EXPLORING THE RELATIONS BETWEEN PRACTICE AND TEST SCORES: A LEARNING ANALYTICS STUDY USING PROCESS MINING AND CORRELATION ANALYSIS WITH AN ONLINE PRACTICE SUPPLEMENT

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

GÖKHAN GÜLFİDAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER EDUCATION AND INSTRUCTIONAL TECHNOLOGY

DECEMBER 2023

Approval of the thesis:

EXPLORING THE RELATIONS BETWEEN PRACTICE AND TEST SCORES: A LEARNING ANALYTICS STUDY USING PROCESS MINING AND CORRELATION ANALYSIS WITH AN ONLINE PRACTICE SUPPLEMENT

submitted by **GÖKHAN GÜLFİDAN** in partial fulfillment of the requirements for the degree of **Master of Science** i**n Computer Education and Instructional Technology, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, **Graduate School of Natural and Applied Sciences** _______________

Prof. Dr. Soner Yıldırım
Head of the Department, **Computer Education and Instructional Technology** _______________

Prof. Dr. Soner Yıldırım
Supervisor, **Computer Education and Instructional Technology, METU** _______________

Assoc. Prof. Dr. Erkan Er
Co-Supervisor, **Computer Education and Instructional Technology, METU** _______________

Examining Committee Members:

Prof. Dr. İhsan Tolga Medeni
Management Information Systems, Ankara Yıldırım Beyazıt University _______________

Prof. Dr. Soner Yıldırım
Computer Education and Instructional Technology, METU _______________

Assoc. Prof. Dr. Tarkan Gürbüz
Computer Education and Instructional Technology, METU _______________

Date: 01.12.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name : Gökhan Gülfidan

Signature :

ABSTRACT

EXPLORING THE RELATIONS BETWEEN PRACTICE AND TEST SCORES: A LEARNING ANALYTICS STUDY USING PROCESS MINING AND CORRELATION ANALYSIS WITH AN ONLINE PRACTICE SUPPLEMENT

Gülfidan, Gökhan
Master of Science, Computer Education and Instructional Technology
Supervisor : Prof. Dr. Soner Yıldırım
Co-Supervisor: Assoc. Prof. Dr. Erkan Er

December 2023, 62 pages

This study was conducted within the scope of Learning Analytics and Process Mining in a learning portal implemented in different private schools in Turkey. The relationship between the first semester exam performances of 364 students studying at the 5th, 6th, 7th, and 8th grade levels and the performance of practice is examined. The variables predicting the exam performance are estimated by examining the process data on the system. In the analysis of the study, Fluxicon Disco software was used to examine the process data. The Pearson Correlation method was used to examine the relationship between the exam performances and practice. According to the research results, there is a positive correlation between student exam performance and practice. In addition, in the process mining analysis, practice positively predicted student performance.

Keywords: Learning Analytics, Process Mining, Correlation Analysis, Data Mining

ÖZ

PRATİK YAPMA VE TEST PUANLARI ARASINDAKİ İLİŞKİLERİN KEŞFEDİLMESİ: ÇEVRİMİÇİ DESTEK MATERYALİYLE SÜREÇ MADENCİLİĞİ VE KORELASYON ANALİZİNİ KULLANAN BİR ÖĞRENME ANALİTİĞİ ÇALIŞMASI

Gülfidan, Gökhan
Yüksek Lisans, Bilgisayar ve Öğretim Teknolojileri Eğitimi
Tez Yöneticisi: Prof. Dr. Soner Yıldırım
Ortak Tez Yöneticisi: Assoc. Prof. Dr. Erkan Er

Aralık 2023, 62 sayfa

Bu çalışma, Türkiye'deki farklı özel okullarda uygulanan bir öğrenme portalında Öğrenme Analitiği ve Süreç Madenciliği kapsamında gerçekleştirilmiştir. 5, 6, 7 ve 8. sınıf düzeyinde öğrenim gören 364 öğrencinin birinci dönem sınav performansları ile pratik yapması arasındaki ilişki incelenmiştir. Sınav performansını yordayan değişkenler sistemdeki süreç verileri incelenerek tahmin edilmiştir. Çalışmanın analizinde süreç verilerinin incelenmesi amacıyla Fluxicon Disco yazılımı kullanılmıştır. Sınav performansları ile uygulama arasındaki ilişkiyi incelemek için Pearson Korelasyon yöntemi kullanıldı. Araştırma sonuçlarına göre öğrencinin sınav performansı ile pratik yapması arasında pozitif bir ilişki vardır. Ayrıca süreç madenciliği analizinde pratik yapma öğrenci performansını olumlu yönde yordamıştır.

Anahtar Kelimeler: Öğrenme Analitiği, Süreç Madenciliği, Korelasyon Analizi, Veri Madenciliği

To my love and family…

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor Prof. Dr. Soner Yıldırım and co-supervisor Assoc. Prof. Dr. Erkan Er for their guidance, advice, criticism, encouragement, and insight throughout the research.

I would also like to thank Prof. Dr. İhsan Tolga Medeni and Assoc. Prof. Dr. Tarkan Gürbüz for their valuable contribution to my study with their evaluations as jury members of my thesis.

In loving memory of my dear mother and dedicated teacher, Kerime Günay, whose strength and wisdom continue to guide me. This thesis is dedicated to her beloved memory, and I hope she is proud of the academic journey I have navigated since her untimely departure during my youth.

I extend heartfelt gratitude to my aunts, Sevgi Gürtuna and Serpil Sarı, and uncle, Kamil Günay, for their unwavering support and love throughout my academic journey. Your guidance and encouragement have been invaluable, and I am deeply thankful for the role each of you played in shaping my life.

My wife, Ayşegül Durmuş Gülfidan, who is one of the biggest supporters of my life, was the most supportive and motivating one during my study. This achievement could not have been possible without her. I cannot express enough gratitude to her for her unwavering patience, understanding, love, and belief in me.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

AI: Artificial Intelligence

DM: Data Mining

EDM: Educational Data Mining

KDD: Knowledge Discovery in Databases

LA: Learning Analytics

LMS: Learning Management System

MOOC: Massive Open Online Course

PLE: Personal Learning Environments

PM: Process Mining

CHAPTER 1

INTRODUCTION

Today, technological advancements have led people to leave copious amounts of data in every field by connecting to the internet via their mobile devices and computers. The number of workspaces with an advantage in using large data sets is high. These workspaces gather data from diverse sources to improve customer service, including social media, health records, purchases, and cell signals. Ensuring the analysis is done correctly for large amounts of data is essential. The right kind of analysis must be carried out to meet the needs of these datasets. At this point, it is necessary to process and understand them with the help of Data Mining (DM) to draw meaningful conclusions in large data sets. DM can be defined as a field that provides meaning to extensive data sets with traditional data analysis processes and artificial intelligence algorithms (Olson, 2007). DM is used in different sectors and business areas, providing significant benefits in the processing and analyzing the data obtained, such as marketing, investment, manufacturing, data cleaning, and financial sectors (Fayyad et al., 1996).

Apart from the mentioned areas, DM has a more important place in education as online learning platforms and interactivity increase and integrate into education (Bartschat et al., 2019). In online learning platforms, DM is essential for understanding students' learning habits and analyzing the information hidden in the data (Mohamad & Tasir, 2013). With the data collected on these platforms, a student-oriented education strategy can be followed by identifying the difficulties experienced by students (Aldowah et al., 2019). Educational Data Mining (EDM), a type of DM, focuses on data analysis and supporting learning in educational settings, such as schools and universities (Romero & Ventura, 2013). EDM is defined as the analysis of large-scale data in education, finding patterns within the framework of these data, research, development, and using these methods to develop online education platforms (Romero et al., 2010).

EDM, the research area that emerged with the idea of using DM methods in education, has been separated as a different area for knowledge discovery and decision-making with training datasets (Peña-Ayala, 2014). A large number of datasets increasing with EDM aims to provide information to instructors and provide better learning support by examining students' behaviors and characteristics (Baker & Inventado, 2014). Learning characteristics, academic achievement, interactions, and individual learning preferences can be collected from online learning platforms and used for effective learning (Aldowah et al., 2019). The field of EDM has influenced many interdisciplinary fields, such as computer science and learning analytics (Siemens & Baker, 2012).

Today, with the rise of online education, educators are worried about whether students genuinely comprehend the content they teach and whether the lesson objectives are being met successfully. With online learning environments, every movement, intervention, and trace left behind by learners in different courses has facilitated data collection to improve the learning environment (Peregrina et al., 2014). However, the critical point here is to perform the logical analysis of the collected data correctly and make the right decisions to improve learning. At this point, Learning Analytics (LA) plays a vital role in analyzing the data and making predictions in line with the analysis.

With all these concerns in mind, the LA area has evolved. LA is the process of gathering, analyzing, and reporting data about learners and their context to understand better and improve learning environments and processes (Long & Siemens, 2011). The life cycle of the LA field can be defined as obtaining, storing, and analyzing the data that learners leave on online learning platforms, making predictions and decisions, and taking action as a result of the analysis (Khalil & Ebner, 2015). The LA field's importance in helping improve learning is evident when instructors identify the weak points of their students and course materials and give feedback (Gašević et al., 2015).

Moreover, LA has many advantages in the field of education for both instructors and teachers. LA offers educators new methods to understand students and provide more effective learning (Clow, 2013). With the help of the analysis obtained from LA, instructors can give personalized feedback according to the learners, predict the students with low success in the course, and intervene and improve learning (Siemens, 2013). In addition, with the help of the LA results, the motivation of the students to the lessons to increase their academic performance also came to the fore as an essential factor for learning (Schumacher & Ifenthaler, 2018). In addition, with the help of analysis results, teaching quality and learning practices can be improved, and more successful results can be obtained for students (Gašević et al., 2015). Furthermore, administrators can decide more easily on subjects such as learning theories and assessment materials for future courses in line with the analysis results and the actual forecasts (Simon, 2017).

Many studies have been carried out on the subject of LA. Most relevant studies on LA attempt to predict learners' academic performance through online learning platforms (Abdous et al., 2012; He, 2013; Romero-Zaldivar et al., 2012; Giesbers et al., 2013; Dietz-Uhler & Hurn, 2013), giving feedback when necessary (Ifenthaler, 2017; Arnold & Pistilli, 2012; Papamitsiou & Economides, 2014; Ali et al., 2012; Clow & Makriyiannis, 2011; Macfadyen & Dawson, 2010; Leong et al., 2012), on evaluation studies (Terzis & Economides, 2011; Kizilcec et al., 2013; Barla et al., 2010; Chen & Chen, 2009; Wilson et al., 2011; Worsley & Blikstein, 2013), on students likely to drop out of course (Dejaeger et al., 2012; Dekker et al., 2009; Giesbers et al., 2013; Gurus et al., 2010; Kizilcec et al., 2013; Freitas et al., 2015), to provide personalized learning (Avella et al., 2016; Siemens, 2013; Romero et al., 2009), study student motivation (Clow & Makriyiannis, 2011; Giesbers et al., 2013) and examine self-awareness using learning analytics (Papamitsiou & Economides, 2014; Fournier et al., 2011; Santos & Boticario, 2012). All studies show that considering the steps and recommendations in the research framework, LA is an essential tool for analyzing data obtained in education and learning platforms and predicting results.

In addition, many people are unaware that every click on the internet leaves behind a trace. This data is known as "log data" and is created by the actions taken and traces left behind (van der Aalst, 2011). The amount of event data generated constantly increases in all industries, economies, organizations, and households, according to Manyika et al. (2011). Examining users' online behavior is crucial, and collecting data plays a significant role. "Process mining" is one area that utilizes data science to analyze user behavior.

