (B-MBrsL) with the text quality for the majority of students. This observation is not consistent with results from other studies that related mean burst length to writing fluency and found a positive association with writing quality (Alves et al., 2007; Rosenqvist, 2015). One alternative justification could be found in a work by Baaijen et al. (2012). They suggest that larger pauses (reflecting planning) before bursts lead to shorter but well-structured bursts. This may indirectly confirm that how having on average shorter bursts (but probably better-formed) increases the chance of producing a high-quality writing as observed in this study. Another justification could be that in this program of research, students were required to write a long essay (1000-word, over 19 days) compared to other studies (e.g., no word limitation, over one hour on average; Alves et al., 2007). Thus, the probability of having a larger variation in their burst length was higher. Therefore, the burst length might not be an indication of writing fluency. In fact, it might be more related to procrastination behaviour. As can be seen in the figure, most of the students who had a larger variation in their burst length (B-SBrsL) were likely to have a better quality of writing, whereas those with less variation could be those who were trying to complete the task in lesser time and type on average in longer bursts. However, verification of such claims needs more evidence and analysis.

Even though there were several features such as percentage of time on small revision (B-TSRev), estimated gamma from reference and body section (R-Pδ, B-Pδ), that were detected as influential, their clear linear association with writing quality cannot be observed. For a subset of students, the higher value was associated with higher quality and for others it was the opposite. Again, this indicates that the importance of these features was impacted by other features. Simple correlation or linear regression analysis may have difficult to uncover the influence of these features.

SHAP algorithm is a powerful technique that enables uncovering the insight and patterns from a machine learning algorithm. The main focus on this program of research is to examine the effect of the *temporal* writing behaviour on writing quality. Therefore, further explanation of the SHAP output in this section (which deals with the time-independent behaviour) is out of the scope of this research. However, the rest of this section provides a very brief overview of how the SHAP algorithm could be used further to provide more insight on the non-linearity of features and toward formulating causal hypotheses on the findings.

SHAP values can be visualized as "forces" that means whether a feature's value increases or decreases the prediction. This visualization is shown in Figure 5-6 for a sample student. Features pushing the prediction higher (higher writing quality) are shown in red, whereas those pushing the prediction lower (lower writing quality) are in blue. Therefore, for each student, it can be determined what features were the most influential ones on the writing quality and in which direction. For instance, for the sample student in Figure 5-6, more time on task (TT) and Time on small revisions (B-TSRev) were pushing the prediction higher, whereas larger variation in the burst length (B-SBrsL) and in the pause durations (B-Pγ) – unsteadiness and non-fluency in writing - were pushing the prediction lower.

Figure 5-6. A representation of force plot for a sample student. Features pushing the prediction higher (higher quality writing) are shown in red, whereas those pushing the prediction lower (lower quality writing) are in blue.

The force plots of all instances can be visualised all at once illustrating how the final prediction was achieved according to a sum of each predictor's attribution. This is shown in Figure 5-7 which has stacked the SHAP values for all students vertically and placed them side by side based on their cluster similarity (SHAP values have the same unit and can be easily clustered - students are being clustered based on their explanation similarity).

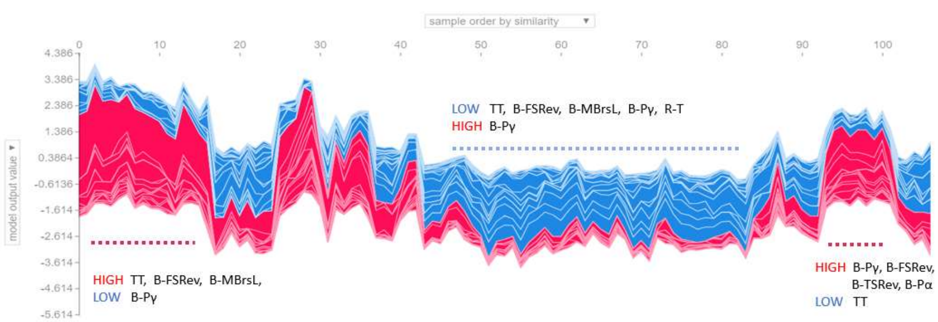


Figure 5-7. A representation of force plot for all students, clustered based on explanation similarity using hierarchical agglomerative clustering. Features pushing the prediction higher (higher quality writing) are shown in red, whereas those pushing the prediction lower (lower quality writing) are in blue.

Three sample clusters are illustrated in Figure 5-7. The leftmost and the rightmost clusters are the groups with a high predicted writing quality, whereas the middle cluster is the group with a low predicted writing quality. Some of the most important features pushing each cluster to the high or low prediction are also illustrated in the figure. For instance, for the left cluster, spending more time on task, more frequent small revision and being more fluent in writing are pushing the prediction higher. This analysis could help to gain more insight regarding various groups of students and how features in each group impact on the prediction.

The interpretation of the importance of a single feature on the prediction could also be provided by SHAP dependence plot as provided for a sample feature in Figure 5-8. In this plot, SHAP value and the feature value are plotted in the y-axis and x-axis respectively. This plot illustrates how the feature importance (SHAP value) changes as its value varies.

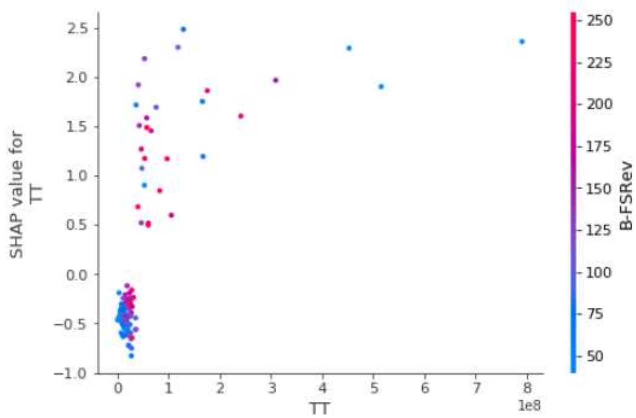


Figure 5-8. SHAP dependence plot that indicates the interpretation of the importance of a single feature of TT (total time on task) on the prediction. Each dot is a student. Dependence plot automatically selects another feature for coloring which indicates the interaction effect of features.

The influence of total time on task (TT) on the prediction is shown in Figure 5-8. Higher SHAP value indicates a higher probability of having high-quality writing due to spending more time on task. Dependence plot automatically selects another feature for coloring that determines the interaction effect of features. Vertical dispersion at a single value of TT, indicates interaction with other features in the data. Coloring each student by the frequency of small revision (B-FSRev) indicates that spending higher time on task (high TT) is influencing the model prediction more significantly, where students also do more frequent small revision (high B-FSRev). In other words, spending more time on task is more effective on having high-quality writing where students also do more small revisions.

As mentioned earlier, the main focus of this study is to gain insight regarding the importance of considering temporal dimension on analyzing students' writing behaviour. Therefore, further explanation of the SHAP output in this section (which deals with the time-independent behaviour) is out of the scope of this study.

5.3.3. Results for Research question 2

In this section the second research question is answered in two steps.

5.3.3.1. Examining the predictive power of features on writing quality over time

Using the XGBoost classification algorithm the predictive power of the extracted feature in this study (combination feature set) was examined on writing quality at each writing episode. The results are reported using the metrics of accuracy and area under the ROC curve (AUC), based on leave one out cross-validation (with 5-fold hyperparameter optimisation). The outcome is shown in Table 5-4. The results demonstrate a high and, in some episodes, moderate success rate in predicting students' writing quality (better than random chance). The highest prediction power is associated with the last episode which involves the complete keystrokes.

Table 5-4. Prediction power of extracted features in each writing episode based on the accuracy and AUC

Episode#	Accuracy	AUC
1	69.16	70.26
2	67.29	67.12
3	69.16	66.01
4	79.43	79.22
5	81.31	82.72

5.3.3.2. Examining the contribution of features on prediction over writing episodes

The next step was to examine the contribution of each feature on the prediction of writing quality at each writing episode and examine whether the contribution varied. For this purpose, an XGBoost model was trained based on the combination feature set in each episode (using 5-fold cross-validation for parameter optimization). Each model was then passed to the SHAP algorithm to explain the importance of each feature on the prediction. The result is reported in Figure 5-9, in which the global contribution of each feature in each writing episode is visualized based on the mean of absolute SHAP values for that feature over all students. A darker colour indicates a higher contribution of that feature on the prediction model in that episode. As can be seen the importance of total time on task (TT) is relatively stable over all the episodes, whereas the importance of the note-taking related features such as percentage of time on note-taking section (N-T), and estimated pause parameters (N-Pδ, N-Pβ) were found to be more influential in the 2nd writing episode. The patterns of pauses in the reference section (i.e., the percentage of time on the reference section (R-T), as well as the estimated gamma and delta parameters (R-Pγ, R-Pδ) also show stronger influence on the prediction of writing quality at the beginning of writing.

The 2nd episode shows interesting sets of contributing features to predicting writing quality. Moreover, in this episode the contributing features are different from the 5th episode (overall keystrokes) analysed in the previous section. Thus, the importance of several contributing features in this episode compared to the 5th episode was decided to be visualized and discussed. The related feature importance is presented in Figure 5-10 using the SHAP algorithm.

