Title	オンライン学習環境における自己調整学習能力認知フレームワークとMarkovモデル
Author(s)	TRAN, MINH TUAN
Citation	
Issue Date	2022-09
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/18131
Rights	
Description	Supervisor:長谷川 忍, 先端科学技術研究科, 博士

A PRINCIPLE FRAMEWORK AND A MARKOV MODEL FOR SELF-REGULATED LEARNING ABILITY RECOGNITION IN ONLINE LEARNING ENVIRONMENTS

TRAN Minh Tuan

Japan Advanced Institute of Science and Technology

Doctoral Dissertation

A PRINCIPLE FRAMEWORK AND A MARKOV MODEL FOR SELF-REGULATED LEARNING ABILITY RECOGNITION IN ONLINE LEARNING ENVIRONMENTS

TRAN Minh Tuan

Supervisor：HASEGAWA Shinobu

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
Information Science

September 2022

Abstract

We are motivated to discover and to know. And one thing that we always do consciously or unconsciously is to learn. Discovering knowledge is motivating. But it is even more desirable to know ourselves, our way of learning, and our learning habits with their strengths and weaknesses so that we can learn effectively, efficiently, patiently, and fruitfully. The modern world today enables us to approach knowledge so quickly that we might rarely think of any roadblock to the learning process. However, the last two years tells us otherwise. COVID-19 pandemic has prevented millions of learners worldwide from knowledge acquisition. Many learners have suffered anxiety and depression from disconnection from knowledge. In this challenging situation, we have realized a fortunate. It is online learning. One of the powerful impact channels for learning is online learning environments (OLEs) used by millions of learners and thousands of educational institutions worldwide. For learners to progress in learning, it is required to grow the mindset and skill set of an active way of learning. Learning in OLEs requires learners to be active and autonomous because of the lack of contact with advisors, teachers, or instructors. Such an active mindset and skill set for learners is named self-regulated learning (SRL).

This research aims to support learners' recognization of their SRL ability in OLEs. We think that adequate support can be done by assisting learners in seeing their online learning history and then helping them understand their learning patterns. To achieve this objective, we highlight two main subjects of this research: the method for modeling online learning behaviors and the framework for explaining the model. We pave the studies on these two subjects with the following research questions.

- RQ1: What intrinsic and extrinsic factors construct and differentiate the SRL ability of a learner?
- RQ2: How can these SRL factors be identified and measured from a learner's learning history?
- RQ3: Under what cognitive or metacognitive conditions are individuals intrinsically/extrinsically motivated to self-regulate their learning?

- RQ4: By what signs can learners' learning history data in OLEs manifest SRL patterns?
- RQ5: How can learning history data from OLEs be synthesized for assessing the SRL ability of a learner?

The research outcome will be twofold:

- An SRL Recognition and Improvement framework (SRL framework for short) which is a source of reference for recognizing the SRL characteristics of a learner and measuring SRL ability;
- A learner Markov model structure for modeling SRL characteristics of learners in OLEs.

The SRL framework contains sound principles for describing SRL ability. The Learn Markov Model refers to both the model of a self-regulated learner and the method to generate the model from online learning data.

To build the SRL framework, we dated back to the basic principle of the mind and then reviewed existing SRL models popular and widely used to analyze their common and unique attributes and root principles. The purpose of the SRL framework is to give a source of explanation for SRL related activities, SRL ability, and SRL modification for improvement.

The learner Markov model is an application built from a reference to the SRL framework to support learners' recognition of their SRL patterns in an online learning context. We demonstrated a procedure to generate resource use sequences from learners' learning history data, suggested how to present SRL activities visually, introduced the SRL profile – a description of learners' SRL characteristics - and especially proposed the quantitative measurement of SRL ability – the SRL index.

The proposed method was applied to an open dataset from the Open University, one of the world's largest universities of online learning, for evaluation. We built prediction models to predict learner performance and compare the prediction results with current approaches to demonstrate the potential of our method. We also discuss the combination of the SRL framework with the SRL profile to support the understanding of learners' SRL.

Since the SRL framework and the proposed method for learner Markov modeling are newly introduced, there are several limitations to the validation of the framework, the specific approaches for SRL improvement, and the application of learning Markov modeling on other learning history data besides OULAD needed to justify its generalization. Such limitations call for future works to test the proposed framework and modeling method on other cases.

With the SRL framework as fundamental and the method for modeling SRL profile, we believe that the outcomes of our research help to ease the understanding of SRL and make the recognition of SRL concrete and the improvement of SRL approachable.

Keywords: Self-regulated learning, SRL profile, SRL ability, Online learning, Learner model.