Process mining uses event logs in modern systems to gain insights, track progress, and optimize processes. This approach enables the discovery, monitoring, and optimization of essential processes (van der Aalst, 2011). Process mining analyze student activities and learning materials to help educators understand student progress and identify areas where assistance may be needed (Uzir et al., 2020). Several process mining tools can be used to visualize student activities (e.g., ProM and Fluxicon Disco) in the literature (dos Santos Garcia et al., 2019; Bogarin et al., 2018). In this way, students' transition and frequency between activities can be examined, and more effective learning can be achieved.

There are various methods to make learning permanent. Rather than presenting learning materials continuously, spaced practice, distributed practice, or spaced repetition learning strategies involve distributing and repeating the materials over time. These strategies enhance cognitive learning and consolidation (Cepeda et al., 2006). On the other hand, repeatedly retrieving information from memory to enhance learning and retention is referred to as retrieval practice (Roediger & Butler, 2011). Retrieval practice employs techniques such as practice tests, flashcards, and open-ended writing to help recall information. Long-term learning benefits are more durable when information is retrieved repeatedly with intervals of time in between compared to simply reviewing the same information over the same intervals (Carpenter et al., 2022).

This study aims to determine whether there is a relationship between practice and exam performance in using online learning platforms among secondary school students. In addition, a process analysis of the use of online learning platforms was carried out with the obtained data sets. This analysis was carried out by dividing the students into two groups according to their exam performance and according to lessons. The process maps obtained from the applied analyses were created, and the factors predicting the success were examined.

1.1 Purpose of the Study

Nowadays, e-learning environments are gaining value with the widespread use of online learning. These platforms, also used in schools, transfer learning between students and educators. With the spread of online learning platforms, data such as students' interaction with course materials and test results can be followed by educators. Today, large amounts of data can be collected with the widespread use of online platforms. In education, learning analytics is essential in making the necessary inferences from these data and making learning more effective. Analyzing the obtained data correctly and making logical inferences is crucial in this context. This study aims to determine whether there is a correlation between students' exam performances and their practice results on the learning portal in Turkish, Mathematics, Science, and Social Science courses. In addition, with the data obtained, the students were divided into two groups according to their exam performance. The factors predicting the success were examined by using process mining techniques to draw the process maps between the learning portal activities.

1.2 Problem Statement

This study examined the relations between students’ practice and exam scores within an online learning portal at the secondary school level. Using process mining and correlation analysis provided a comprehensive understanding of young learners' behavior in online learning environments.

1.3 Research Questions

- Is there a correlation between students' exam scores and practice scores obtained in the learning portal in Turkish, Mathematics, Science, and Social Science courses respectively?
- How do the frequency and sequence of transitions between lecture, practice and exam activities relate to level of student achievement?

1.4 Significance of the Study

In recent years, the importance of data mining is evident with the increase in the use of online platforms and the collection of data for more efficient learning. In this study, a process analysis of the students was carried out with the data obtained from the learning portal. The analysis examined students' behaviors on the learning portal according to the success groups. Students' learning habits were determined, and inferences were made for better learning on online learning platforms. In addition, in light of the results obtained, educators and administrators have valuable information on what needs to be done for better learning on online platforms. Moreover, the relationship between practice and the exam was examined using the learning portal. With the results obtained from the study, those working in the field of education gain important information about the use of practice in improving learning performance. Additionally, there are few studies in the literature on young learners' behavior in online learning environments, highlighting the importance of this study.

1.5 Limitations of the Study

- The study is limited to 364 secondary school students using the learning portal in Turkey.
- The students were chosen by convenience sampling.
- The study is limited to students' success in Turkish, Mathematics, Science, and Social Science courses.
- It was assumed that the students in the study used the learning portal as desired.
- The research data was collected during the Covid-19 pandemic.
- Students' practice and exam scores are collected in an uncontrolled learning portal.

CHAPTER 2

LITERATURE REVIEW

In this section, the processing of large-scale data, the areas where it is used, and the Data Mining concept were discussed in this context. In addition, the intersections of the field of education and large-scale data and their benefits were mentioned. Moreover, the current importance and usage of the Learning Analytics and Process Mining fields were explained. Finally, research related to the study was examined, spaced and retrieval practice was discussed in detail.

2.1 Data Mining

Today, the increase in the amount of data and the interpretation of this data have increased the interest in Data Mining (DM). Institutions and individuals who had to deal with large data sets faced the challenge of reaching conclusions with a manual analysis. DM is a field that enables to make inferences from large data sets (Kriegel et al., 2007). DM, also known as Knowledge Discovery in Databases (KDD), can be defined as the process of discovering usable and meaningful data from data using specific algorithms and patterns (Fayyad et al., 1996). DM has proven its worth in almost every academic discipline in which it has been used. These fields include banking, health systems, financial analysis, insurance, and education (Olson, 2007).

The steps to extract the necessary information from the data can be defined as data selection, cleaning and preprocessing, data transformation, searching for patterns in the data, evaluating the data, and extracting the information (D'oca & Hong, 2015). The main objectives of the DM field are divided into prediction and description, which form the basis of the methods used in this field (Fayyad et al., 1996). The primary methods used in data mining are characterization, classification, evaluation, matching patterns, clustering, and visualizing data (Liao et al., 2012).

DM is essential in education as it is used in many business areas. In order to process the large data sets collected in the field of education, make them meaningful, and contribute to this field, the field of Educational Data Mining (EDM) has developed over time, and a new research area has been formed.

2.2 Educational Data Mining

The increase in the number of online environments used in education, such as the Learning Management System (LMS) and the Massive Open Online Course (MOOC), and the increase in the amount of data collected can be used to solve different educational problems (Romero & Ventura, 2013). When students complete the course requirements in online learning environments, they leave statistically essential data such as time spent and course grades. Every educator wants their students' learning to become more efficient. For this reason, the data left in the online learning environments where the course takes place can be used with the help of EDM to improve course quality and for better learning in the future. EDM is significant in examining students' learning behaviors, especially in online learning environments (Mohamad & Tasir, 2013). It can also be explained as an application of Data Mining techniques used to find answers to educational questions with data sets collected in educational environments (Romero & Ventura, 2013). EDM is a field that develops on finding patterns and making predictions by examining learners' characteristic behaviors, achievements, course content, and educational concepts such as assessment materials from the perspectives of design models, methods, and algorithms (Peña-Ayala, 2014).

EDM enables learners to create self-awareness learning environments by analyzing the obtained data. These environments allow learners to design patterns according to their needs and provide feedback for different learning styles and groups (Papamitsiou & Economides, 2014).

Although EDM uses data mining, statistics, and visualization techniques, it needs more customized research as a separate and advanced field (Romero & Ventura, 2007). EDM stages can be listed as preparation of the environment in the learning process, preliminary preparation of the process, data collection and reporting, and estimation of the results (Papamitsiou & Economides, 2014). With the data collection techniques in online learning environments, text mining, pattern mining, classification, and data visualization can be performed (Aldowah et al., 2019).

2.3 Learning Analytics

Interest in managing large data sets has increased with the rise of online platforms in education. (Clow, 2013). Technology-driven learning environments provide an excellent opportunity to analyze students' behavior and reshape it accordingly (Schumacher & Ifenthaler, 2018). However, unlike other sectors, there have been some difficulties in analyzing data obtained in education and developing learning environments according to these analyses (Siemens, 2013). Although high amounts of data were obtained from online learning systems such as LMS and Personal Learning Environments (PLE), data for learning and teacher development was limited (Greller & Drachsler, 2012).

The data collected during the interactions in education and information technologies has recently attracted attention within the framework of the Learning Analytics (LA) research field in terms of understanding subjects such as the learning process (Gašević et al., 2015). According to Long and Siemens (2011), 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 "(p. 34). With the rich data sources obtained, the goal of understanding the learning methods of the learners and performing better teaching has become possible in the LA field (Gursoy et al., 2016).

LA is close to several related research areas, such as Academic Analytics and Educational Data Mining (Viberg et al., 2018). LA aims to develop a better learning, teaching, and learning process by creating and analyzing a pattern of the learning process from the data obtained by the actions of students, teachers, and administrators in online education environments (Peregrina et al., 2014).

Khalil and Ebner (2015) describe the life cycle of the LA approach as four main stages:

- Production of Data
- Data Store
- Analysis
- Act

LA life cycle starts in online learning environments, and data are obtained from different environments such as LMS and MOOC. In these environments, learners leave traceable data behind, which indicates that learners are data sources. Obtained data sets create patterns and are analyzed with analysis methods and techniques. As a result of the analysis, some forecasting, recommendations, and interventions are determined to improve learning (Khalil & Ebner, 2015).

LA has several advantages in education for learners, teachers, and administrators. In terms of educators, it provides the opportunity to monitor the real-time performance of the learners and intervene with learning activities on time. At the same time, for the students, it ensures that they are motivated by getting information about their situation and helping each other. It enables administrators to compete with other institutions and not be economically challenged (Long & Siemens, 2011). With the help of the data obtained with LA, learner-oriented education is facilitated during the process, and personalization increases (Greller & Drachsler, 2012). In this way, students are evaluated by examining the results of the LA analysis, and the education given is improved over time.

Educators have professional responsibilities to increase students' learning levels, and the LA research field includes powerful tools and methods to achieve this (Clow, 2013). Educators can make their lessons more productive and provide better learning with LA. In addition, by analyzing student behavior, educators can predict their success and grades, monitor course progress, and track time spent in the online learning environment with timestamps. (AbuKhousa & Atif, 2016; Cambruzzi et al., 2015; Harrison et al., 2015; Menchaca et al., 2015; Vahdat et al., 2015; Wise, 2014). In light of the data obtained, educators can improve learning by identifying students' weak points according to the activities performed and giving feedback on where they are stuck (Gašević et al., 2015). In terms of administrators, a result of LA analysis can allow educators to change the course in this direction by making the necessary decisions so that fewer students will fail in future courses (Simon, 2017).

On the other hand, there are some critical issues while collecting data and estimating the analysis results, such as information, location confidentiality, and ethical and time problems (Slade & Prinsloo, 2013). In addition, institutional decisions should be made within the framework of ethical principles, and institutions should act and guide following the general decision (West et al., 2016). On the other hand, according to the analyzed data, the support and intervention of the learners should be done on time and according to specific needs (Siemens, 2013).

2.4 Related Studies about Learning Analytics

In the literature, when relevant review articles are examined (Papamitsiou & Economides, 2014; Banihashem et al., 2018) generally uses quantitative data such as age, gender, previous academic achievements, family status. Some of the research topics of the articles are divided into paragraphs and explained in detail.

Firstly, learning analytics demonstrate efficacy in providing timely feedback and effective interventions, predicting student achievements, analyzing learning habits, and fostering a supportive and personalized educational environment. To improve the accuracy of predicting students' performance, removing answers with extremely high or low probabilities of being correct is recommended. It can be achieved using forecasting techniques considering sequential information (Ouyang et al., 2023; Ifenthaler, 2017; Arnold & Pistilli, 2012; Papamitsiou & Economides, 2014; Ali et al., 2012; Clow & Makriyiannis, 2011; Macfadyen & Dawson, 2010; Leong et al., 2012; Terzis & Economides, 2011; Kizilcec et al., 2013; Barla et al., 2010; Chen & Chen, 2009; Wilson et al., 2011; Worsley & Blikstein, 2013; Shih et al., 2008; Abdous et al., 2012; He, 2013; Romero-Zaldivar et al., 2012; Giesbers et al., 2013; Dietz-Uhler & Hurn, 2013).

Several studies have highlighted the significance of learning analytics in improving students' self-awareness, fostering collaborative learning, predicting dropouts, and enabling personalized learning experiences. The findings indicated that learning analytics in education could yield significant advantages, including boosted student engagement, enhanced learning achievements, and the identification of students who may be at risk (Yang & Ogata, 2023; Papamitsiou & Economides, 2014; Fournier et al., 2011; Santos & Boticario, 2012; Dejaeger et al., 2012; Dekker et al., 2009; Giesbers et al., 2013; Guruler et al., 2010; Kizilcec et al., 2013; Freitas et al., 2015; Avella et al., 2016; Siemens, 2013; Romero et al., 2009).