Unlike the 5th episode, in this episode the estimated pause parameter in the body section (B-Pδ) has been detected as the most influential feature. However, no linear association with writing quality could be inferred.

The estimated beta parameter (skewness of distribution) from pauses inside the note-taking section (N-Pβ), was not found as influential in the 5th episode. This feature shows a positive relationship with the chance of having high-quality writing for almost all students. This could indicate students who had relatively extreme large pauses with respect to their normal behaviour while taking notes are more likely to be successful. As mentioned before, a work by Baaijen et al. (2012) associated large pauses to planning. Since this pattern is observed in the note-taking section, so there exists more evidence to connect these extreme pauses to large thinking periods on planning, and its direct association with writing quality if it happens at the early stage of the writing.

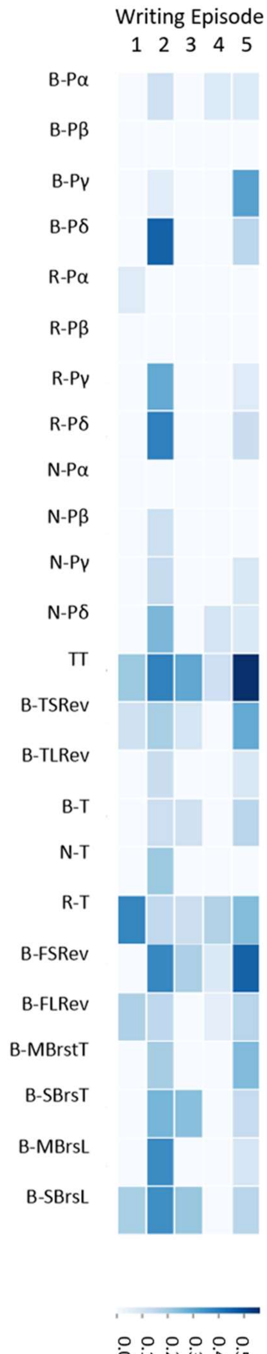


Figure 5-9. Visualization of the global importance of each feature on the prediction in each writing episode based on the SHAP values. Darker colors indicate higher contribution of that feature on prediction in that episode

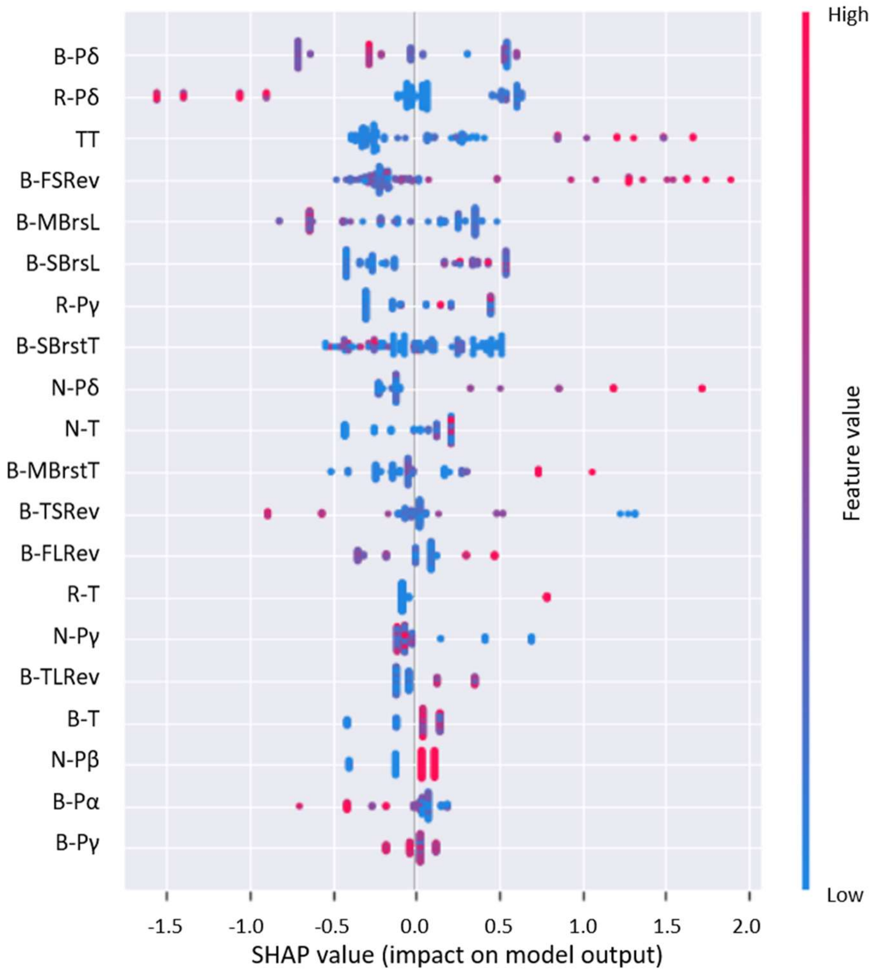


Figure 5-10. SHAP values for the 2nd writing episode indicating the influence of each feature on writing quality for each student.

The second most influential factor is R-Pδ indicating the peaks in pauses of the reference section. Smaller delta (i.e. a faster writing tempo) was found to increase the chance of producing high-quality writing for a subset of students. In the 5th episode, this pattern was found not to be as influential as this episode. Moreover, unlike the 5th episode, the percentage of time on reference organization (R-T) shows a clear direct association with a higher chance of producing high-quality writing. This could indicate that it is more likely for high-quality writing to be produced if more effort is allocated to reference gathering or organization in the early stages of the writing.

An interesting observation is the negative relation of the estimated alpha parameter in the body section (B-Pα) with the chance of having high-quality writing for most students. The relation of this feature was found to be the opposite for the 5th episode. When this feature is considered at the early stage of the writing it suggests that successful students had relatively larger pauses while transcribing. This could indicate that having a higher tempo in transcribing does not always imply higher quality writing. It suggests that this, in fact, depends on what moment during writing this is examined. As mentioned before, larger pauses could reflect planning behaviour (Baaijen et al., 2012). Thus, this pattern could emphasise the importance of planning at the early stage of writing for having high-quality writing.

Similar to the 5th episode, total time on task (TT) and frequency of small revision (B-FSRev) are among the most influential factors with the direct association with writing quality.

Overall, the influential features were quite different and for many features no clear linear trend regarding their association with writing quality could be observed. Again, this indicates that the influence of features on writing quality was not linear, rather it was mainly based on combination with other features.

The temporal analysis reveals that the importance of specific writing behaviour on writing quality may change over time. Thus, characterizing students' writing process irrespective of time may hide the effect of meaningful and predictive writing behaviours. Moreover, this analysis could be used to predict students writing quality over time and act as a filter for early targeting of students with different writing qualities opening up feedback opportunities.

5.4. Discussion

This study was based on a fully online web authoring tool called Cadmus. Like typical word processing software it allowed the users to type, highlight, delete and update their writing. The availability of separate sections for body text, note-taking and referencing allowed us to separately extract the keystroke logs of different sections. From these logs this study was able to engineer multiple sets of features capturing different aspects of students' writing processes ranging from patterns of pauses, burst and revision summaries as well as time dedicated to note-taking, reference organization, small and large revision, and transcribing behaviours, each of which reflects specific aspects of the writing process.

The performance of a model trained on the pause pattern related features extracted from overall keystrokes (baseline model) was compared against the performance of a model trained on pause pattern related features of each writing section separately. The section-specific model performed slightly better. This indicates that the baseline model may miss or "overlook" on specific moments where the features associated with certain writing phases become more important. The performance of the section specific model was further improved by adding features including burst and revision summaries and percentage of time on specific activities.

The SHAP algorithm was used to interpret the contribution of features on the model's prediction for each student. Visualizing the feature importance of the resultant model, this study found which students' writing characteristics were more associated with writing quality. Consistent with prior work, the total time on task was found as the strongest predictor of text quality. The feature importance was visualized for every student showing the non-linear relationship between some influential features and writing quality (i.e. how specific behaviour that have a positive effect for one student may have a negative effect for another because of the impact of other features). This is consistent with a theory of writing which suggests there exist considerable interaction and overlap between writing phases

(Biggs, 1988). This also emphasises the necessity of using models that capture the interrelationship between features rather than simple correlation and regression analysis.

The current study also investigated students' essay writing process using a temporal approach, utilising learning analytics rather than self-report data. This examined whether during specific moments of writing, the influence of extracted features on writing quality varies. For this purpose, students' keystrokes were divided into several writing episodes which were defined based on the keystrokes used to complete the fixed number of words. This allowed developing a machine learning model to predict the quality of the writing task from an early point and to understand which aspects of writing become important during specific writing episodes. Based on the results, the influence of several features varied across writing episodes indicating the importance of taking the temporal aspects of writing process into account.

This study had several limitations. The first limitation arises from considering all the activities in the note-taking and reference sections to be associated with the related phase of writing. This association is irrespective of the actual written text. Another limitation is the generalizability of the interpretations regarding the influential features. This study was based on the essay writing of undergraduate students with diverse writing and language backgrounds. The influential features might differ for data from a more diverse set of students and across variations in topic, genre and prompts.