Dedication

"You have accomplished all I have done." (Isaiah 26:12)

To my parents, who give me unconditional love and providence.
To my wife, who is willing to leave everything behind to walk with me and share the better and the worse with me until the end.

Acknowledgments

I heartily wish to thank my advisor, Professor Shinobu Hasegawa, for his devotion to guidance, support, and constructive advice throughout my journey of study and research. He has instructed me step-by-step, enlightened me, and kept supporting me, especially in the challenges of academic research. He has sponsored and supported me financially and academically. Without Hasegawa sensei, I could not reach this far in my study. Any achievements in my study that I have earned in Japan are enabled by Sensei. Thank you, Sensei, for everything.

I also wish to express my sincere gratitude to Professor Kiyoaki Shirai and Professor Minh Le Nguyen for your valuable feedback and advice on my research during the mid-term review process to steer my work in the right direction.

I would like to thank Professor Razvan Beuran and Professor Yasuo Tan for guiding me, easing my way through the minor research project, and giving me helpful advice on the project and my research journey.

I would like to thank you, Professor Takeshi Matsuda, Professor Minh Le Nguyen, Professor Kokolo Ikeda, and Professor Shogo Okada, for dedicating your valuable time and expertise to examining my thesis, giving me advice and instructions to correct the research to a better shape.

I would like to thank the Japan Advanced Institute of Science and Technology (JAIST) and the University of Science, Vietnam National University – Ho Chi Minh City (HCMUS-VNU) for supporting me during my learning journey. I would also like to thank the JAIST staff for helping me adapt to a new culture at JAIST and in Japan.

I would like to express my gratitude to all instructors from JAIST and HCMUS-VNU for your devotion to teaching me.

I appreciate the sponsorship from JAIST and Hasegawa-lab for my trips to the international conferences to strengthen my research works. I am also grateful to Research Center for Advanced Computing Infrastructure (RCACI) for providing me with infrastructure that supports my experiment.

From the bottom of my heart, I am grateful to my family for providing me with spiritual support during my time away from home. Knowing that my family is always with me, no matter how far I am away from home, I am motivated to strive forwards.

Thank you so much, my family, for your loving sacrifices during the hardship of the pandemic to take care of everything back home.

I am grateful and blessed to have my better half always by my side to share the better and the worse, the happy and the sad, the sweet and the bitter in the past years throughout the Ph.D. journey and the future to come. My wife is always my only best friend, mentor, and partner who supports my learning journey at JAIST. Without you, my better half, I do not think I have made it through this journey.

I would like to thank my friends from Hasegawa-lab for assisting me during my study.

Finally, to many others who have silently helped me have a joyful life with unforgettable experiences in Japan, I would like to thank you and keep all of the support in my heart.

Definitions, acronyms, and abbreviations

Clickstream	A sequence of accesses to online learning resources done via mouse clicks
Cognitive score	A performance score that a learner earns from completing an assignment or an exam to indicate their learning performance
Learner model	A partial representation of a learner from a particular aspect
Learning pattern	A repetitive arrangement of learning activities or learning resource access
Learner profile	A visual or written description of characteristics of a learner
Learning traces	Learning activity data stored in an OLE
LMM	Learner Markov model
LMS	Learning management system
Log SRL index	Logarithm to the base 2 of the SRL index
Markov chain	A process of states satisfying the Markov property, which is that a state in a process only depends on a state right before it and ignores all the other previous states.
Markov model	A model for an activity sequence, developed by applying Markov chain
Metacognitive score	A score that indicates a learner's awareness of his or her learning process.
MM	Markov model

OLE	Online learning environment
Online learning	The state of approaching and obtaining knowledge via the internet channel.
Online learning context	Online learning-related events, conditions, or environments
Online learning resource	Learning materials in multimedia forms
SR	Self-regulation
SRL	Self-regulated learning
SRLer	Self-regulated learners
SRL ability	A quality of execution of SRL capability
SRL capability	A learning skill or power that a self-regulated learner can perform
SRL character	A collection of characteristics or attributes that a self-regulated learner has
SRL characteristic	Marks for recognizing a self-regulated learners
SRL framework	A set of principles for the existence and operation of SRL
SRL index	A measurement of SRL ability
SRL profile	A description of a learner's SRL ability

Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1. Background, challenges, and opportunities for research

To know ourselves and seek knowledge are two everlasting desires of ours. That is why any brick walls on the road to knowledge cause us confusion and anxiety. During the last two years, COVID-19 has prevented millions of pupils and students worldwide from going to school, causing tremendous impacts on learning progress for individual learners. It might be the first time many learners feel depressed because of being blocked from learning. Fortunately, nowadays, paths to knowledge have been vastly and effectively supported by learning courses, materials, and visual and audio lectures delivered via the high-speed internet. That is online learning which is the state of approaching and gaining knowledge via the internet channel. Online learning supports our learning process beyond the time and space and learning material types constraints. Besides being an effective channel to transfer knowledge, online learning also poses the potential to assist learners in recognizing their way of learning.