It has been demonstrated through multiple studies that learning analytics can inform the development of curriculum, improve the design of learning experiences, and offer cost-effective educational solutions based on data-driven insights. By providing data-driven insights to support decision-making, learning analytics can positively impact curriculum development (Avella et al., 2016; Reyes, 2015; Freitas et al., 2015; Lockyer et al., 2013; Dekker et al.,2009; Dinu et al., 2017).

Lastly, numerous studies have shown that learning analytics can increase student motivation and provide valuable information on achievement rates and exam performances, creating a more engaging and informed learning environment (Ifenthaler et al., 2023; Giesbers et al., 2013 Clow & Makriyiannis, 2011; Lykourentzou et al., 2009).

As a result of the related studies in the literature, some conclusions can be reached by making inferences. One of them, LA research, shows that online education platforms can collect and analyze data about learners and educators within the course framework. With the analysis results obtained, the students' academic performance can be observed, and interventions can be made to the learners when necessary, such as feedback. Therefore, this situation can be prevented by recognizing students with deficient performance who are likely to drop out of the course. In addition, with the developed online learning portals, the self-awareness feature of the students can be improved, and their motivation for the lesson can be increased.

Moreover, educators can monitor and analyze students' performance throughout the semester with these portals. In this way, the learners' performance increase can be observed, and personalized learning can be provided. However, it should be noted that these analyses must be done carefully and diligently in order to make the right interventions for learners.

It is crucial to consider cognitive strategies that can enhance learning outcomes to improve the effectiveness of interventions. As such, the next section examines the existing literature on spaced and retrieval practices - two techniques shown to improve information retention and retrieval significantly.

2.5 Cognitive Strategies for Learning

There are six main strategies for effective learning. Firstly, elaboration integrates new information with pre-existing knowledge by expressing it in one's own words and establishing connections with the existing knowledge, facilitating comprehension and retention (Brown et al., 2014). Secondly, interleaving can be employed, incorporating varied problems or materials into study sessions to enhance learning and retention (Rohrer D., 2012). Thirdly, using concrete examples to exemplify abstract concepts is a commonly employed technique that aids in enhancing comprehension (Rawson et al.,2015). Fourthly, dual coding involves using visual and verbal methods to improve memory retention and comprehension of information (Mayer & Anderson, 1992). The last two strategies, spaced and retrieval practice, are the most frequently used in online learning environments. These strategies are explained in detail in the following sections.

2.5.1 Spaced Practice

In the field of education, both students and teachers have sought methods for effective learning. One of the basic requirements of learning is to practice. The timing of the practice is crucial and can significantly impact the outcome (Kang, 2016). The learning strategy known as spaced practice, distributed practice, or spaced repetition involves distributing and repeating learning materials over time rather than presenting them continuously, consolidated, and intended to enhance cognitive learning (Cepeda et al., 2006). This approach differs from massed practice, which presents information in a concentrated and uninterrupted manner (Toppino & Cohen, 2010). Revisiting important concepts over time can significantly improve student retention (Cepeda et al., 2006).

In education, it is necessary to implement spaced practice with appropriate methods to be applied effectively in schools and online learning platforms (Petersen-Brown et al., 2019). There are several ways to space out the exposure to content for better retention. One effective method is to review the essential concepts from a lecture at the end of each class and then review them again at the beginning of the next class (Dunlosky & Rawson, 2015). It has been found that when information is retrieved repeatedly with intervals of time in between, there are long-term benefits to learning that are more durable than simply reviewing the information over the same intervals. Additionally, retrieving information over longer spacing intervals is more effective than doing so after shorter intervals (Carpenter et al., 2022).

2.5.2 Retrieval Practice

Learning can be defined as acquiring new information or skills and adapting existing knowledge. Enhancing the learning process has always been crucial for learners, instructors, policymakers, and researchers (Yang et al., 2021). In the field of education, different strategies have been developed to make learning more permanent. Some effective strategies are summarization, keyword mnemonic, rereading, note-taking, and practice testing (Dunlosky et al., 2013). Nowadays, testing is generally used to measure students' performance in the measurement and evaluation part. However, studies show retrieval practice is more effective than other strategies to facilitate long-term learning and mastery of new information (Roediger & Karpicke, 2006; Yang et al., 2017). The term retrieval practice is repeatedly retrieving information from memory to improve learning and retention (Roediger & Butler, 2011). Practice tests, flashcards, and open-ended writing are all different methods of retrieval practice used to recall information (Carpenter et al., 2022). Testing is an effective way to improve memory retention. It enhances retention more than restudy, even when done without feedback. This phenomenon is called the testing effect (Roediger & Karpicke, 2006).

Learning designers can use retrieval practice to enhance online courses' interactivity and involvement (Millet et al., 2021). If teachers identify critical knowledge and skills their students should learn, they can emphasize and test them repeatedly to ensure mastery. (Roediger & Karpicke, 2006). With the spread of online learning platforms, a more interactive learning environment has emerged for learners. Students may not effectively use retrieval practice during the study if they hold inaccurate metacognitive beliefs, which may lead them to avoid repeated self-testing of material. Research suggests that people tend to restudy material rather than repeatedly self-test, and even when they do self-test, they tend to drop the material after being recalled once. (Karpicke, 2009; Kornell & Bjork, 2008). Although these environments benefit from technology, a design in the light of learning strategies is essential for efficient learning. It is aimed at making learning more efficient with different interactive practice techniques. The learner must also decide when to end or adjust their study. In general, it is not only necessary to teach individuals different methods of acquiring information, but it is also crucial to instruct them on how to assess the efficacy of these methods. (King et al., 1980).

Three instructional conditions need to be met to maximize the positive effect of retrieval practice on learning outcomes. These conditions are spacing, feedback, and desirable difficulties. Scholars recommend that retrieval practice be spaced equally with longer intervals for maximum effectiveness. Feedback is essential in recognition tests, such as multiple-choice or true/false questions, because incorrect information can be exposed. By increasing the difficulty of the practice, learners' initial retrieval and storage strength can be lowered, leading to better long-term retention. (Millet et al., 2021)

In online learning platforms, self-regulation is expected from students. For this reason, to improve learning on these platforms, some instructions are applied that students should follow. Strategies developed for students to repeat study materials have been observed to affect learning (Ariel & Karpicke, 2018). As crucial as ensuring that students' progress in a correct sequence with the necessary instructions, the activities applied have an essential place in learning. Structuring or scheduling learning activities over time is known as spacing, while retrieval practice refers to a learning activity that can be included in a more comprehensive and organized plan for effective learning (Carpenter et al., 2022). It is seen that learning is more efficient when study and testing, which are learning stages, are applied consecutively in a sequence (Roedieger & Butler, 2011). It has been observed in recent experiments that providing explicit directions regarding the benefits of repeated retrieval practice and how to employ this technique can modify and enhance students' self-regulated learning behavior. A straightforward instructional intervention can enhance learning results and encourage the use of effective self-regulated learning techniques that endure over a prolonged period (Ariel & Karpicke, 2018). For this reason, in online learning platforms, it can be ensured that students follow a specific process to perform more efficient learning and implement practice activities. The paths students follow on these platforms can be examined using process mining.

2.6 Process Mining

Recent technological advancements have enabled individuals to carry out their daily tasks using online platforms. Even though most people are unaware of it, every click on the Internet leaves a trace behind. The data created due to the actions taken and the traces left behind is called "log data." This event data constantly increases exponentially across all industries, economies, organizations, and households (Manyika et al., 2011). Collecting this data is extremely important to examine users' online behavior. One of the fields that analyzes user behavior with data science is "process mining."

Process mining involves utilizing event logs that are easily accessible in modern systems to gain insights, track progress, and enhance actual processes. This approach enables the discovery, monitoring, and optimization of essential processes (van der Aalst, 2011). It can discover more information about events by referring to the event logs. Process mining techniques frequently use this data, which includes details like the identity of the resource (whether it is a person or device) that performed or triggered the activity, the event's timestamp, and any recorded data elements associated with the event (van der Aalst, 2012). Teachers can utilize the learning behaviors of their students in order to gain insight into and facilitate the learning process. However, analyzing various learning pathways can prove challenging for instructors to undertake manually in widespread education (Moreno et al., 2021).

In event log structure, there are three significant components: Case ID specifies the instance or case to which an event or action belongs; activity, which captures the action performed by the event; and timestamp, which represents the time at which the event occurred (Moreno et al., 2021). These elements are essential parts of process analysis. With this data, analyzes can be carried out in the field of education.

Process mining can be used in education to analyze how students learn. It involves looking at students' activities (e.g., viewing, attempting, and updating) and learning materials (e.g., assignments, quizzes, resources, and forums). By combining this information, educators can better understand how students are progressing and what they might need help with (Uzir et al., 2020). After converting the data into the required format of an "event log," process mining techniques and tools can create a model without relying on any pre-existing information (van der Aalst, 2011).

Process mining models have gained popularity in learning analytics for their ability to describe and analyze learning processes. This approach has proven helpful for understanding the dynamics of how students learn and can provide valuable insights for educational practitioners (Bogarin et al., 2018). In online learning, the teacher typically gives students guidelines to follow to gain specific competencies and skills. However, each student's learning style, pace, and preferred media content may differ, which means that the path they take to acquire these skills may not be the same as the one initially outlined in the online learning platforms. Combining process mining and learning analytics techniques to analyze students' behavior patterns could provide numerous advantages (Moreno et al., 2021).

Data and process mining algorithms can predict which online learning students may fail to meet course requirements (Umer et al., 2017). Gasevic et al. (2017) examined the correlation between students' reported learning strategies and those detected from trace data and their impact on academic performance. The use of process mining techniques in software development education can provide insight into the behavior of both students and developers during coding sessions, ultimately leading to enhanced coding processes (Ardimento et al., 2019).

Over the past few years, process mining techniques have increasingly been utilized in education to analyze, enhance, and visualize learning processes. Numerous studies have been conducted using these techniques (dos Santos Garcia et al., 2019; Bogarin et al., 2018).

Firstly, several studies have used process mining to identify bottlenecks in the self-regulated learning curriculum, revealing insights into educational process efficiency and potential areas for improvement. Comparisons between the paths taken by successful and less successful students have been conducted in certain studies, highlighting differences that could be utilized to enhance the curriculum (Hobeck et al., 2022; Juhanak et al., 2019; Intayoad et al.,2018; Maldonado-Mahauad et al., 2018; Qiao & Hu, 2018; Rodrigez et al., 2018; Cameranesi et al., 2017; Mukala et al., 2017; Neyem et al., 2017; Alvarez et al., 2016; Aisa et al., 2015; Beheshitha et al., 2015; Trcka & Pechenizkiy, 2009; Priyambada et al., 2017; Azzini et al., 2016; Trcka et al., 2010; Wang & Zaiane, 2015).

Numerous publications have investigated the connection between students' learning patterns and academic performance using process mining techniques. These studies have provided valuable insights into the relationship between learning behaviors and success. The studies also have highlighted the potential of educational process mining in understanding and enhancing students' learning patterns and performance (Sedrakyan et al., 2016; Vidal et al., 2016; Mukala et al., 2015; Sedrakyan et al., 2014).

Several studies have utilized process mining to examine students' learning behavior in specific educational activities. These studies aimed to improve understanding and optimize instructional strategies based on observed behavior patterns (Bala et al., 2022; Maita et al., 2017; Sobocinski et al., 2017; Wang et al., 2017; Castillo, 2016; Doleck et al., 2016).