5.5. Chapter summary

The goal of this study was to examine the influence of a well-engineered feature set characterizing students' writing process on writing quality over specific drafts of writing (word-based episodes). The result indicated the power of using temporal representations of students' writing process from keystroke logs. Through this study, there is a hope to develop a better understanding of students' writing process in authentic educational settings. The findings can enable educators to identify and then support students' writing processes, thereby promoting higher writing quality. Investigation of the influence of each writing characteristic on the writing quality of each individual student over time, enables educators to provide personalized and strategic instructions to students to support their development of the skills needed to produce high-quality writings.

Chapter 6

Conclusion and Future Directions

This program of research investigated different students' study behaviours in relation to study outcomes and obtained understanding regarding the meaningfulness of their behaviours at different levels and structures when engaging with digital educational environments. In particular, three study behaviours were examined in two different educational contexts: 1) new understanding regarding session behavioural data (how students organised their workload into study sessions along the course weeks) was obtained in a MOOC, 2) The importance and meaningfulness of the sequential nature of students' behaviour was examined toward completing assessment tasks in a MOOC, 3) The importance of the temporal nature of students' writing behaviour was investigated when completing a writing task in a digital writing software.

This chapter summarises the contributions of this thesis, discusses some limitations of the works, and explains potential directions for future studies.

6.1. Summary of contributions

In Chapter 3 of this thesis, the importance and meaningfulness of session behavioural data in online learning have been studied. Although the concept of "session" has been previously considered in research as a unit of analysis, it rarely has been explicitly examined in relation to learning outcome in online learning. This gap has been addressed in this research by an exploratory temporal approach using clustering and group comparison tests, to gain insight into how students organised their session along a MOOC over time and which types of activities they prioritised within these sessions.

This work, through applying temporal analysis, is one way to progress in the direction of characterizing students' time management strategies and understanding the effect of the process used by students during learning. As a result, several patterns associated with learning outcome were found; Successful students had more frequent and longer sessions across the course, mixed up activities within sessions and changed the focus of activities within sessions across the course. These all together can explain high learning outcome based on higher engagement, longer time to engage with more challenging tasks, applying different learning strategies, and focus on whatever activity that is perceived as the most effective one at a particular moment of learning which are consistent with educational theories.

Two major applications of these findings include: First, educators and designers could apply the proposed approach to identify patterns of session behaviours in their courses. This would allow them to identify which session behaviours require support – and (most importantly) when – so they can select relevant interventions to include in their courses. Second, they could implement interventions to support learners' time management and effort regulation skills and expect a positive impact on learners' session behaviour. Such interventions could include providing study plans and tools for time management or helping learners to track and review time dedicated studying and planning ahead, that could lead to improvement of time management skills measures.

In Chapter 4, the importance of sequential behavioural data in online learning was investigated in relation to the learning outcome. This study focused on relating students' behaviours prior to assessment submissions to their subsequent assessment outcomes, that could provide better insight regarding their progress over time. Previous research in online learning has mostly focused on predefined aggregated factors of learners' behaviours and evaluating their effect on their learning outcome. Various prediction models were used and compared for this purpose. While comparing and selecting an appropriate algorithm is a challenging task, switching between algorithms usually results in a minimal change, whereas a proper representation of data can lead to a considerable difference and can be more challenging. In this matter, comparing the effect of modelling students based on different representations of the same behaviour within MOOCs (e.g., sequential and aggregated forms) was understudied by the literature.

This gap has been addressed in this research by a range of learning analytics techniques including prediction models and sequential pattern mining methods based on a MOOC dataset. Through a range of prediction models, a students' profile (prediction model) was built – in two forms of aggregate and sequential – to predict their performance in assessment tasks. This study also investigated whether the most recent activities prior to submissions were more influential on assessment outcomes compared to incorporating the more historic activities applied by students. Additionally, sequences of students' assessment preparation behaviours were explored using a sequential pattern mining method, to expose underlying factors that might influence students' assessment results.

The results suggested it was possible to successfully predict students' assessment outcomes, particularly if the sequential aspects of students' behaviour were represented in the machine learning model. Moreover, the findings confirmed the importance of the overall process students applied in the course as a significant indicator of learning outcome rather than only the most recent activities prior to submissions. Additionally, the results showed how some behavioural patterns reflecting specific learning strategies may be more effective in promoting or undermining learning than others.

A clear implication of the research presented in this study is that the sequential dimension in modelling students' behaviour is important, and there may be a need to account for it in learning analytics research. In addition, the findings of pattern extraction analysis provide insights into how a framework for both educators and adaptive environments could be developed to identify which behaviours require support so that relevant interventions could be provided.

In Chapter 5, the meaningfulness and importance of the temporal nature of students' writing behaviour were explored in relation to their writing outcome using their keystroke logs in a word processing software. Even though there exists much research examining students' writing process in relation to writing outcome, there are only a few that combine writing process and temporal analysis, these are mostly focused on think-aloud procedures and offline measurements that have been criticized as being imprecise representations of the writing process.

This gap has been addressed in this research, by a temporal approach combining prediction models and an additive feature interpretation methods. The analysis was carried out by characterizing students' writing processes connected to isolated writing phases, and examined whether the significance of specific writing characteristics (that predict students' writing outcome) varies at specific moments of writing using a feature interpretation method. The result indicated the power of using temporal representations of students' writing process from keystroke logs. Investigation of the influential writing characteristics for individual students clarified the non-linear relationship between several characteristics and writing quality. This is consistent with a theory of writing which suggests there exist considerable overlap and interaction between writing phases. This also reveals the necessity of using models that capture the interrelationship between features rather than simple correlation or linear regression analysis.

Through this study, there is a hope to develop a better understanding of students' writing processes in authentic educational settings. The findings can enable educators to identify and then support students' writing processes, thereby promoting higher writing quality. Investigation of the influence of each writing characteristic on the writing quality of each individual student over time, enables educators to provide personalized and strategic instructions to students to support their development of the skills needed to produce high-quality writing.

6.2. Future directions

The studies presented in this program of research, addressed different aspects of students' study behaviour based on various data levels, structures and learning contexts in relation to their study outcome. However, there are many potential works left in the context of LA, study behaviour, and study outcome to be addressed in the future.

Chapter 3 and Chapter 4 of this thesis were conducted based on only one MOOC, and the study conducted in Chapter 5 was based on an essay writing task of a subject. Therefore, they have the limitations commonly found in most LA studies, that is the generalizability of the analysis and interpretations of the patterns found. The findings might differ for data from a more diverse set of students and across variations in topic, genre or prompts. Replication considering other MOOCs or writing tasks, particularly with different foci and learning designs, would contribute to a better understanding of the different contextual influences on students' study behaviours, and their relationship with performance.

Moreover, some learners' personal factors related to behaviour and performance in MOOCs and writing task were not examined as they were beyond the scope of this thesis. These include, for example, learners' individual differences, such as age, level of education and course goals, previous online learning experience, and their motivation prior to and throughout the course. Future studies could focus on further examining the impact of these variables on different aspects of students behaviour (e.g., session behaviour, use of specific learning or writing strategy).

Additionally, there may be different approaches to examining students' study behaviours. These include, for example, a qualitative investigation of learners' contribution to the discussion forum and social media that could provide further insight into learners' use of help seeking and peer learning strategies, such as content analysis or natural language processing. Another example, could be differentiating students' writing skills in specific writing phases by including the actual text that they enter, including the body text they produce, their written notes and collected reference materials using content analysis. This may help to distinguish between weak and strong planning behaviour and the related impact on writing quality. Another investigation could be investigating the patterns of students writing bursts, particularly considering the distance from the assignment deadline which may help to distinguish between bursts reflecting fluency of writing and procrastinating behaviour.

Furthermore, future research in online learning could not only benefit from considering different levels and structure of data, but could also focus on examining the relationship between them. For instance, research could be conducted to incorporate findings related to session behavioural data (Chapter 3) into studies that examine students' study behaviour in relation to the use of specific learning strategies (e.g., the study provided in Chapter 4). This would provide a more comprehensive and possibly more actionable suite of interventions for learners taking into consideration how they manage their time when learning online. For example, making suggestions of task sequences (e.g., "How about taking this activity now?") could become specific recommendations considering the length and frequency of sessions at different times of the course and learners' history of session behaviour (e.g., "How about a 50 minutes session this week including activities X, Y and Z?" to a student who has mainly engaged in short sessions that week). In this case, predictive models and the use of adaptive digital environments could be helpful to provide personalised interventions and help learners persist with longer sessions. For example, Mao et al. (2013) created models in the field of online volunteering that predicted the exact moment when volunteers would drop out of sessions. Their model considered features such as time on task, number of sessions, time on each session, time on breaks and features of tasks. The current study provides a first step towards implementing a similar approach in an online educational setting, as we have found sessions to be an important feature related to achievement.

Studying session behaviour as a unit of analysis has not been addressed in *writing analytics* so far. An interesting direction for future research could also be investigating students' writing management over time by considering the temporal features of their writing sessions (the interrupted episodes students dedicate to their writing). This could provide insight into whether the organization and length of students' writing sessions impacts on their writing process and the quality of the writing outcome they produce.

Future research could also be conducted to assess the effectiveness of the findings (behaviours related to students' study outcome) for feedback via a randomized controlled study. In addition, there is a need for deeper understanding of the risks and opportunities of learning analytics from an ethical point of view and more explicit discussions regarding the conflicts and trade-offs between values. For example, the emphasis of making the prediction of complex models explainable and individual-based, and in turn, using the interpretations for automated decision making could improve students' study experiences on one side, but raise the issue of personal data privacy on the other side.