To widen and deepen knowledge, we need to be aware of and understand our learning habits with their strengths and weaknesses. To aid the process of recognizing personal self-regulated learning capability, in this study, we propose a method for modeling online learning behaviors and a framework for explaining the model. The proposals strive to assist learners in discovering knowledge about their online learning patterns and supporting their awareness of their online learning habits.

Self-regulated learning (SRL) and online learning support are the two research areas that have recently attracted significant growth in research. SRL can be seen as both an active and proactive learning process or an active and proactive learning capability. Although SRL is innate, SRL ability varies from one learner to another. SRL ability refers to how effectively and efficiently learners regulate their learning journey toward a goal. Research in SRL has stretched from models for representing SRL processes to methods and tools for recording, measuring, and supporting SRL abilities [1]–[6]. One of our publications [7] empirically shows that SRL is a positive intertwinement of cognition and metacognition observable in online learning environments.

From the online learning perspective, learners get more engaged in online learning environments (OLEs), where there are conditions, multimedia materials, and services to support learning over the internet. Bernacki et al. [5] illustrated the compatibility of OLEs and SRL in which the organization of learning materials in OLEs helps learners to self-regulate their learning, and learners with SRL ability can learn effectively in OLEs. OLEs preserve data of learners' interaction with online learning resources - a valuable source of feedback about learners' learning processes. Various studies have analyzed such a type of history data for indicators about how learners have used learning resources to predict future learning performance, potential dropouts, or learners at risk of failing so that educators and instructors can offer the learners with necessary support [8]–[11]. Although gaining high accuracy in such a prediction, current research encounters challenges in providing a rational explanation for the correlation between online learning traces and learning habits. Hence, it is not helpful to assist learners in building good learning habits to improve learning performance.

Current researches also tend to support SRL rather than first helping learners understand their SRL characteristics and abilities. We believe that such an understanding is a prerequisite for adequate support that follows. Understanding one's learning habits is a source of proper regulation to improve learning performance. In modern days, one's learning process with the support of OLEs such as learning management systems (LMS) or massive online open courses (MOOCs) leaves various learning traces informative for analysis to gain knowledge about one's learning habits. Such data open opportunities to know learners' way of learning somehow and support their improvement.

1.2. Research objectives and research questions

With the purpose of supporting learners' recognition of their SRL from their learning behavior data, we have started this research. To understand SRL, it is necessary for us to describe and measure it. Thus, we focus this research on a single objective that is to describe and measure learners' SRL ability in OLEs.

To achieve this objective, we highlight two main subjects of this research: the method for modeling online learning behaviors and the framework for explaining the model. We pave the studies on these two subjects with the following research questions.

- RQ1: What intrinsic and extrinsic factors construct and differentiate the SRL ability of a learner?
- RQ2: How can these SRL factors be identified and measured from a learner's learning history?
- RQ3: Under what cognitive and metacognitive conditions are individuals intrinsically/extrinsically motivated to self-regulate their learning?
- RQ4: By what signs can learners' learning history data in OLEs manifest SRL patterns?
- RQ5: How can learning history data from OLEs be synthesized for assessing the SRL ability of a learner?

The research outcome will be fourfold:

- the SRL Recognition and Improvement framework, which is a source of reference for recognizing the SRL characteristics of a learner and measuring SRL ability,
- the learner Markov model, which is generated by the application of the Markov chain to model learnerts' online learning patterns,
- the self-regulated learning (SRL) profile, which is the presentation of the model to support learners' understanding of their learning patterns,
- and the SRL index, which is a scalar measurement of the effectiveness of the learning patterns.

The innovative points of this research are highlighted in the following two contributions. First is the structure used for modeling online learning behavior data. And second is the principles on which the modeled data are explained. Current research usually analyzes attributes of online learning behaviors separately or without academic principles to relate such attributes to one another; therefore, the analysis outcome does not reflect actual learning habits. Proposing a new modeling structure and principles for explaining observation data, this research strives to give online learning behaviors data a reliable description, which is helpful for learners to understand their learning habits.

1.3. Dissertation structure

Chapter 1. Introduction

This chapter strives for readers' interest in the research topic of SRL and appropriate support for self-regulated learners (SRLers) in online learning. In this chapter, the research background, motivation, challenges, and opportunities for the authors to carry out this research are explained. The research objectives, research questions to be addressed, and the expected outcome are stated in this chapter. This chapter also describes the structure of the dissertation and a guideline for reading the document.