Numerous research have utilized process mining to cluster and map students' profiles, which has led to significant improvements in the educational process. These studies have aimed to create more personalized and effective educational experiences by leveraging data-driven insights. Using clustering and mapping techniques is particularly effective in mining educational processes (Ariouat et al., 2016; Cairns et al., 2015; Bogarin et al., 2014).

Lastly, some studies have delved into process mining to identify learning strategies in educational processes. These studies have also focused on time management. The findings of these studies have contributed to a better understanding of how students interact with learning materials and manage their time. As a result, they offer valuable insights for instructional design and support (Matcha et al., 2019 and time management: Uzir et al., 2020).

Analyzing process mining data can contain numerous challenges. These challenges include refining event data and dealing with complex event logs. It is essential to ensure representative benchmarks and balance quality criteria. Combining process mining with other analyses can be difficult, and usability should be improved for non-experts (van der Aalst, 2012).

In recent years, multiple tools have been developed to aid process mining. When examining education, the ProM framework and Fluxicon Disco tools are commonly cited in publications (dos Santos Garcia et al., 2019; Bogarin et al., 2018). With the process mining tools, models are created from the data obtained. With these models, students' transitions between activities and the frequencies of these transitions are visualized. In this way, a process model is created, and necessary investigations can be made.

In summary, in this literature review, the concept of data mining is first reviewed, and how data is obtained from large-scale organizations is examined. In education, the traces left behind by students on the online learning platforms and how they were analyzed were also mentioned. Secondly, the concept and development of learning analytics, one of the main elements of the study, was mentioned. LA approach consists of data acquisition, storage, analysis, and action parts. The studies' topics and results in this field's literature are also mentioned. Some of the studies in this field are on examining students' behavior, improving their academic performance, giving feedback, and how to make necessary interventions. Then, the concept of process mining was investigated in the context of education. In online learning platforms, students' activities can be observed throughout the learning process using log data and timestamps. Curriculum, learning patterns and performance, student habits, and time management are some concepts discussed in the studies carried out in this field. Finally, spaced and retrieval practice studies in education have examined how learning can become more effective and permanent.

Only a few studies in the literature analyze the path students take in online learning platforms. To fill this gap, in this study, students were grouped according to their exam performance success, and the paths used by these groups and the transition processes between activities were analyzed. The relationship between retrieval practice and students' exam performances was also examined.

CHAPTER 3

METHOD

This study collected data on students' performance scores and interaction in exams, practice, and lecture sections with the learning portal utilized by a private school in the first semester of the 2021-2022 academic year. The learning portal is a software that enables students to attend courses, view the course content, assess performance by practice and exam, and, at the same time, instructors can track and report learners' information. The learning portal used in this study is the online practice supplement used by private schools and in addition to the Ministry of National Education curriculum. Based on these data, procedures were carried out to analyze the paths followed by students between activities (such as lecture, practice, and exam) using learning portal in Mathematics, Turkish, Science, and Social Science courses and to determine the correlation between practice scores and exam scores.

3.1 Data Collection Environment

The K12 level learning portal is an optional product used in private schools in Turkey and sold by a private company. Schools use this platform to support students' courses and take practice and exams for some courses through the system. Permission was obtained from the company to use secondary school students' data for a certain period through the system.

The learning portal includes different types of educational content within the scope of Turkish, Mathematics, Science, and Social Science courses.

The learning portal also includes many test questions in different subjects of the specified courses. When students solve these tests, true, false, and blank questions are recorded. Teachers, students, and parents can monitor students' success in each course. The timestamps when students started and finished the activities were also recorded. In this way, the process data the students followed from the moment they entered the system were collected.

The research was carried out on the exams applied in the study group's first semester of the 2021-2022 academic year and the practice activities performed by the students before the exam. The research was conducted with the data obtained by an anonymized system. The exams were applied to measure the success of 5th, 6th, 7th, and 8th grades in different subjects in Mathematics, Turkish, Science and Social Science lessons. On the other hand, practices are designed to help students prepare and reinforce their knowledge before taking these exams. The practice activities include interactive exercises such as drag and drop, fill in the blanks, and matching. This study divided secondary school students into two groups based on their average exam scores in each course. The data collected was analyzed using process mining techniques to examine the paths taken by the groups between activities. In addition, the correlation between the practice and exam score's part was also examined according to the courses.

3.2 Data Source

The study's data was obtained from the online learning portal.The study consists of 364 students in 5th, 6th, 7th, and 8th grade from different private schools in Turkey using the learning portal.

Table 3.1 *Number of students by class levels*

Level	Total Students	Percentage
5th Grade	137	37.64%
6th Grade	86	23.62%
7th Grade	65	17.86%
8th Grade	76	20.88%

3.3 Data Collection Procedures

Exam performance, practice performance, and system usage data of students using learning portal were obtained from the system's database. The necessary SQL queries were created, and the raw data was obtained as Microsoft Excel in ".xlsx" format. The data is divided into sections, such as lecture, practice, and exam, according to the parts obtained from the system. These sections are collected in different Excel sheets. The data collected from the system was used to understand the correlation between students' success in practice and exam performance. In addition, students' starting and finishing timestamp data were also collected to understand the effect of transitions between lecture, practice, and exam activities on success via process mining.

3.4 Measures

All variables in the Excel document have been renamed for easier understanding and reporting in analysis programs in the study.

The first variable in the raw data is the exam performance data of the students, which is included as the dependent variable in the study. In order to measure the success of the students in the courses included in the learning portal, the number of correct, incorrect, and blank questions applied in different subjects and the score data obtained from the exam were used in the study.

The practices applied by the students in the learning portal to prepare for the exam were used as an independent variable in the study. These practices include drag and drop, fill-in-the-blanks, and matching activities. Students can log in to the same practice more than once. When the same practice is applied again, the places of the questions and choices change. The students' scores from the practices they applied in different courses and on different subjects were used in the study.

The learning portal has a section for students to view the course content and listen to the lecture part. This section is aimed to enable students to listen and learn the narration of different subjects belonging to different courses with course materials. This variable was used as an independent variable in the process analysis.

In addition, the starting and finishing times of all the activities specified by the students are found in raw data as timestamps. The path students took after entering the learning portal was analyzed using this data.

3.4.1 Data Pre-Processing

All data obtained from the learning portal were not included in the study. First, the variable names in the raw data were rearranged and made understandable. Missing values were determined, and the Excel file was rearranged to analyze the path they followed during the learning portal by grouping the secondary school students according to success. The histograms were created by separately calculating the exam score averages of the students in Turkish, Mathematics, Science, and Social Science lessons. With the help of the histograms, the students were divided into two groups: medium and high performers. The starting and finishing timestamps of these two groups of students for learning portal’s activities (lecture, practice, and exam) were rearranged in Excel documents. This way, the raw data was ready for the process analysis study that the students followed in the learning portal according to the success groups.

In addition, in the learning portal, secondary school students' practice, and exam scores are grouped according to the courses in the Excel document to analyze the correlation between these variables. Missing values in practice and exam scores of students in Mathematics, Turkish, Science, and Social Science courses were determined. Students' exam and practice score averages in these courses were calculated without missing values. The students' practice and exam averages were matched in the Excel document, and the data was made ready for the correlation analysis between practice and exam scores.

3.5 Data Analysis

In this study, the Fluxicon Disco tool was used to analyze the path followed by the students during the use of the learning portal. In addition, the Pearson Correlation was used to analyze the relationship between practice and exam scores. The Fluxicon Disco tool and Pearson Correlation were explained in detail in this part of the study.

3.5.1 Process Mining Analysis in Learning Portal Usage

Modern systems provide event logs that can be used for process mining to gain insights, track progress, and improve actual processes. According to van der Aalst (2011), this approach makes it possible to discover, monitor, and optimize crucial processes. In the structure of an event log, there are three essential components. The Case ID specifies the instance or case to which an event or action belongs. The activity component captures the action performed by the event. Finally, the timestamp represents the event's exact time (Moreno et al., 2021). Various process mining tools have been developed in recent years. ProM framework and Fluxicon Disco are frequently used in education research for creating process models and investigating students' activity transitions and frequencies (dos Santos Garcia et al., 2019; Bogarin et al., 2018). This study used the Fluxicon Disco algorithm to analyze the learning process and draw the paths (https://fluxicon.com/disco/). This software has been used in studies with similar research objectives. The software was selected because it has an appropriate algorithm for conducting process analysis in the study. The algorithm is designed to produce logical and understandable results by processing complex processes. The tool visualizes the results as a process map, adapting the desired analysis with measurements and filters to the questions to be investigated. Utilizing the Disco algorithm, the study demonstrated the analysis of the process involving the students' lecture, practice, and exam activities within the learning portal. The students' use of which activity and how much and the transitions between the activities were visualized with the Disco tool analysis. The percentages of the resulting numerical values were added to the process map. In every lesson, two separate analyses were made with the Disco algorithm, and process maps were visualized for the students, who were divided into two data groups as medium and high performers according to the average exam scores they obtained in these lessons. In this way, students' paths in different achievement groups in the learning process were analyzed.

3.5.2 Correlation Analysis Between Exam and Practice Scores

Pearson correlation was used in the study to understand the relationship between students' exams and practice scores. Pearson correlation, which assesses the direction and strength of a two-variable linear relationship, is one of the most used correlation methods (Gravetter & Wallnau, 2014). The correlation coefficient evaluates how often two variables' values increase or decrease together (Boslaugh, 2012). This method is preferred because it can accurately determine the degree, direction, presence, or absence of a relationship between two variables. In this study, students can interactively repeat the information they learned in the lessons in the learning portal’s practice section and prepare for the exams. The average of the scores they obtained from the practice activities and the average of their scores in the exams were arranged in the Excel document and grouped according to the courses. Pearson correlation analysis was applied to the obtained raw data using SPSS software. In this way, the relationship between secondary school students' practice and exam scores was analyzed according to the courses.

CHAPTER 4

RESULTS

The results of the process mining and correlation analyses carried out are explained in detail.

4.1 Process Mining Analysis Results

As stated in the Method section, the students were divided into two groups according to their exam scores in Turkish, Mathematics, Science, and Social Science lessons. Process analysis was run with the Disco algorithm to extract the students' activities into process maps according to achievement levels for each lesson. The exam score average of the students in the learning portal was used as a performance variable. The division was made according to the 25th percentile of the exam scores, and the students were divided into two groups according to the histograms. The cut point was determined according to the shape of the histograms obtained from exam averages. Therefore, those with 75 points and above are considered high performers, and those with 75 points below are medium performers. While the students were divided into two groups according to their exam score averages, the analysis did not include students with missing values in the raw data and students with zero points in the first exam.

The process mining analysis examines students' transitions between activities within the learning portal by creating a process map. Secondary school students were separately divided into two performance groups according to their exam score averages in Turkish, Mathematics, Science, and Social Science lessons. In each group, the analysis of the paths followed by the students on the learning portal between activities was visualized. The activities in the learning portal consist of lecture, practice, and exam sections and are symbolized by boxes. The darkness of the box color and the numbers inside the box represent the frequency of the activity. The lines with arrows between the boxes represent students' transitions between different activities. The thickness of these lines indicates the frequency of transitions between activities.

In the Fluxicon Disco tool, the arrows between the boxes representing the activities and next to the arrows symbolizing the transitions between the activities show the absolute frequencies. The percentages in the boxes show the activity numbers converted to percentages (%). The percentages displayed alongside the arrows that indicate the movement from one activity to another represent the transitions between activities converted into percentages.