6.3. Conclusion

This thesis studied students' study behaviour in digital educational environments in relation to their study outcome. Findings from this research raise the awareness of LA and EDM research regarding the need to account for various levels of conceptualizing data and different dimensions when studying learners' behaviour to understand their performance in digital educational environments. This work contributes to the advance of the current knowledge base of digital education, learning analytics and student learning, and hopes to lay the foundations for future research.

Bibliography

Agrawal, A., & Gupta, H. (2013). Global K-Means (GKM) clustering algorithm: a survey. *International Journal of Computer Applications*, *79*(2), 20–24. https://doi.org/10.5120/13713-1472

Ainley, M., Corrigan, M., & Richardson, N. (2005). Students, tasks and emotions: Identifying the contribution of emotions to students' reading of popular culture and popular science texts. *Learning and Instruction*, *15*(5), 433–447. https://doi.org/10.1016/j.learninstruc.2005.07.011

Al-Shabandar, R., Hussain, A., Laws, A., Keight, R., Lunn, J., & Radi, N. (2017). Machine learning approaches to predict learning outcomes in Massive open online courses. *The 2017 International Joint Conference on Neural Networks (IJCNN 2017)*, Alaska, USA, May 14–19. https://doi.org/10.1109/IJCNN.2017.7965922

Aldrich, F., Rogers, Y., & Scaife, M. (1998). Getting to grips with "interactivity": helping teachers assess the educational value of CD-ROMs. *British Journal of Educational Technology*, *29*(4), 321–332. https://doi.org/10.1111/1467-8535.00078

Algarni, A. (2016). Data mining in education. *International Journal of Advanced Computer Science and Applications*, *7*(6), 456–461. https://doi.org/10.14569/IJACSA.2016.070659

Allen, G., Davies, R., Ball, N., Albrecht, C., & Bakir, N. (2019). Using educational data mining to identify and analyze student learning strategies. *The 52nd Hawaii International Conference on System Sciences*, Grand Wailea, Maui, Jan 8-11. https://doi.org/10.24251/HICSS.2019.004

Almond, R., Deane, P., Quinlan, T., Wagner, M. & Sydorenko, T. (2012). A preliminary analysis of keystroke log data from a timed writing task. (*ResearchReportNo.RR-12-23*). (Dec. 2012), Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2012.tb02305.x

Alves, R. A., Castro, S. L., De Sousa, L., & Strömqvist, S. (2007). Influence of typing skill on pause-execution cycles in written composition. In Torrance M., van Waes L., & Galbraith D. (Eds.), *Studies in Writing*. *Writing and cognition: Research and applications* (pp. 55–65). Amsterdam: Elsevier.

Arthur, D., & Vassilvitskii, S. (2007). K-means++: the advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (SODA 2007), New Orleans, Louisiana, USA, Jan 7-9.

https://doi.org/10.1145/1283383.1283494

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, *130*(2), 224–237. https://doi.org/10.1037//0096-3445.130.2.224

Astani, M., Ready, K. J., & Duplaga, E. A. (2010). Online course experience matters: investigating students' perceptions of online learning. *Issues in Information Systems*, *11*(2), 14–21.

Baaijen, V. M., Galbraith, D., & de Glopper, K. (2012). Keystroke analysis: Reflections on procedures and measures. *Written Communication*, *29*(3), 246–277. https://doi.org/10.1177/0741088312451108

Baker, R. S. J. d., Inventado, P.S. (2014) Educational data mining and learning analytics. In Larusson J.A., White B. (Eds.), *Learning Analytics: From Research to Practice*. Berlin, Germany: Springer. http://www.columbia.edu/~rsb2162/Educational%20Data%20Mining%20and%20Learning%20Analytics%20-%20DRAFT.pdf

Baker, R. S. J. d., & Yacef, K. (2009). The state of educational data mining in 2009: a review and future visions. *Journal of Educational Data Mining*, *1*(1), 3–17. http://www.columbia.edu/~rsb2162/BakerYacef.pdf

Bakhshinategh, B., Zaïane, O., Elatia, S., & Ipperciel, D. (2017). Educational data mining applications and tasks: a survey of the last 10 years. *Education and Information Technologies*, *23*(1), 537–553. https://doi.org/10.1007/s10639-017-9616-z

Balfour, S. (2013). Assessing writing in MOOCs: Automated essay scoring and calibrated peer review. *Research & Practice in Assessment*, *8*, 40–48. Retrieved from http://www.rpajournal.com/dev/wp-content/uploads/2013/05/SF4.pdf

Barbour, M. (2016). Emergence and innovation in digital learning: foundations and applications. In Veletsianos G. (Eds.), *Emergence and Innovation in Digital Learning: Foundations and Applications*. Athabasca University Press. https://doi.org/10.15215/aupress/9781771991490.01

Barkaoui, K. (2010). Think-aloud protocols in research on essay rating: An empirical study of their veridicality and reactivity. *Language Testing*, *28*(1), 51–75. https://doi.org/10.1177/0265532210376379

Bauckhage, C., Drachen, A., & Sifa, R. (2015). Clustering game behavior data. *IEEE Transactions on Computational Intelligence and AI in Games*, *7*(3), 266–278. https://doi.org/10.1109/TCIAIG.2014.2376982

Bektik, D. (2019). Issues and challenges for implementing writing analytics at higher education. In Ifenthaler D., Mah D. K., & Yau, J. K. (Eds.) *Utilizing*

Learning Analytics to Support Study Success. Springer International Publishing, pp. 143–155. https://doi.org/10.1007/978-3-319-64792-0_9

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*(1), 281–305. Retrieved from http://dblp.uni trier.de/db/journals/jmlr/jmlr13.html#BergstraB12

Bertolaccini, L., Solli, P., Pardolesi, A., & Pasini, A. (2017). An overview of the use of artificial neural networks in lung cancer research. *Journal of Thoracic Disease*, *9*(4), 924–931. https://doi.org/10.21037/jtd.2017.03.157

Biggs, J. (1988). The role of metacognition in enhancing learning. *Australian Journal of Education*, *32*(2), 127–138. https://doi.org/10.1177/000494418803200201

Blanco, A., Serrano, A., Freire, M., Martinez-Ortiz, I., & Fernández-Manjón, B. (2013). E-learning standards and learning analytics. Can data collection be improved by using standard data models? *IEEE Global Engineering Education Conference (EDUCON)*, Berlin, Germany, March 13–15. https://doi.org/10.1109/EduCon.2013.6530268

Bland, J. M., & Altman, D. G. (1995). Statistics notes: multiple significance tests: the Bonferroni method. *British Medical Journal (BMJ)*, *310*(6973), 170–170. https://doi.org/10.1136/bmj.310.6973.170

Bousbia, N., & Belamri, I. (2014). Which contribution does EDM provide to computer-based learning environments? *Studies in Computational Intelligence*, *524*, 3–28. https://doi.org/10.1007/978-3-319-02738-8-1

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. https://doi.org/10.1109/TSP.2016.2546228

Brinton, C. G., Rill, R., Ha, S., Chiang, M., Smith, R., & Ju, W. (2015). Individualization for education at scale: MIIC design and preliminary evaluation. *IEEE Transactions on Learning Technologies*, *8*(1), 136–148. https://doi.org/10.1109/TLT.2014.2370635

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

Brooker, A., Corrin, L., de Barba, P., Lodge, J., & Kennedy, G. (2018). A tale of two MOOCs: how student motivation and participation predict learning outcomes in different MOOCs. *Australasian Journal of Educational*

Technology, *34*(1), 73–87. https://doi.org/10.14742/ajet.3237

Buckingham S. S., 2012. Learning analytics: policy brief. Moscow: UNESCO Institute for Information Technologies in Education, www.iite.unesco.org/pics/publications/en/fi les/3214711.pdf

Castro, F., Vellido, A., Nebot, A., Mugica, F. (2007): Applying Data Mining Techniques to e-Learning Problems. In Jain L.C., & Tedman R.A., & Tedman D.K. (Eds.) E*volution of Teaching and Learning Paradigms in Intelligent Environment. Studies in Computational Intelligence*, 62 (183-221). Springer-Verlag Berlin Heidelberg, ISBN 978-3-540-71973-1

Chanquoy, L. (2009). Revision processes. In *The SAGE Handbook of Writing Development* (80–97), In Beard R., Myhill D., Riley J., & Nystrand M. (Eds.). London EC1Y 1SP United Kingdom: London: SAGE Publications Ltd. https://doi.org/10.4135/9780857021069.n6

Chatti, M., Dyckhoff, A., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, *4(5/6)*, 318–331. https://doi.org/10.1504/IJTEL.2012.051815

Chen, C. S. (2002). Self-regulated learning strategies and achievement in an introduction to information systems course. *Information Technology Learning and Performance Journal*, *20*(1), 11–25. https://pdfs.semanticscholar.org/d0b2/97b5bf6e03cbc18e9eb7a8b936f47a2f3227.pdf

Chen, T., & Guestrin, C. (2016). XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, California, USA, Aug 13-17. https://doi.org/10.1145/2939672.2939785