4.1.1 Process Mining Analysis Result for Turkish Lesson

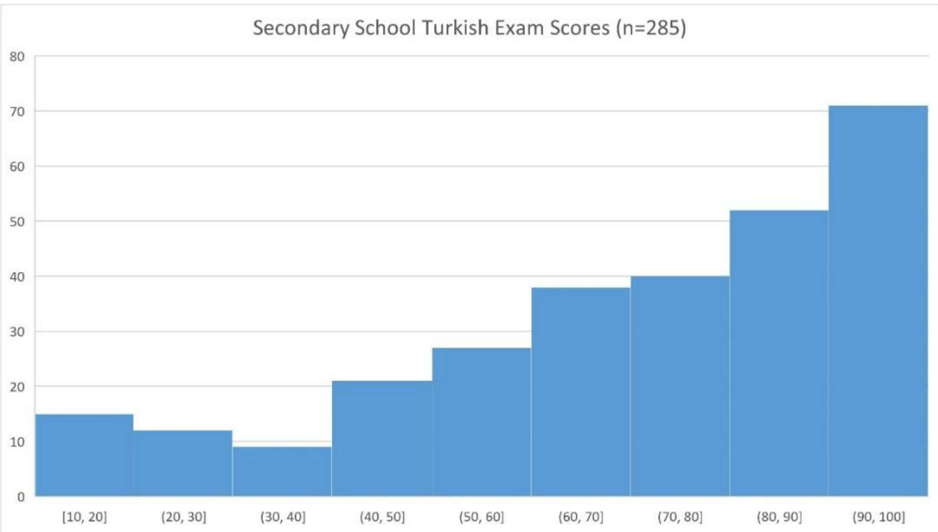

Figure 4.1 Secondary School Turkish Exam Scores

After the students were divided into two groups, process analysis was applied to these groups with the Disco algorithm separately.

As seen in the results of the Turkish Lesson analysis (Figure 4.2), in the high performers' group, the percentage of students performing the practice activity is 36.10%. In comparison, in the medium performers group, it is 31.19%. Furthermore, it is observed that the students in the high performers group repeat the practice (15,84%) activity. On the other hand, this percentage is not found in the medium performers group. Moreover, in the medium performers' group, it was seen that the students preferred the Lecture→Exam (7,58%) path. On the other hand, it is seen that the high performers' group does not prefer the Lecture→Exam path.

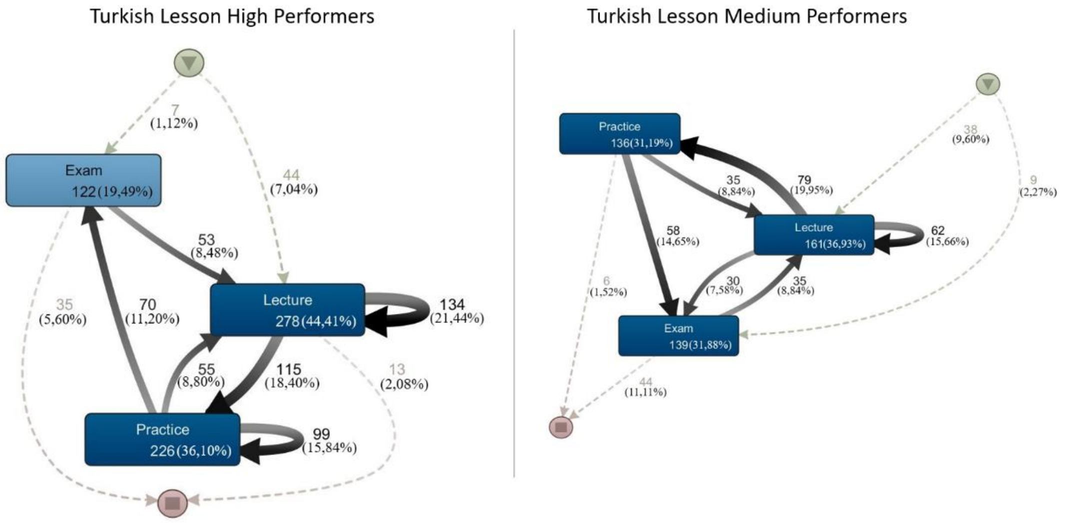

Figure 4.2 Process Analysis for Turkish Lesson

4.1.2 Process Mining Analysis Result for Mathematics Lesson

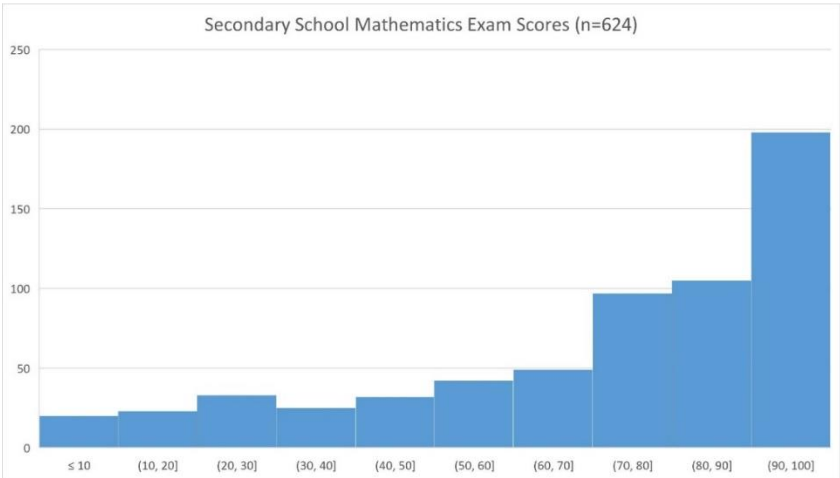

Figure 4.3 Secondary School Mathematics Exam Scores

After the students were divided into two groups, process analysis was applied to these groups with the Disco algorithm separately.

As seen in the results of the Mathematics Lesson analysis (Figure 4.4), the percentage of students performing the practice activity is 32.92% in the high performers' group. In comparison, in the medium performers' group, it is 31.71%. Moreover, the Practice→Exam path is preferred 0.90% more in the high performers group than the medium performers group.

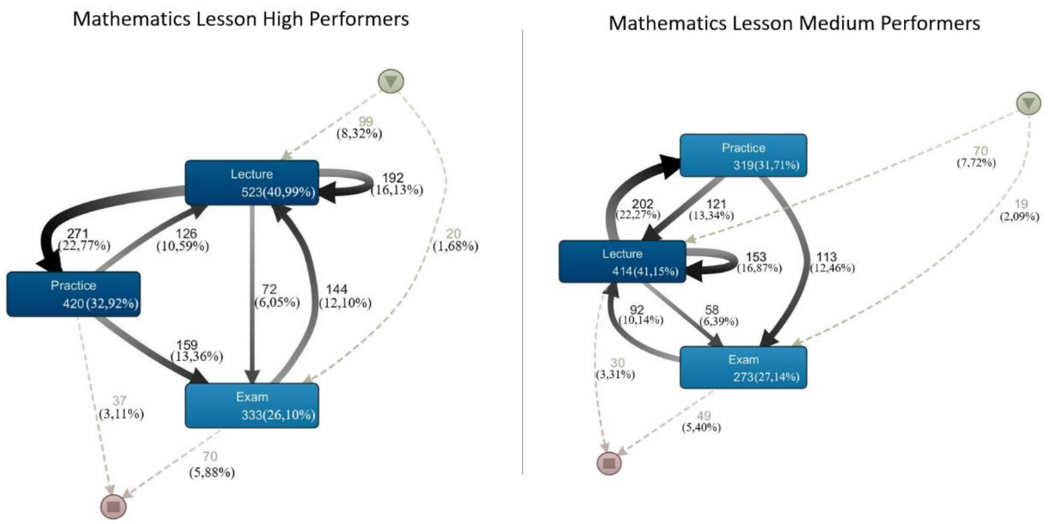

Figure 4.4 Process Analysis for Mathematics Lesson

4.1.3 Process Mining Analysis Result for Science Lesson

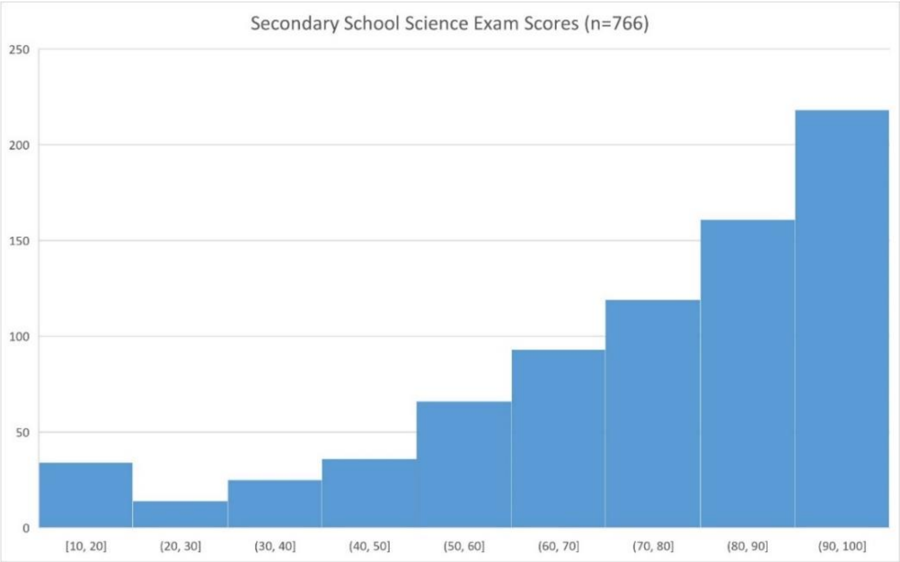

Figure 4.5 Secondary School Science Exam Scores

After the students were divided into two groups, process analysis was applied to these groups with the Disco algorithm separately.

As seen in the results of the Science Lesson analysis (Figure 4.6), in the medium performers' group, it was seen that the students preferred the Lecture→Exam (6.37%) path. On the other hand, it is seen that the high performers' group does not prefer the Lecture→Exam path. In addition, the Practice→Exam path is preferred 2.03% more in the high performers' group than the medium performers' group. Moreover, it is observed that the students in the high performers group repeat the exam (9.87%) activity. On the other hand, this percentage is not found in the medium performers group.

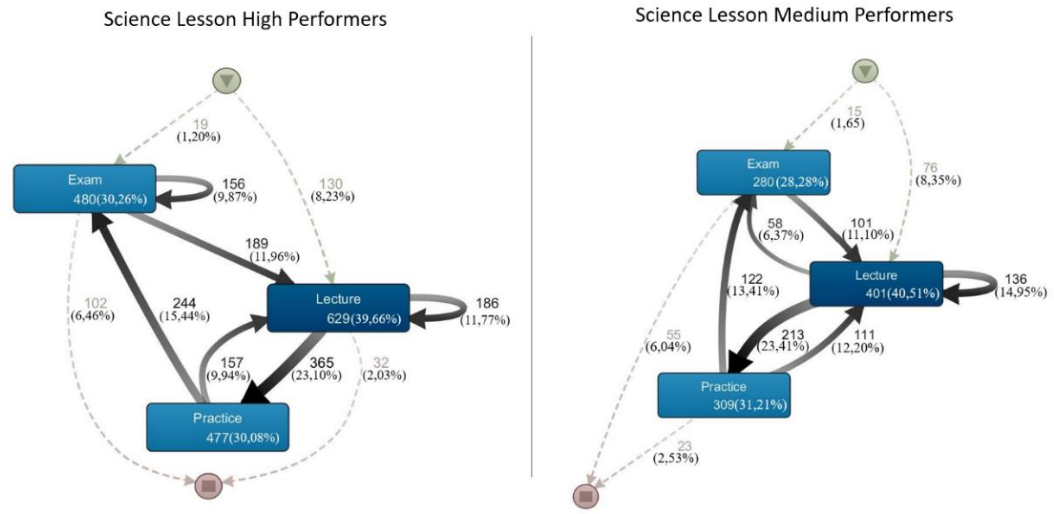

Figure 4.6 Process Analysis for Science Lesson

4.1.4 Process Mining Analysis Result for Social Science Lesson

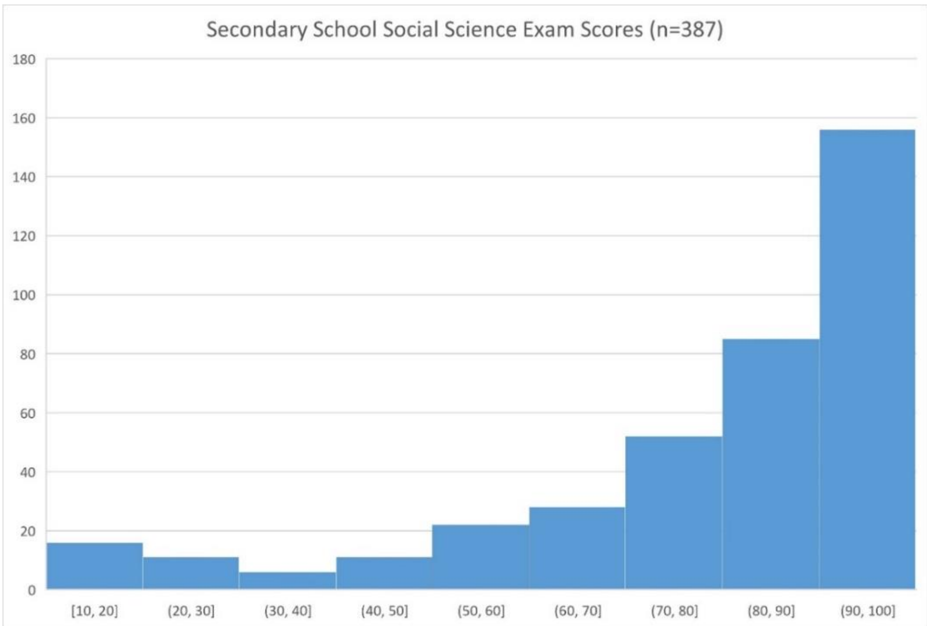

Figure 4.7 Secondary School Social Science Exam Scores

After the students were divided into two groups, process analysis was applied to these groups with the Disco algorithm separately.

As seen in the results of the Social Science Lesson analysis (Figure 4.8), the percentage of students performing the Practice activity is 29.68% in the high performers' group. In comparison, in the medium performers' group, it is 21.69%. In addition, the high performers’ group followed the path of Lecture→Practice (19,66%) and Practice→Exam (17,20%). On the other hand, the medium performers' group followed the path of Lecture→Practice (14,29%) and Practice→Exam (10,26%). High performers' path of Lecture→Practice→Exam (36,86%) is 12.31% higher than the medium performers' path of Lecture→Practice→Exam (24,55%). Moreover, in the high performers group, it was seen that the students preferred the Practice→Lecture (6,84%) path. On the other hand, it is seen that the Practice→Lecture path is not observed in the medium performers’ group.

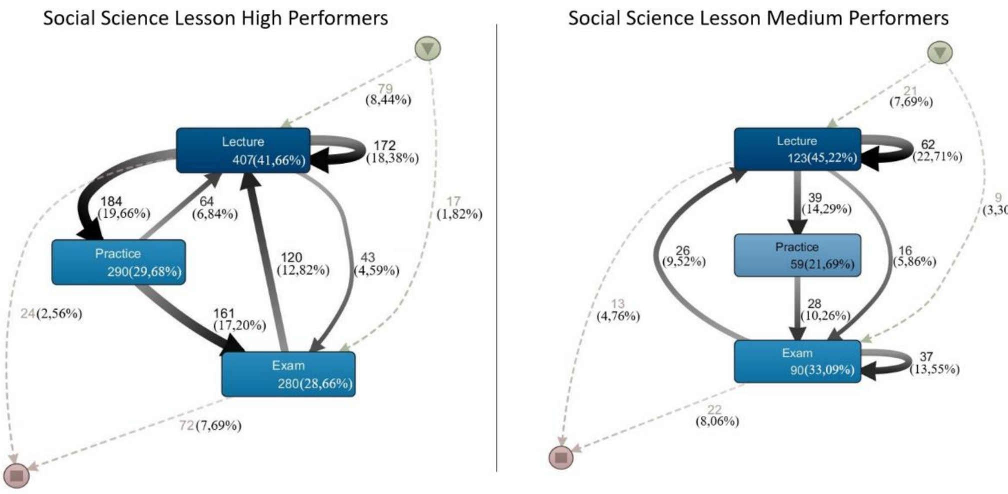

Figure 4.8 Process Analysis for Social Science Lesson

4.2 Correlation Analysis Results

In the study, Pearson Correlation analysis was applied to examine the relationship between the exam and the practice scores of secondary school students conducted in the learning portal in Turkish, Mathematics, Science, and Social Science courses.

In order to improve the transparency and rigor of the analysis, it is crucial to thoroughly examine the linearity assumption before analyzing the Pearson correlation results. This is accomplished by utilizing linearity assumption tables, which provide a comprehensive overview of the linear relationships among the variables being considered. In this study, before investigating the correlation between practice and exam scores, linearity was examined for each course. The tables presented beforehand establish a strong foundation for interpreting Pearson correlation results. This ensures detailed, well-informed, and meaningful conclusions from the study's findings.

Analysis was done in SPSS software, and analysis outputs were added as tables. At the secondary school level, linearity assumption, Pearson Correlation analysis result tables, and explanations applied separately for each course were explained in detail.

4.2.1 Correlation Analysis Result for Turkish Lesson

Table 4.1 *Linearity Assumption for Turkish Lesson*

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
TurkishExam * TurkishPractice	Between Groups	(Combined)	39443,350	66	597,627	1,769	,068
		Linearity	12179,161	1	12179,161	36,053	,000
		Deviation from Linearity	27264,189	65	419,449	1,242	,292
	Within Groups		7431,826	22	337,810		
	Total		46875,177	88			

Table 4.1 examines the assumption of linearity between the practice and exam scores of students in the Turkish lesson. It can be seen in the Deviation from the Linearity, 0.292 > 0.05, which concluded a linear relationship between the two variables.

Table 4.2 *Correlations for Turkish Lesson*

	Practice	Exam
Practice	.	
Exam	.51**	.

Note. N = 89. **Correlation is significant at the 0.001 level (2-tailed)

A Pearson correlation coefficient was computed to assess the linear relationship between practice score and exam score in Turkish Lesson at secondary school level.

There was a positive correlation between the two variables, $r = +0.51$, $n = 89$, $p < .001$, two tails.

4.2.2 Correlation Analysis Result for Mathematics Lesson

Table 4.3 *Linearity Assumption for Mathematics Lesson*

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
MathematicsExam * MathematicsPractice	Between Groups	(Combined)	69445,327	112	620,048	1,214	,197
		Linearity	13914,647	1	13914,647	27,237	,000
		Deviation from Linearity	55530,680	111	500,276	,979	,545
	Within Groups		33717,592	66	510,873		
	Total		103162,920	178			

In Table 4.3, the linearity assumption between the Mathematics lesson practice and exam scores of the students is explored. The Deviation from Linearity, which is 0.545, indicates a linear relationship between the two variables as it is more significant than 0.05.

Table 4.4 *Correlations for Mathematics Lesson*

	Practice	Exam
Practice	.	
Exam	.37**	.

Note. N = 179. **Correlation is significant at the 0.001 level (2-tailed)

A Pearson correlation coefficient was computed to assess the linear relationship between practice score and exam score in Mathematics Lesson at secondary school level.

There was a positive correlation between the two variables, $r = +0.37$, $n = 179$, $p < .001$, two tails.

4.2.3 Correlation Analysis Result for Science Lesson

Table 4.5 *Linearity Assumption for Science Lesson*

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
ScienceExam * SciencePractice	Between Groups	(Combined)	50857,656	111	458,177	1,192	,182
		Linearity	11737,025	1	11737,025	30,539	,000
		Deviation from Linearity	39120,631	110	355,642	,925	,656
	Within Groups		40353,975	105	384,324		
	Total		91211,631	216			

Table 4.5 examines the linear relationship between students' science lesson practice and exam scores. Deviation from Linearity, 0.656 > 0.05, indicating a significant linear relationship.

Table 4.6 *Correlations for Science Lesson*

	Practice	Exam
Practice	.	
Exam	.36**	.

Note. N = 217. **Correlation is significant at the 0.001 level (2-tailed)

A Pearson correlation coefficient was computed to assess the linear relationship between practice score and exam score in Science Lesson at secondary school level.

There was a positive correlation between the two variables, $r = +0.36$, $n = 217$, $p < .001$, two tails.

4.2.4 Correlation Analysis Result for Social Science Lesson

Table 4.7 *Linearity Assumption for Social Science Lesson*

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
SocialScienceExam * SocialSciencePractice	Between Groups	(Combined)	22846,378	62	368,490	1,093	,375
		Linearity	5325,455	1	5325,455	15,801	,000
		Deviation from Linearity	17520,923	61	287,228	,852	,725
	Within Groups		16514,888	49	337,039		
	Total		39361,266	111			

Table 4.7 examines the correlation between students' social science lesson practice and exam scores. Deviation from linearity (0.725 > 0.05) indicates a significant linear relationship.

Table 4.8 *Correlations for Social Science Lesson*

	Practice	Exam
Practice	.	
Exam	.37**	.

Note. N = 112. **Correlation is significant at the 0.001 level (2-tailed)

A Pearson correlation coefficient was computed to assess the linear relationship between practice score and exam score in Social Science Lesson at secondary school level.

There was a positive correlation between the two variables, $r = +0.37$, $n = 112$, $p < .001$, two tails.

A positive correlation was found in each lesson between the reinforcement activities carried out in the practices and the grades that the students got from the exams of the lessons. As explained in the results of the Pearson Correlation analysis, the practice made had a positive impact on learning in Turkish, Mathematics, Science and Social Science lessons' exam scores.

CHAPTER 5

DISCUSSION AND CONCLUSION

In this study, the processes between lecture, practice, and exam activities in learning portal were examined according to success groups, and it was determined how the frequency and sequence of transitions between activities relate to students' achievement level in Turkish, Mathematics, Science, and Social Studies courses. Moreover, the research examined the data obtained from the learning portal used by middle school students, the correlation between the exam scores, and the practice scores in the system. The study utilized learning analytics, process mining techniques, and Pearson Correlation for analysis. When analyzing the students' activities in the learning portal, their starting and finishing times were taken into consideration. In this way, the sequence and frequency of activities were analyzed through path analysis. Moreover, correlation analysis was applied to examine the relationship between exam score and practice score. The research questions were thoroughly analyzed, and as a result, the answers to these questions were uncovered.

The study's results suggest a significant positive correlation between the students' exam scores and their practice scores in Turkish, Mathematics, Science, and Social Science courses at the secondary school level in learning portal. The Pearson Correlation analysis was utilized for each lesson separately, and a positive relationship was observed in all courses (see Chapter 4.2). Furthermore, in the learning portal, the students' process mining analysis in the activities was examined separately for each course in two success groups: high performers and medium performers (see Chapter 4.1). As shown in Figure 4.2, Figure 4.4, and Figure 4.8, the results suggest that the high performer group engages in more practice activities than the medium performer group. Another finding is that the high performer group takes exams more frequently after practice sessions than the medium performer group (see Figure 4.4, Figure 4.6, and Figure 4.8). Lastly, the data suggest that students in the medium performer group commonly skipped the practice activity and proceeded directly to the exam after the lecture (see Figure 4.2 and Figure 4.6).