Chen, W., Brinton, C. G., Cao, D., Mason-Singh, A., Lu, C., & Chiang, M. (2018). Early detection prediction of learning outcomes in online short-courses via learning behaviors. *IEEE Transactions on Learning Technologies*, *12*(1), 44–58. https://doi.org/10.1109/TLT.2018.2793193

Christianson, K., Mestre, J. P., & Luke, S. G. (2012). Practice makes (nearly) perfect: solving 'students-and-professors'-type algebra word problems. *Applied Cognitive Psychology*, *26*(5), 810–822. https://doi.org/10.1002/acp.2863

Chukharev-Hudilainen, E. (2014). Pauses in spontaneous written communication: a keystroke logging study. *Journal of Writing Research*, *6*(1), 61–84. https://doi.org/10.17239/jowr-2014.06.01.3

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. In S. Buckingham Shum, D. Gašević, & R. Ferguson (Eds.), In *Proceedings of the*

2nd International Conference on Learning Analytics and Knowledge (pp. 134-138). New York, NY: ACM Digital Library. doi: 10.1145/2330601.2330636

Coffrin, C., Corrin, L., de Barba, P., & Kennedy, G. (2014). Visualizing patterns of student engagement and performance in MOOCs. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge (LAK)*. Indianapolis, USA, March 24–28. https://doi.org/10.1145/2567574.2567586

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, *4*(4), 253–278. https://doi.org/10.1007/BF01099821

D. Cummins, K. Yacef, I. Koprinska (2006). A sequence based recommender system for learning resources, *Australian Journal of Intelligent Information Processing Systems*, 9(2), 49–56. http://www.it.usyd.edu.au/~irena/ajiips06.pdf

Dabbagh, N., & Kitsantas, A. (2004). Supporting self-regulation in student-centered web-based learning environments. *International Journal on E-Learning*, *3*(1), 40–47. Retrieved from https://eric.ed.gov/?id=EJ723806

Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth. *Paradox and Possibility*, *25*, 321–323. https://doi.org/10.5334/2012-18

Daniel, W.W. 1990. Spearman rank correlation coefficient. *Applied Nonparametric Statistics* (2nd ed.). 358–365. Boston, MA: PWS-Kent. https://lib.ugent.be/en/catalog/rug01:000283035

Davis, D., Chen, G., van der Zee, T., Hauff, C., & Houben, G. J. (2016). Retrieval practice and study planning in MOOCs: exploring Classroom-based self-regulated learning strategies at scale. In *Proceedings of the 11th European Conference on Technology Enhanced Learning (EC-TEL 2016)*, Lyon, France, Sep 13-16 . https://doi.org/10.1007/978-3-319-45153-4_5

de Barba, P. G. (2017). Autonomous learning and achievement motivation in online learning environments. PhD thesis, The University of Melbourne, Australia. http://hdl.handle.net/11343/212132

de Barba, P. G., 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. https://doi.org/10.1111/jcal.12130

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, *18*(1), 7–24. https://doi.org/10.1016/j.asw.2012.10.002

Deane, P. (2014). Using writing process and product features to assess writing quality and explore how those features relate to other literacy tasks. *ETS Research Report Series 14-03*, Princeton, NJ: Educational testing Service. https://doi.org/10.1002/ets2.12002

DeBoer, J., Stump, G. S., Seaton, D., & Breslow, L. (2013). Diversity in MOOC students' backgrounds and behaviors in relationship to performance in 6.002 x. In *Proceedings of the 6th International Conference of MIT's Learning International Networks Consortium (LINC)*, Cambridge, Massachusetts, June 16–19.

Dembo, M. H., & Seli, H. (2012). Motivation and learning strategies for college success: a focus on self-regulated learning (4th Ed.). New York, NY, USA: Routledge. https://routledgetextbooks.com/textbooks/_author/dembo-9780415894203/

Van den Bergh, H., & Rijlaarsdam, G. (1996). The dynamics of composing: Modeling writing process data. In Levy C. M., & Ransdell S. (Eds.), *The science of writing: theories, methods, individual differences, and applications* (pp. 207–232). New York: Erlbaum. https://hdl.handle.net/11245/1.126018

Di, W., Bhardwaj, A., & Wei, J. (2018). Deep learning essentials: Your hands-on guide to the fundamentals of deep learning and neural network modeling. Packt Publishing, Birmingham. https://dl.acm.org/citation.cfm?id=3208611

Diana, N., Eagle, M., Stamper, J., & Koedinger, K. R. (2016). Extracting measures of active learning and student self-regulated learning strategies from MOOC data. In *Proceedings of the 9th International Conference on Educational Data Mining* (EDM). North Carolina, USA, June 29–July 2. Retrieved from https://dblp.org/db/conf/edm/edm2016

Du, M., Liu, N., & Hu, X. (2018). Techniques for interpretable machine learning, arXiv preprint. arXiv:1808.00033

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). Pattern classification (2nd ed.). New York, NY, USA: Wiley-Interscience.

Farrow, R. (2019). Massive open online courses for business learning: key research, best practices and pathways to innovation. Pressbooks. http://oro.open.ac.uk/58992/

Fautley, M., & Savage, J. (2008). Assessment for learning and teaching in secondary schools. Exeter, UK: Learning Matters. http://dx.doi.org/10.4135/9781446278642

Feiz, P., Hooman, H. A., & kooshki, S. (2013). Assessing the motivated strategies for learning questionnaire (MSLQ) in iranian students: construct validity and reliability. *Procedia - Social and Behavioral Sciences*, *84*(9), 1820–1825. https://doi.org/10.1016/j.sbspro.2013.07.041

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

Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G. (2019). Educational data mining: predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research*, *94*, 335–343. https://doi.org/10.1016/j.jbusres.2018.02.012

Fincham, O., Gasevic, D., Jovanovic, J., & Pardo, A. (2018). from study tactics to learning strategies: An analytical method for extracting interpretable representations. *IEEE Transactions on Learning Technologies, 12*(1), 59–72 https://doi.org/10.1109/TLT.2018.2823317

Fisher, R. A. (1922). On the interpretation of χ 2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society*, *85*(1), 87–94. https://doi.org/10.2307/2340521

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*(4), 365–387. Retrieved from http://www.jstor.org/stable/356600

Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge discovery in databases: an overview. *AI Magazine*, *13*(3), 57–70. https://doi.org/10.1609/aimag.v13i3.1011

Gagne, R. M. (1977). The conditions of learning (3rd Ed.), New York Holt. *Rinehart and Winston*.

Gardner, J., & Brooks, C. (2017). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction, 28*(2), 127–203 https://doi.org/10.1007/s11257-018-9203-z

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: learning analytics are about learning. *TechTrends*, *59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Gasevic, D., Jovanovic, J., Pardo, A., & Dawson, S. (2017). Detecting learning strategies with analytics: links with self-reported measures and academic performance. *Journal of Learning Analytics*, *4*(2), 113–128. https://doi.org/10.18608/jla.2017.42.10

Gers, F., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: continual prediction with LSTM. *Neural Computation*, *12*(10), 2451–2471. https://doi.org/10.1162/089976600300015015

Gibbs, G., & Simpson, C. (2005). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, (1), 3–31. https://doi.org/10.1007/978-3-8348-9837-1

Gibson, A., & Lang, C. (2018). The pragmatic maxim as learning analytics research method, In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK18)*, 7-9 March 2018, Sydney, NSW, Australia. https://doi.org/10.1145/3170358.3170384

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: a review of the literature. *Computers & Education*, *57*(4), 2333–2351. https://doi.org/10.1016/j.compedu.2011.06.004

Goldberg, Y. (2017). Neural network methods for natural language processing. *Synthesis Lectures on Human Language Technologies*, *10*(1), 1–309. https://doi.org/10.2200/S00762ED1V01Y201703HLT037

Graham, S., & Harris, K. R. (2013). Designing an effective writing program. In S. Graham, C. MacArthur, & J. Fitzgerald (Eds.), *Best practices in writing instruction* (2nd ed., pp. 3–25). New York: Guilford.

Graham, S., Harris, K. R., & Santangelo, T. (2015). Research-based writing practices and the common core: meta-analysis and meta-synthesis. *The Elementary School Journal*, *115*(4), 498–522. https://doi.org/10.1086/681964

Green, D. W., & Wason, P. C. (1982). Notes on the psychology of writing. *Human Relations*, *35*(1), 47–56. https://doi.org/10.1177/001872678203500104

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: a generic framework for learning analytics. *Educational Technology & Society*, *15*(42), 42–57. https://eric.ed.gov/?id=EJ992502

Grigorova, K., Malysheva, E., & Bobrovskiy, S. (2017). Application of data mining and process mining approaches for improving e-learning processes, In *Proceedings of the 3rd International conference Information Technology and Nanotechnology (ITNT), Samara, Russia, Apr 25-27*. https://doi.org/10.18287/1613-0073-2017-1903-115-121

Gulbahar, Y., & Ilgaz, H. (2014). Premise of learning analytics for educational context: Through concept to practice. *Bilişim Teknolojileri Dergisi*, *7*(3), 12–20. https://doi.org/10.12973/bid.2020

Guo, H., Deane, P. D., van Rijn, P. W., Zhang, M., & Bennett, R. E. (2018). Modeling basic writing processes from keystroke logs. *Journal of Educational Measurement*, *55*(2), 194–216. https://doi.org/10.1111/jedm.12172

Guo, P., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos, In *Proceedings of the first ACM conference on Learning @ scale conference*, Atlanta, Georgia, USA, March 4-5. https://doi.org/10.1145/2556325.2566239

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, *17*(2/3), 107–145. https://doi.org/10.1023/A:1012801612483

Hayes, J., & Gradwohl Nash, J. (1996). On the nature of planning in writing. In *The Science of Writing:Theories, Methods, Individual Differences, and Applications.* (pp. 29–55). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc. 29–55

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication*, *29*(3), 369–388. https://doi.org/10.1177/0741088312451260

He, J., Bailey, J., Rubinstein, B. I. P., & Zhang, R. (2015). Identifying at-risk students in massive open online courses. In *29th AAAI Conference on Artificial Intelligence*, Austin, Texas, Jan 25–29. http://dl.acm.org/citation.cfm?id=2886521.2886563

Howard, S. K., & Mozejko, A. (2015). Considering the history of digital technologies in education. In Henderson M., & Romero G. (Eds.), *Teaching and Digital Technologies: Big Issues and Critical Questions* (157-168). Port Melbourne, Australia: Cambridge University Press. https://ro.uow.edu.au/sspapers/1831/

Hussain, M., Zhu, W., Zhang, W., Abidi, S. M. R., & Ali, S. (2018). Using machine learning to predict student difficulties from learning session data. *Artificial I*ntelligence *Review*, 52(1), 381–407. https://doi.org/10.1007/s10462-018-9620-8

Hyland, K. (2013). Writing in the university: education, knowledge and reputation. *Language Teaching*, 46(1), 53–70. https://doi.org/10.1017/S0261444811000036

Indrisano, R., & Squire, J. R. (2000). Perspectives on writing: research, theory, and practice. Newark, Dela, International Reading Association.

Itani, A., Brisson, L., & Garlatti, S. (2018). Understanding learner's drop-out in MOOCs. I*n Proceedings of the 19th International Conference on Intelligent Data Engineering and Automated Learning*, Madrid, Spain, Nov 21–23. https://doi.org/10.1007/978-3-030-03493-1_25

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys*, 31(3), 264–323. https://doi.org/10.1145/331499.331504

Jiang, S., Williams, A., Schenke, K., Warschauer, M., & O'dowd, D. (2014). Predicting MOOC performance with week 1 behavior. I*n Proceedings of the*

7th International Conference on Educational Data Mining (EDM'14), London, UK, July 4-7.

Jones, M., Alexander, J., & Estell, D. (2010). Homophily among peer groups members' perceived self-regulated learning. *The Journal of Experimental Education*, *78*(3), 378–394. https://doi.org/10.1080/00220970903548020

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

Judd, T., & Kennedy, G. (2011). Measurement and evidence of computer-based task switching and multitasking by 'net generation' students. *Computers & Eduction*, 56(3), 625–631. https://doi.org/10.1016/j.compedu.2010.10.004

Kayapinar, U. (2014). Measuring essay assessment: intra-rater and inter-rater reliability. *Eurasian Journal of Educational Research*, 14(57), 113–135. https://doi.org/10.14689/ejer.2014.57.2

Ke, G., Meng, Q., Finely, T., Wang, T., Chen, W., Ma, W. Liu, T.Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Proceedings of the 31st Conference on Neural Information Processing Systems (N*IPS 2017), Long Beach, CA, USA. Dec 4-9. https://dl.acm.org/citation.cfm?id=3295074

Kennedy, G. (2004). Promoting cognition in multimedia interactivity research. *Journal of Interactive Learning Research*, 15(1), 43–61. https://www.learntechlib.org/p/4530

Khan, I., & Pardo, A. (2016). Data2U: scalable real time student feedback in active learning environments, In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK'16)*, Edinburgh, United Kingdom, April 25 - 29. https://doi.org/10.1145/2883851.2883911

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. https://doi.org/10.1016/j.compedu.2016.10.001

Kizilcec, R., 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 (LAK'13)*, Leuven, Belgium, April 8-13. https://doi.org/10.1145/2460296.2460330

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of cluster in k-means clustering. *International Journal of Advanced Research in Computer Science and Management Studies*, 1(6), 90–95.

http://www.ijarcsms.com/

Koedinger, K. R., D’Mello, S., McLaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(4), 333–353. https://doi.org/10.1002/wcs.1350

Kotsiantis, S., & Pintelas, P. (2005). Predicting students marks in hellenic open university. In *Proceedings of the 5th IEEE International Conference on Advanced Learning Technologies (ICALT'05)*, Kaohsiung, Taiwan, July 5-8. https://doi.org/10.1109/ICALT.2005.223

Kovačić, Z., & Nz. (2010). Early prediction of student success: mining students enrolment data, In *Proceedings of Infoming Science & IT Education Conference (inSITE)*, Cassino, Italy, Jun 19-24. https://doi.org/10.28945/1281

Kovanovic, V., Gasevic, D., Dawson, S., Joksimovic, S., Baker, R., & Hatala, M. (2015). Does time-on-task estimation matter? implications on validity of learning analytics findings. *Journal of Learning Analytics, 2*(3), 81–110. https://doi.org/10.18608/jla.2015.23.6

Kuh, G., Kinzie, J., Buckley, J., Ed Bridges, & Ed Hayek. (2007). Piecing together the student success puzzle: research, propositions, and recommendations. *ASHE Higher Education Report*, 32(5), 1–182. ASHE Higher Education Report. https://eric.ed.gov/?id=EJ791634

Larusson, J. A., & White, B. (2014). Learning analytics: from research to practice. New York: Springer. https://doi.org/10.1007/978-1-4614-3305-7

Latif, M. M. A. (2008). A state-of-the-art review of the real-time computer-aided study of the writing process. *International Journal of English Studies*, 8(1), 29–50. https://doi.org/10.6018/ijes.8.1.49081

Leijten, M., & Van Waes, L. (2005). Inputlog: a logging tool for the research of writing processes. Research paper, University of Antwerp, department of management, faculty of applied economics.

Grabowski, J. (1996). Writing and speaking: Common grounds and differences towards a regulation theory of written language production. Levy, C. M., & Ransdell, S. E. (Eds.). *The science of writing: theories, methods, individual differences, and applications* (pp. 73-92). Mahwah, NJ: Lawrence Erlbaum Associates.

Liñán, L. C., & Pérez, Á. A. J. (2015). Educational data mining and learning analytics: differences, similarities, and time evolution. *International Journal of Educational Technology in Higher Education*, *12*(3), 98–112. https://doi.org/10.7238/rusc.v12i3.2515

Ling, C. X., Huang, J., & Zhang, H. (2003). AUC: a statistically consistent and more discriminating measure than accuracy. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI'03)*, Acapulco, Mexico, Aug 9-15. https://dl.acm.org/citation.cfm?id=1630659

Liyanagunawardena, T. R., Adams, A. A., & Williams, S. A. (2013). MOOCs: a systematic study of the published literature 2008-2012. *International Review of Research in Open and Distributed Learning,* 14(3), 202–227. http://centaur.reading.ac.uk/33109/

Lockwood, J. (2014). Handbook of automated essay evaluation current applications and new directions. Shermis, M. D. & Burstein J. (eds.) (2013) New York: Routledge. *Writing & Pedagogy*, 6(2), 437–441. https://doi.org/10.1558/wap.v6i2.437

Long, P. & Siemens, G. (2011). Penetrating the fog: analytics in learning and education. *TD Tecnologie Didattiche,* 22(3), 132–137 https://doi.org/10.17471/2499-4324/195

López de Prado, M. (2018). Advances in financial machine learning (Edition, 2). John Wiley & Sons.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*, Long Beach, CA, USA, Dec 4-9. https://dl.acm.org/citation.cfm?id=3295230

Ma, X. (2018). Machine learning with adversarial perturbations and noisy labels. PhD Thesis, Univeristy of Melbourne. Australia. http://hdl.handle.net/11343/219680

Mao, A., Kamar, E., & Horvitz, E. (2013). Why stop now? predicting worker engagement in online crowdsourcing. In *Proceedings of the 1st AAAI Conference on Human Computation and Crowdsourcing*, Palm Springs, California, USA, Nov 7-9. https://www.aaai.org/

Matsuhashi, A. (1981). Pausing and planning: the tempo of written discourse production. *Research in the Teaching of English*, *15*(2), 113–134. http://www.jstor.org/stable/40170920

Mccardle, L., & Hadwin, A. (2015). Using multiple, contextualized data sources to measure learners' perceptions of their self-regulated learning. *Metacognition and Learning*, *10*(1), 43–47. https://doi.org/10.1007/s11409-014-9132-0

McCutchen, D. (2000). Knowledge, processing, and working memory: implications for a theory of writing. *Educational Psychologist*, 35(1), 13–23. https://doi.org/10.1207/S15326985EP3501_3

Merceron, A., & Yacef, K. (2010). Measuring correlation of strong symmetric association rules in educational data. In Romero, C., Ventura S., Pechenizikiy M., Baker R. S. J. d. (Eds.), *Handbook of Educational Data Mining* (245–256). Boca Raton, FL: CRC Press, https://doi.org/10.120/b10274-20

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52(*4), 463–477. https://doi.org/10.1016/j.jml.2004.12.001

Minn, S., Yu, Y., Desmarais, M. C., Zhu, F., & Vie, J. J. (2018). Deep knowledge tracing and dynamic student classification for knowledge tracing. *In Proceedings of the 20th International Conference on Data Mining (ICDM), Istanbul, Turkey, Jul 23-24*.