These findings support that practice activities positively impact students' exam performance, and those who practice are more likely to succeed. The findings of this research are consistent with the studies in the literature. Schwerter's (2022) research findings reveal that voluntary practice positively impacts exam performance, suggesting that practicing leads to improved grades. In addition, according Roediger and Karpicke (2006), practice testing has shown a positive impact in making learning effective, and this is called the testing effect.

The analysis results by creating the process map suggest that practice is essential for efficient learning and positively predicts exam performance (see Chapter 4.1). According to an online interaction data study from Sun et al. (2019), consolidating knowledge by focusing on content learning, assignments, and quizzes is common among high-achieving students. On the other hand, low-achieving students tend to be inactive in their learning process but become opportunistic during exams. The process maps that emerged from the analysis suggest that students in the high performers group have a higher transition rate from Practice to Exam (see Figure 4.4, Figure 4.6, and Figure 4.8). This result indicates a positive relationship between practice activities and exam performance. According to Roedieger and Butler (2011), it has been observed that learning is more efficient when the study and testing phases are applied consecutively during learning. This study supports that students' exam performance increases when they apply to practice and exam applications one after another on online learning platforms. In addition, as shown in Figure 4.2 and Figure 4.6, process mining analysis shows that the medium performers' group students tend to switch directly to the Exam activity after the Lecture activity. Research suggests that practicing is an effective method for long-term learning and mastery of new information (Yang et al., 2017). The recent findings further emphasize the significance of engaging in practice activities.

The importance of this study is that, although many studies examine the relationship between practice and exam, process mining analysis determines students' learning habits according to their success groups and supports this with correlation analysis between variables. As process mining is a developing field, it is crucial to bridge the knowledge gap and contribute to the education of instructors and researchers. Moreover, there are few studies in the literature on young learners and their behaviors in online learning environments, which highlights the significance of this study.

In the study, Disco software was used to analyze the obtained data. This tool maps the process on the system with the starting and finishing timestamps of the activities. One of the software's advantages is that it shows the frequency values and sequence on the process map. This software was preferred because it has the most suitable algorithm for the process analysis to be performed in the study. In addition, the Pearson Correlation method was preferred to examine the relationship between exam and practice performance data. This method is preferred because of its advantage in determining the degree, direction, presence, or absence of the relationship between two variables. The study determined that it was the most appropriate analysis method for examining the relationship between dependent and independent variables in the obtained data. For this reason, these analysis methods were applied in the study.

In summary, the results highlight two important implications. First, practice activities significantly impact the realization of efficient learning. Students who practice and repeat tend to have more effective learning. Secondly, the order and frequency of the activities carried out in education are essential for efficient learning. The correct order and frequency of lecture, practice, and exam activities support more effective and long-lasting student learning.

The study has some limitations. The research was conducted with a limited number of students and specific courses. The larger sample size and different courses in the learning portal may provide more proper results. Additionally, it was assumed that the students in the study used the learning portal as desired. The research data was collected during the Covid-19 pandemic. Students’ practice and exam scores are collected in an uncontrolled learning environment.

In future studies, students should be required to complete practice activities in the correct sequence and frequency within the online learning environments. In recent years, Artificial intelligence (AI) has been employed in various fields, including education. Incorporating AI technology into online learning platforms may optimize the learning experience and improve efficiency. AI can be utilized to examine the learning habits of students and analyze the frequency of their practice activities. Furthermore, with the help of AI, it can be recommended to follow the lecture, practice, and exam path in the correct sequence on these platforms. By analyzing students' learning habits with the practice and exam scores obtained, immediate interventions can be made with AI to suggest optimal usage frequency of the activities. Additionally, the spaced practice effect can be observed in the future studies by controlling students’ practice times. Finally, the results should be linked to providing more effective and permanent learning by finding methods to enable students to perform their practice activities.

REFERENCES

Aalst, W. van der. Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer-Verlag, Berlin, 2011.

Abdous, M. H., Wu, H., & Yen, C. J. (2012). Using data mining for predicting relationships between online question theme and final grade. *Journal of Educational Technology & Society*, *15*(3), 77.

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in human behavior*, *31*, 542-550.

Ahmad Uzir, N. A., Gašević, D., Matcha, W., Jovanović, J., & Pardo, A. (2020). Analytics of time management strategies in a flipped classroom. Journal of Computer Assisted Learning, 36(1), 70-88.

Aisa, V., & Prima, A. (2015). Evaluation of the online assessment test using process mining. In International Conference on Information and Communication Technology.

Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, *37*, 13-49.

Alejandro Bogarín, Rebeca Cerezo, and Cristóbal Romero. 2018. A survey on educational process mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8, 1 (2018), e1230.

Ali, L., Hatala, M., Gašević, D., & Jovanović, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education*, *58*(1), 470-489.

Alvarez, P., Fabra, J., Hernández, S., & Ezpeleta, J. (2016, September). Alignment of teacher's plan and students' use of lms resources. analysis of moodle logs. In 2016 15th international conference on information technology based higher education and training (ithet) (pp. 1-8). IEEE.

Ardimento, P., Bernardi, M. L., Cimitile, M., & De Ruvo, G. (2019, June). Learning analytics to improve coding abilities: a fuzzy-based process mining approach. In 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-7). IEEE.

Ariel, R., & Karpicke, J. D. (2018). Improving self-regulated learning with a retrieval practice intervention. *Journal of Experimental Psychology: Applied*, *24*(1), 43.

Ariouat, H., Cairns, A. H., Barkaoui, K., Akoka, J., & Khelifa, N. (2016, June). A two-step clustering approach for improving educational process model discovery. In 2016 ieee 25th international conference on enabling technologies: Infrastructure for collaborative enterprises (wetice) (pp. 38-43). IEEE.

Arnold, K. E., & Pistilli, M. D. (2012, April). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 267-270).

Avella, J. T., Kebritchi, M., Nunn, S. G., & Kanai, T. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning*, *20*(2), 13-29.

Azzini, A., Ceravolo, P., Scarabottolo, N., & Damiani, E. (2016, April). On the predictive power of university curricula. In 2016 IEEE global engineering education conference (educon) (pp. 929-932). IEEE.

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning analytics* (pp. 61-75). Springer, New York, NY.

Bala, S., Revoredo, K., & Mendling, J. (2022, October). Process Mining for Analyzing Open Questions Computer-Aided Examinations. In *International*

Conference on Process Mining (pp. 565-576). Cham: Springer Nature Switzerland.

Banihashem, S. K., Aliabadi, K., Pourroostaei Ardakani, S., Delaver, A., & Nili Ahmadabadi, M. (2018). Learning analytics: A systematic literature review. *Interdisciplinary Journal of Virtual Learning in Medical Sciences*, *9*(2).

Barla, M., Bieliková, M., Ezzeddinne, A. B., Kramár, T., Šimko, M., & Vozár, O. (2010). On the impact of adaptive test question selection for learning efficiency. *Computers & Education*, *55*(2), 846-857.

Bartschat, A, Reischl, M, Mikut, R. Data mining tools. *WIREs Data Mining Knowl Discov*. 2019; 9:e1309. https://doi.org/10.1002/widm.1309

Beheshitha, S. S., Gašević, D., & Hatala, M. (2015, March). A process mining approach to linking the study of aptitude and event facets of self-regulated learning. In Proceedings of the fifth international conference on learning analytics and knowledge (pp. 265-269).

Bogarín, A., Cerezo, R., & Romero, C. (2018). A survey on educational process mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(1), e1230.

Bogarín, A., Romero, C., Cerezo, R., & Sánchez-Santillán, M. (2014, March). Clustering for improving educational process mining. In Proceedings of the fourth international conference on learning analytics and knowledge (pp. 11-15).

Brown, P. C., Roediger III, H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*. Harvard University Press.

Cairns, A. H., Gueni, B., Hafdi, H., Joubert, C., & Khelifa, N. (2015, July). Towards a distributed computation platform tailored for educational process discovery and analysis. In 2015 International Conference on Protocol Engineering

(ICPE) and International Conference on New Technologies of Distributed Systems (NTDS) (pp. 1-8). IEEE.

Cameranesi, M., Diamantini, C., Genga, L., & Potena, D. (2017, June). Students' careers analysis: a process mining approach. In Proceedings of the 7th International Conference on Web Intelligence, Mining and Semantics (pp. 1-7).

Carpenter, S. K., Pan, S. C., & Butler, A. C. (2022). The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology*, *1*(9), 496-511.

Castillo, L. (2016, September). A virtual laboratory for multiagent systems: Joining efficacy, learning analytics and student satisfaction. In 2016 International Symposium on Computers in Education (SIIE) (pp. 1-6). IEEE.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. Psychological bulletin, 132(3), 354.

Chen, C. M., & Chen, M. C. (2009). Mobile formative assessment tool based on data mining techniques for supporting web-based learning. *Computers & Education*, *52*(1), 256-273.

Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. john Wiley & sons.

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, *18*(6), 683-695.

Clow, D., & Makriyannis, E. (2011, February). iSpot analysed: participatory learning and reputation. In *Proceedings of the 1st international conference on learning analytics and knowledge* (pp. 34-43).

D'Oca, S., & Hong, T. (2015). Occupancy schedules learning process through a data mining framework. *Energy and Buildings*, *88*, 395-408.

De Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., ... & Arnab, S. (2015). Foundations of dynamic learning analytics: Using university student data to increase retention. *British journal of educational technology*, *46*(6), 1175-1188.

Dejaeger, K., Goethals, F., Giangreco, A., Mola, L., & Baesens, B. (2012). Gaining insight into student satisfaction using comprehensible data mining techniques. *European Journal of Operational Research*, *218*(2), 548-562.

Dekker, G. W., Pechenizkiy, M., & Vleeshouwers, J. M. (2009). Predicting Students Drop Out: A Case Study. *International Working Group on Educational Data Mining*.

Dietz-Uhler, B., & Hurn, J. E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of interactive online learning*, *12*(1), 17-26.

Dinu, V. E., Papuc, D., Gheorghiu, A., Dascalu, M. I., Moldoveanu, A., & Moldoveanu, F. (2017). BIOMETRIC DATA IN LEARNING ANALYTICS: A SURVEY ON EXISTING APPLICATIONS. *eLearning & Software for Education*, *2*.

Doleck, T., Jarrell, A., Poitras, E. G., Chaouachi, M., & Lajoie, S. P. (2016, February). Examining diagnosis paths: A process mining approach. In 2016 Second International Conference on Computational Intelligence & Communication Technology (CICT) (pp. 663-667). IEEE.

dos Santos Garcia, C., Meincheim, A., Junior, E. R. F., Dallagassa, M. R., Sato, D. M. V., Carvalho, D. R., ... & Scalabrin, E. E. (2019). Process mining techniques and applications–A systematic mapping study. Expert Systems with Applications, 133, 260-295.

Dunlosky, J., & Rawson, K. A. (2015). Practice tests, spaced practice, and successive relearning: Tips for classroom use and for guiding students' learning. *Scholarship of Teaching and Learning in Psychology*, *1*(1), 72.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public interest*, *14*(1), 4-58.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, *17*(3), 37-37.

Fournier, H., Kop, R., & Sitlia, H. (2011, February). The value of learning analytics to networked learning on a personal learning environment. In *Proceedings of the 1st international conference on learning analytics and knowledge* (pp. 104-109).

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64-71.

Gašević, D., Mirriahi, N., Dawson, S., & Joksimović, S. (2017). Effects of instructional conditions and experience on the adoption of a learning tool. Computers in Human Behavior, 67, 207-220.

Giesbers, B., Rienties, B., Tempelaar, D., & Gijselaers, W. (2013). Investigating the relations between motivation, tool use, participation, and performance in an e-learning course using web-videoconferencing. *Computers in Human Behavior*, *29*(1), 285-292.

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

Gursoy, M. E., Inan, A., Nergiz, M. E., & Saygin, Y. (2016). Privacy-preserving learning analytics: challenges and techniques. *IEEE Transactions on Learning technologies*, *10*(1), 68-81.