Molnar, C. (2019). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. Accessed: April. 16, 2020. [Online] Retrieved from https://christophm.github.io/interpretable-ml-book/

Morgan, M. (1987). Self-monitoring and goal setting in private study. *Contemporary Educational Psychology*, 12(1), 1–6. https://doi.org/10.1016/S0361-476X(87)80033-4

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*. Seoul, Republic of Korea, Apr 7-11. https://doi.org/10.1145/2567948.2580054

Nguyen, Q., Huptych, M., & Rienties, B. (2018). Using temporal analytics to detect inconsistencies between learning design and student behaviours. *Journal of Learning Analytics*, 5(3), 120–135. https://doi.org/10.18608/jla.2018.53.8

Okubo, F., Yamashita, T., Shimada, A., & Ogata, H. (2017). A neural network approach for students' performance prediction. In *Proceedings of the 7th International conference on Learning Analytics & Knowledge (LAK'17)*, Vancouver, British Columbia, Canada, March 13–17. https://doi.org/10.1145/3027385.3029479

Oyelade, J., Oladipupo, O., & Obagbuwa, I. (2010). Application of k-means clustering algorithm for prediction of students academic performance. *International Journal of Computer Science and Information* Security, 7(1), 292–295. https://arxiv.org/ftp/arxiv/papers/1002/1002.2425.pdf

Pal, S. K., & Mitra, S. (1992). Multilayer perceptron, fuzzy sets, and classification. *IEEE Transactions on Neural Networks*, 3(5), 683–697. https://doi.org/10.1109/72.159058

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: a systematic literature review of empirical evidence. *Journal of Educational Technology & Society*, *17*(4), 49–64. https://www.jstor.org/stable/jeductechsoci.17.4.49

Pardo, A., Han, F., & Ellis, R. A. (2016). Exploring the relation between self-regulation, online activities, and academic performance. In *Proceedings of the 6th International Conference on Learning Analytics & Knowledge* (LAK'16). Edinburgh, Scotland, UK, Apr 25-26. https://doi.org/10.1145/2883851.2883883

Pardos, Z., Bergner, Y., Seaton, D., & Pritchard, D. (2013). Adapting bayesian knowledge tracing to a massive open online course in edx. In *Proceedings of the 6th International Conference on Educational Data Mining (EDM'13)*. Memphis, TN, USA, July 6-9. https://doi.org/10.1.1.668.6548

Park, J., Denaro, K., Rodriguez, F., Smyth, P., & Warschauer, M. (2017). Detecting changes in student behavior from clickstream data. In *Proceedings of the 7th International Conference on Learning Analytics & Knowledge (LAK'17)*, New York, NY, USA. https://doi.org/10.1145/3027385.3027430

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaïane, O. R. (2009). Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(6), 759–772. https://doi.org/10.1109/TKDE.2008.138

Peters, O. (2000). Digital learning environments: New possibilities and opportunities. *The International Review of Research in Open and Distributed Learning*, 1(1). 1–19. https://doi.org/10.19173/irrodl.v1i1.3

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In *Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS'15)*. Montreal, Canada, Dec 7–12. https://arxiv.org/abs/1506.05908

Pintrich, P. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology, 25* (1), 92–104. https://doi.org/10.1006/ceps.1999.1017

Pintrich, P., & Duncan, T. (1991). Student goal orientation and self-regulation in the college classroom. In Maehr M. L., & Pintrich P. R. (Eds.), *Advances in Motivation and Achievement* (371–402). Greenwich, CT: JAI Press.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In Boekaerts M., Pintrich P. R., & Zeidner M. (Eds.), Handbook of self-regulation (451–502). Academic Press. https://doi.org/10.1016/B978-012109890-2/50043-3

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., & DeGroot, E. (1990). Quantitative and qualitative perspectives on student motivational beliefs and self-regulated learning. I*n Annual American Educational Research Association*, Boston, Massachusetts, Apr 16-20.

Pintrich, P. R., Ryan, A. M., & Patrick, H. (1998). The differential impact of task value and mastery orientation on males' and females' self-regulated learning. In Interest and learning: In Hoffmann A. K. L., Renninger K. A., & Baumert J. (Eds.), *Interest and learning* (pp. 337–52). Kiel: Institute for Science Education (IPN)

Pintrich, P., Smith, D., Duncan, T., & Mckeachie, W. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ). Ann Arbor. Michigan.

Poellhuber, B., & Roy, N. (2019). Understanding Participant's behaviour in massively open online courses. *International Review of Research in Open and Distance Learning*, *20*(1), 222–242. https://doi.org/10.19173/irrodl.v20i1.3709

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

Qiu, J., Tang, J., Liu, T., Gong, J., Zhang, C., Zhang, Q., & Xue, Y. (2016). Modeling and predicting learning behavior in MOOCs. In *Proceedings of he 9th ACM International Conference on Web Search and Data Mining (WSDM'16)*, San Francisco, CA, USA, Feb 22–25. https://doi.org/10.1145/2835776.2835842

Richter, S. (2012). Learning tasks. In Seel N. M. (Eds.), *Encyclopedia of the sciences of learning* (1975–1979). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4419-1428-6_342

Riel, J., Lawless, K., & Brown, S. (2018). Timing matters: approaches for measuring and visualizing behaviours of timing and spacing of work in self-paced online teacher professional development courses. *Journal of Learning Analytics, 5*(1), 25–40. https://doi.org/10.18608/jla.2018.51.3

Rijlaarsdam, G., Weijen, D., & Bergh, H. (1994). Relations between writing processes and text quality: When and how? *Cognition and Instruction*, *12*(2), 103–123. https://doi.org/10.1207/s1532690xci1202_2

Rizvi, S., Rienties, B., Rogaten, J., & Kizilcec, R. F. (2019). Investigating variation in learning processes in a FutureLearn MOOC. *Journal of*

Computing in Higher Education, 12, 1–20. https://doi.org/10.1007/s12528-019-09231-0

Rodrigo, M. M. T., & Beck, J. B. (2014). Understanding wheel spinning in the context of affective factors. *International Journal of People-Oriented Programming*, *3*(2), 75–90. https://doi.org/10.4018/IJPOP.2014070104

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rogers, Y., & Scaife, M. (1998). How can interactive multimedia facilitate learning. In Lee J. (Eds.) Intelligence and Multimodality in Multimedia Interfaces: *Research and Applications*. AAAI. Press: Menlo Park, CA.

Rohman, D. G. (1965). Pre-writing the stage of discovery in the writing process. *College Composition and Communication*, 16(2), 106–112. http://www.jstor.org/stable/354885

Rohrer, D. (2015). Student instruction should be distributed over long time periods. *Educational Psychology Review*, 27(4), 635–643. https://doi.org/10.1007/s10648-015-9332-4

Romero, C., & Ventura, S. (2007). Educational data mining: a survey from 1995 to 2005. *Expert Systems with Applications*, *33*(1), 135–146. https://doi.org/10.1016/j.eswa.2006.04.005

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12–27. https://doi.org/10.1002/widm.1075

Romero, C., Ventura, S., Espejo, P., & Martínez, C. (2008). Data mining algorithms to classify students. In *Proceedings of the 1st International Conference on Educational Data Mining (EDM'08)*, Montreal, Québec, Canada, Jun 20-21.

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. (2010). *Handbook of educational data mining*. Boca Raton: CRC Press.

Rosenqvist, S. (2015). Developing pause thresholds for keystroke logging analysis. B.A. thesis. University of Umea, Sweden. Retrieved from http://umu.diva-portal.org/smash/get/diva2:834468/FULLTEXT01.pdf.

Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Rubens, W. (2014). Improving the learning design of massive open online courses. *Turkish Online Journal of Educational Technology-TOJET*, *13*(4), 71–80.

https://www.learntechlib.org/p/155523/

Rusli, N., Ibrahim, Z., & Mohd Janor, R. (2008). Predicting students' academic achievement: comparison between logistic regression, artificial neural network, and Neuro-fuzzy. In *Proceedings of International Symposium on Information Technology (ITSim),* Kuala Lumpur, Malaysia, Aug 26–29. https://doi.org/10.1109/ITSIM.2008.4631535

Russell, D. R. (1997). Rethinking genre in school and society: an activity theory analysis. *Written Communication*, 14(4), 504–554. https://doi.org/10.1177/0741088397014004004

Saha, S., & Waheed, S. (2017). Credit risk of bank customers can be predicted from customer's attribute using neural network. *International Journal of Computer Applications*, 161(3), 39–43. https://doi.org/10.5120/ijca2017913170

Sara, N.-B., Halland, R., Igel, C., & Alstrup, S. (2015). High-School dropout prediction using machine learning: a danish large-scale Study. In *Proceedings of the 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, Belgium, Apr 22-24. http://www.i6doc.com/en/.