Guruler, H., Istanbullu, A., & Karahasan, M. (2010). A new student performance analysing system using knowledge discovery in higher educational databases. *Computers & Education*, *55*(1), 247-254.

He, W. (2013). Examining students' online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior*, *29*(1), 90-102.

Hobeck, R., Pufahl, L., & Weber, I. (2022, October). Process Mining on Curriculum-Based Study Data: A Case Study at a German University. In *International Conference on Process Mining* (pp. 577-589). Cham: Springer Nature Switzerland.

Ifenthaler, D. (2017). Are higher education institutions prepared for learning analytics?. *TechTrends*, *61*(4), 366-371.

Ifenthaler, D., Schumacher, C., & Kuzilek, J. (2023). Investigating students' use of self-assessments in higher education using learning analytics. *Journal of Computer Assisted Learning*, *39*(1), 255-268.

Intayoad, W., Kamyod, C., & Temdee, P. (2018, February). Process mining application for discovering student learning paths. In 2018 International Conference on Digital Arts, Media and Technology (ICDAMT) (pp. 220-224). IEEE.

Juhaňák, L., Zounek, J., & Rohlíková, L. (2019). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. Computers in Human Behavior, 92, 496-506.

Kang, S. H. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 12-19.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, *138*(4), 469.

Khalil, M., & Ebner, M. (2015, June). Learning analytics: principles and constraints. In *EdMedia+ Innovate Learning* (pp. 1789-1799). Association for the Advancement of Computing in Education (AACE).

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *The American journal of psychology*, 329-343.

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

Krause, U. M., Stark, R., & Mandl, H. (2009). The effects of cooperative learning and feedback on e-learning in statistics. *Learning and instruction*, *19*(2), 158-170.

Kriegel, H. P., Borgwardt, K. M., Kröger, P., Pryakhin, A., Schubert, M., & Zimek, A. (2007). Future trends in data mining. *Data Mining and Knowledge Discovery*, *15*(1), 87-97.

Kornell, N., & Bjork, R. A. (2008). Optimising self-regulated study: The benefits—and costs—of dropping flashcards. *Memory*, *16*(2), 125-136.

Leong, C. K., Lee, Y. H., & Mak, W. K. (2012). Mining sentiments in SMS texts for teaching evaluation. *Expert Systems with Applications*, *39*(3), 2584-2589.

Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications–A decade review from 2000 to 2011. *Expert systems with applications*, *39*(12), 11303-11311.

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*, *57*(10), 1439-1459.

Lykourentzou, I., Giannoukos, I., Mpardis, G., Nikolopoulos, V., & Loumos, V. (2009). Early and dynamic student achievement prediction in e-learning courses using neural networks. *Journal of the American Society for Information Science and Technology*, *60*(2), 372-380.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & education*, *54*(2), 588-599.

Maita, A. R., Fantinato, M., Peres, S. M., Thom, L. H., & Hung, P. C. (2017, May). Mining unstructured processes: An exploratory study on a distance learning domain. In 2017 international joint conference on neural networks (ijcnn) (pp. 3240-3247). IEEE.

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

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity.

Matcha, W., Gašević, D., Uzir, N. A. A., Jovanović, J., & Pardo, A. (2019, March). Analytics of learning strategies: Associations with academic performance and feedback. In Proceedings of the 9th international conference on learning analytics & knowledge (pp. 461-470).

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of educational Psychology*, *84*(4), 444.

Millet, A., Turcotte, N., & Yan, S. (2021). Retrieval Practice and Online Learning. In *International Perspectives in Online Instruction* (Vol. 40, pp. 95-112). Emerald Publishing Limited.

Mohamad, S. K., & Tasir, Z. (2013). Educational data mining: A review. *Procedia-Social and Behavioral Sciences*, *97*, 320-324.

Moore, M. G. (2001). Surviving as a distance teacher.

Moreno, M., Exposito, E., & Gueye, M. (2021, January). Process Mining Model to visualize and analyze the Learning Process. In REES AAEE 2021 conference: Engineering Education Research Capability Development: Engineering Education Research Capability Development (pp. 698-706). Perth, WA: Engineers Australia.

Mukala, P., Buijs, J. C. A. M., Leemans, M., & van der Aalst, W. (2015). Learning analytics on coursera event data: A process mining approach. In 5th International Symposium on Data-Driven Process Discovery and Analysis (SIMPDA 2015 (pp. 18-32). CEUR-WS. org.

Mukala, P., Cerone, A., & Turini, F. (2017). An empirical verification of a-priori learning models on mailing archives in the context of online learning activities of participants in free\libre open source software (FLOSS) communities. Education and Information Technologies, 22, 3207-3229.

Neyem, A., Diaz-Mosquera, J., Munoz-Gama, J., & Navon, J. (2017, March). Understanding student interactions in capstone courses to improve learning experiences. In Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education (pp. 423-428).

Olson, D. L. (2007). Data mining in business services. *Service Business*, *1*(3), 181-193.

Ouyang, F., Wu, M., Zheng, L., Zhang, L., & Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education*, *20*(1), 1-23.

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.

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, *41*(4), 1432-1462.

Petersen-Brown, S., Lundberg, A. R., Ray, J. E., Dela Paz, I. N., Riss, C. L., & Panahon, C. J. (2019). Applying spaced practice in the schools to teach math vocabulary. *Psychology in the Schools*, *56*(6), 977-991.

Priyambada, S. A., Mahendrawathi, E. R., & Yahya, B. N. (2017). Curriculum assessment of higher educational institution using aggregate profile clustering. Procedia Computer Science, 124, 264-273.

Qiao, C., & Hu, X. (2018, July). Discovering student behavior patterns from event logs: Preliminary results on a novel probabilistic latent variable model. In 2018 ieee 18th international conference on advanced learning technologies (icalt) (pp. 207-211). IEEE.

Rawson, K. A., Thomas, R. C., & Jacoby, L. L. (2015). The power of examples: Illustrative examples enhance conceptual learning of declarative concepts. *Educational Psychology Review*, *27*, 483-504.

Reyes, J. A. (2015). The skinny on big data in education: Learning analytics simplified. *TechTrends*, *59*(2), 75-80.

Rodríguez, M. C., Nistal, M. L., Fonte, F. A. M., Penín, M. L., & Molina, M. M. (2018, April). Exploring the application of process mining to support self-regulated learning: An initial analysis with video lectures. In 2018 IEEE global engineering education conference (EDUCON) (pp. 1766-1774). IEEE.

Roediger III, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on psychological science*, *1*(3), 181-210.

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

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, *24*, 355-367.

Romero C, Ventura S, Pechenizky M, Baker R. *Handbook of Educational Data Mining*. Data Mining and Knowledge Discovery Series. Boca Raton, FL: Chapman and Hall/CRC Press; 2010.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert systems with applications*, *33*(1), 135-146.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12-27.

Romero, C., Ventura, S., Zafra, A., & De Bra, P. (2009). Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems. *Computers & Education*, *53*(3), 828-840.

Romero-Zaldivar, V. A., Pardo, A., Burgos, D., & Kloos, C. D. (2012). Monitoring student progress using virtual appliances: A case study. *Computers & Education*, *58*(4), 1058-1067.

Santos, O. C., & Boticario, J. (2012, September). Affective issues in Semantic Educational Recommender Systems. In *RecSysTEL@ EC-TEL* (pp. 71-82)

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in human behavior*, *78*, 397-407.

Schwerter, J., Dimpfl, T., Bleher, J., & Murayama, K. (2022). Benefits of additional online practice opportunities in higher education. *The Internet and Higher Education*, *53*, 100834.

Sedrakyan, G., De Weerdt, J., & Snoeck, M. (2016). Process-mining enabled feedback:“tell me what I did wrong” vs.“tell me how to do it right”. Computers in human behavior, 57, 352-376.

Sedrakyan, G., Snoeck, M., & De Weerdt, J. (2014). Process mining analysis of conceptual modeling behavior of novices–empirical study using JMermaid modeling and experimental logging environment. Computers in Human Behavior, 41, 486-503.

Shih, B., Koedinger, K. R., & Scheines, R. (2008). A Response Time Model For Bottom-Out Hints as Worked Examples. *EDM*, *2008*, 117-126.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, *57*(10), 1380-1400.

Siemens, G., & Baker, R. S. D. (2012, April). Learning analytics and educational data mining: towards communication and collaboration. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 252-254).

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, *46*(5), 30.

Simon, J. (2017). A priori knowledge in learning analytics. In A. Peña-Ayala (Ed.), Learning Analytics: Fundaments, Applications, and Trends (Vol. 94, pp. 199–227). Cham, Switzerland: Springer. doi: doi.org/10.1007/978-3-319-52977-6.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510-1529.

Sobocinski, M., Malmberg, J., & Järvelä, S. (2017). Exploring temporal sequences of regulatory phases and associated interactions in low-and high-challenge collaborative learning sessions. Metacognition and Learning, 12, 275-294.

Sun, D., Cheng, G., Xu, P., Zheng, Q., & Chen, L. (2019). Using HMM to compare interaction activity patterns of student groups with different achievements in MPOCs. *Interactive Learning Environments*, *27*(5-6), 766-781.

Sun, P. C., Tsai, R. J., Finger, G., Chen, Y. Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & education*, *50*(4), 1183-1202.

Terzis, V., & Economides, A. A. (2011). The acceptance and use of computer based assessment. *Computers & Education*, *56*(4), 1032-1044.

Toppino, T. C., & Cohen, M. S. (2010). Metacognitive control and spaced practice: clarifying what people do and why. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1480.

Trcˇka, N., Pechenizkiy, M., & van der Aalst, W. (2010). Process mining from educational data. Handbook of educational data mining, 123-142.

Trcka, N., & Pechenizkiy, M. (2009, November). From local patterns to global models: Towards domain driven educational process mining. In 2009 Ninth international conference on intelligent systems design and applications (pp. 1114-1119). IEEE.

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.

Uzir, N. A. A., Gašević, D., Jovanović, J., Matcha, W., Lim, L. A., & Fudge, A. (2020, March). Analytics of time management and learning strategies for effective online learning in blended environments. In Proceedings of the tenth international conference on learning analytics & knowledge (pp. 392-401).

Van Der Aalst, W. (2012). Process mining: Overview and opportunities. ACM Transactions on Management Information Systems (TMIS), 3(2), 1-17.

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, *89*, 98-110.

Vidal, J. C., Vázquez-Barreiros, B., Lama, M., & Mucientes, M. (2016). Recompiling learning processes from event logs. Knowledge-Based Systems, 100, 160-174.

Wang, R., & Zaïane, O. R. (2015). Discovering Process in Curriculum Data to Provide Recommendation. In EDM (pp. 580-581).

Wang, Y., White, W. M., & Andersen, E. (2017, May). Pathviewer: Visualizing pathways through student data. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 960-964).

West, D., Huijser, H., & Heath, D. (2016). Putting an ethical lens on learning analytics. *Educational Technology Research and Development*, *64*(5), 903-922.

Wilson, K., Boyd, C., Chen, L., & Jamal, S. (2011). Improving student performance in a first-year geography course: Examining the importance of computer-assisted formative assessment. *Computers & Education*, *57*(2), 1493-1500.

Worsley, M., & Blikstein, P. (2013, April). Towards the development of multimodal action based assessment. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 94-101).

Yang, C. C., & Ogata, H. (2023). Personalized learning analytics intervention approach for enhancing student learning achievement and behavioral engagement in blended learning. *Education and Information Technologies*, *28*(3), 2509-2528.

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, *147*(4), 399.

Yang, C., Potts, R., & Shanks, D. R. (2017). The forward testing effect on self-regulated study time allocation and metamemory monitoring. Journal of Experimental Psychology: Applied, 23(3), 263.