Schunk, D. H., & Zimmerman, B. J. (1994). *Self-regulation of learning and performance: Issues and educational applications*. Hillsdale, NJ: Erlbaum Associates.

Scriven, M., & Paul, R. (1987). Defining critical thinking. *In 8th Annual International Conference on Critical Thinking and Education Reform*, Summer 1987.

Sedgwick, P. (2012). Pearson's correlation coefficient. *British Medical Journal (BMJ)*, 345, e4483–e4483. https://doi.org/10.1136/bmj.e4483

Shah, D. (2018). By the numbers: MOOCS in 2017. *Class Central Report.* https://www.class-central.com/report/mooc-stats-2017/

Shaikhina, T., & Khovanova, N. A. (2017). Handling limited datasets with neural networks in medical applications: a small-data approach. *Artificial Intelligence in Medicine*, 75, 51–63. https://doi.org/10.1016/j.artmed.2016.12.003

Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2(1), 41–50. https://doi.org/10.1109/TETCI.2017.2772792

Siemens, G., & Gasevic, D. (2012). Guest editorial-learning and knowledge analytics. *Journal of Educational Technology & Society*, 15(3), 1–2.

Smith, D., Campbell, J., & Brooker, R. (1999). The impact of students' approaches to essay writing on the quality of their essays. *Assessment & Evaluation in Higher Education*, *24*(3), 327–338. https://doi.org/10.1080/0260293990240306

Sommers, N. (1980). Revision strategies of student writers and experienced adult writers. *College Composition and Communication, 31*(4), 378–388. http://www.jstor.org/stable/356588

Son, L., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. journal of experimental psychology. *Learning, Memory, and Cognition, 26*(1), 204–221. https://doi.org/10.1037/0278-7393.26.1.204

Stallard, C. K. (2019). An analysis of the writing behavior of good student writers [microform]. *Research in the Teaching of English*, *8*(2), 206–218. https://www.jstor.org/stable/40170525

Struyf, A., Hubert, M., & Rousseeuw, P. (1996). Clustering in an object-oriented environment. *Journal of Statistical Software*, 1(4), 1–30. https://doi.org/10.18637/jss.v001.i04

Tabuenca, B., Kalz, M., Drachsler, H., & Specht, M. (2015). Time will tell: The role of mobile learning analytics in self-regulated learning. *Computers & Education*, 89, 53–74. https://doi.org/10.1016/j.compedu.2015.08.004

Tanuar, E., Heryadi, Y., Lukas, Abbas, B. S., & Gaol, F. L. (2018). Using machine learning techniques to earlier predict student's performance. In *Proceedings of 2018 Indonesian Association for Pattern Recognition International Conference (INAPR)*, Jakarta, Indonesia, Sep 7-8.

Taylor, R. (2012). Review of the motivated strategies for learning questionnaire (MSLQ) using reliability generalization techniques to assess scale reliability. PhD Thesis, Auburn University, Alabama. http://hdl.handle.net/10415/3114

Tillema, M., van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2011). Relating self reports of writing behaviour and online task execution using a temporal model. *Metacognition and Learning, 6(*3), 229–253. https://doi.org/10.1007/s11409-011-9072-x

Tough, A. (1979). The adult's learning projects: a fresh approach to theory and practice in adult learning (2nd ed.). Toronto: Ontario Institute for Studies in Education.

Trezise, K., de Barba, P. G., Jennens, D., Zarebski, A., Russo, R., & Kennedy, G. (2017). A learning analytics view of students' use of self-regulation strategies for essay writing. In *Proceedings of the 34th International Conference of Innovation, Practive and Research in The use of Educational Technologies in Tertiary Education (ASCILITE 2017)*, Toowoomba, QLD, Australia, Dec

4-6.

Tukey, J. W.(1977). Exploratory data analysis. reading, MA: Addison-Wesley.

Umer, R., Susnjak, T., Mathrani, A., & Suriadi, S. (2017). On predicting academic performance with process mining in learning analytics. *Journal of Research in Innovative Teaching & Learning*, *10*(2), 160–176. https://doi.org/10.1108/JRIT-09-2017-0022

Vahdat, M. (2017). Learning analytics and educational data mining for inquiry-based learning. PhD Thesis, Eindhoven University of Technology, Netherlands https://pure.tue.nl/ws/files/59844819/20170403_Vahdat.pdf

Van den Beemt, A., Buijs, J., & Van der Aalst, W. (2018). Analysing structured learning behaviour in massive open online courses (MOOCs): an approach based on process mining and clustering. *The International Review of Research in Open and Distributed Learning*, 19(5), 35–60. https://doi.org/10.19173/irrodl.v19i5.3748

Vavoula, G., & Sharples, M. (2002). KLeOS: a personal, mobile, knowledge and learning organisation system, In *Proceedings of IEEE International Workshop on Wireless and Mobile Technologies in Education*, Tokushima, Japan, Aug 30. https://doi.org/10.1109/WMTE.2002.1039239

Viola, P., & Wells, W. M. I. I. I. (1995). Alignment by maximization of mutual information. In *Proceedings of IEEE International Conference on Computer Vision*, MA, USA, Jun 20-23. https://doi.org/10.1109/ICCV.1995.466930

Wang, C., Shannon, D., & Ross, M. (2013). Students' characteristics, self-regulated learning, technology self-efficacy, and course outcomes in online learning. *Distance Education, 34*(3), 302–323. https://doi.org/10.1080/01587919.2013.835779

Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In Wittock M. C. (Eds.), *Handbook of Research on Teaching*, New York, MacMillan company.

Wengelin, Å. (2006). Examining pauses in writing: theories, methods and empirical data. In Sullivan K. P. H., & Lindgren E. (Eds.), C*omputer Key-stroke Logging and Writing: Methods and Applications. 18:* 107–130. Oxford: Elsevier.

Whipp, J. L., & Chiarelli, S. (2004). Self-regulation in a web-based course: a case study. *Educational Technology Research and Development*, 52(4), 5–21. https://link.springer.com/article/10.1007/BF02504714

William G, C. (1952). The chi-square test of goodness of fit. *The Annals of Mathematical Statistics,* 23(3), 315–345. https://doi.org/10.1214/aoms/1177729380

Winne, P. H., & Hadwin, A. F. (1998). *Studying as self-regulated learning. In Metacognition in educational theory and practice*. (277–304). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Wittrock, M. C. (1977). Learning and instruction. Berkeley, CA: McCutchan Publishing Corporation.

Xu, C., & Ding, Y. (2014a). An exploratory study of pauses in computer-assisted EFL writing. *Language, Learning and Technology*, 18(3), 80–96. https://eric.ed.gov/?id=EJ1046527

Xu, C., & Ding, Y. (2014b). An exploratory study of pauses in computer-assisted EFL writing. *Language Learning & Technology*, 18(3), 80–96. http://llt.msu.edu/issues/october2014/xuding.pdf

Yang, S. J. H., Lu, O. H. T., Huang, A. Y. Q., Huang, J. C. H., Ogata, H., & Lin, A. J. Q. (2018). Predicting students' academic performance using multiple linear regression and principal component analysis. *Journal of Information Processing*, 26, 170–176. https://doi.org/10.2197/ipsjjip.26.170

You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *The Internet and Higher Education*, 29, 23–30. https://doi.org/10.1016/j.iheduc.2015.11.003

Yuan, L., & Powell, S. (2013). MOOCs and open education: implications for higher education. *Centre for Educational Technology and Interoperability Standards*, Bolton, England. https://doi.org/10.13140/2.1.5072.8320

Yudelson, M., Hosseini, R., Vihavainen, A., & Brusilovsky, P. (2014). Investigating automated student modeling in a java MOOC. In *Proceedings of the 7th International Conference on Educational Data Mining (EDM'14)*. London, UK, Jul 4-7. http://d-scholarship.pitt.edu/id/eprint/21833

Zhang, M., Hao, J., Deane, P., & Li, C. (2018). Defining personalized writing burst measures of translation using keystroke logs. In *Proceedings of the 11th International Conference on Educational Data Mining* (EDM'18). Buffalo NY, USA, Jul 15-18.

Zhang, M., Hao, J., Li, C., & Deane, P. (2016). Classification of writing patterns using keystroke logs. In van der Ark L. A., Bolt D. M., Wang W. C., Douglas J. A., & Wiberg M., (Eds.), *Quantitative psychology reseach* (299-314). New Yprk, NY: Springer. https://doi.org/10.1007/978-3-319-38759-8_23

Zhang, M., Hao, J., Li, C., & Deane, P. (2018). Defining personalized writing burst measures of translation using keystroke logs. In *Proceedings of the International Conference on Educational Data Mining (EDM'18)*, Buffalo NY, USA, Jul 15-18.

Zhang, Y., Chan, W., & Jaitly, N. (2017). Very deep convolutional networks for end-to-end speech recognition. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*. New Orleans, LA, USA, Mar 5-9. https://doi.org/10.1109/ICASSP.2017.7953077

Zimmerman, B. (2015). Self-regulated learning: theories, measures, and outcomes. *International Encyclopedia of the Social & Behavioral Sciences (2nd ed.),* 541-546. https://doi.org/10.1016/B978-0-08-097086-8.26060-1

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Zurada, J. M. (1992). Introduction to artificial neural systems. St. Paul, MN, USA: West Publishing Corporation.