Chapter 2. Literature review

This chapter reviews research on SRL and online learning. Firstly, it presents the idea of SRL and the outcomes of SRL research relating to traditional education and online learning. The review strives to justify the important role of SRL for self-study in online learning contexts, presents available SRL-related models that can be applied to supporting learners in online learning contexts, and points out the gaps and opportunities for improvement which this research would fill in.

Then, the chapter describes the characteristics of self-study in online learning by demonstrating the correlation between SRL and online learning, analyzes data on online learning behaviors, and presents current supports for online learning and opportunities for developing ideas and methods to support learners further in online learning.

The chapter also briefly reviews studies in learning analytics and educational data mining, learning tactics and strategies, and asynchronous and synchronous online learning to scope this research in an appropriate expertise and application".

Chapter 3. SRL Recognition and Improvement framework

This chapter introduces and describes one of the outcomes of this research, the SRL recognition and improvement framework. The framework demonstrates how SRL in online learning contexts should be formed and measured, what SRL characteristics are, and how they can be improved. This chapter demonstrates the manifestation of SRL in online learning contexts, which is written in chapter 2, and then mentions a need for modeling online learning-related data for further application, which will be discussed in Chapter 4.

Chapter 4. Formulation of SRL related problems in online learning contexts

This chapter formulates SRL-related problems and presents the use of techniques and theories in information science to model online learning data. Markov chain and related algorithms are the main approaches for analyzing and manipulating online learning data for modeling SRL and later supporting SRL in online learning contexts. In this chapter, the Markov chain is briefly reviewed and followed by the application of these techniques to this research.

Specifically, this chapter formulates the problem of modeling a self-regulated learner in an online learning context, develops a Markov model (MM) for a learner, shows how the parameters of a learner Markov model (LMM) are estimated and how steady states of the LMM, representing online learning habits of learners, are calculated.

From LMM, this chapter presents the SRL profile – a description of an SRLer in OLEs – and the SRL index – measurement of SRL ability. Together with the SRL recognition and improvement framework, the SRL profile and the SRL index are the primary outcomes of this research. They are used to describe and measure SRL ability.

Chapter 5. Method evaluation and discussion

This chapter presents the application of the proposed method to an open dataset named OULAD [12] in order to model SRLers and learners' SRL ability and develop learning performance prediction models. The proposed method is compared with existing approaches to evaluate its effectiveness and rationale.

Chapter 6. Conclusion, limitations, and future works

Chapter 6 summarizes the dissertation contents, distills theories and fundamentals used to develop the research, highlights outcomes and achievements, points out limitations, and recommends future research and development.

Chapter 2. Literature review

We tend to begin changes after we have understood the reasons for change. In online learning contexts, learners leave behind redundancy of learning history data. There are studies and applications about using online learning history data to assist learners; however, it seems that the learners could not make proper changes for improvement. We think this is because current approaches might not have the data made sense to the learners. To understand their SRL from their learning history data, learners need to have their learning history data arranged and organized in a certain manner and an underlying set of principles to explain the data. With such goals in mind, we would like to conduct a literature review on studies in SRL and studies and applications from the analysis of online learning data.

We will journey into the formation of theory about self-regulation, self-regulated learning (SRL), and the development of SRL research, specifically SRL models, to illustrate the differences and similarities of the SRL models and the need for an SRL framework. After reviewing research in SRL, we would like to go into current research in online learning data analysis with existing results and outcomes for supporting learners. We also present recent research on SRL in online learning, challenges, and opportunities in assessing SRL ability and helping learners self-regulate their learning.

2.1. SRL models

This section is a slightly modified version of the authors' article [13] published in 2021 and has been reproduced here with the copyright holder's permission.

2.1.1. Self-regulated learning

We find it necessary to comprehend what SRL is, how it operates, and why it is worth studying; therefore, we would like to start the literature review by presenting the understanding of SRL from leading researchers in this field of research.

Self-regulation (SR) has become an attractive research topic since the 1980s, and studies on self-regulation, specifically on SRL, have grown significantly since the 2000s. Despite its recent attraction for research, SR was thought about in ancient times and is known as the human's ability of self-consciousness [14]. What is self-regulation?

According to the Oxford English Dictionary, to regulate is to control things so that they behave properly. The Latin origination of the verb regulate means to rule and direct. SR is a state of adjusting, ruling, and directing oneself in a certain activity or process, or procedure.

Theories of SR have been developed, each of which observes SR from different perspectives. The theories are significantly applied to SRL. There are theories formulating SRL as a goal-directed process. Other theories view SRL as a series of reciprocal interactions between a person and his surroundings when performing a task. Still, other theories center SRL around personal self-awareness. Because of various aspects of SR, definitions of SRL also vary.

According to professor Boekaerts [15], SRL is difficult to define. There are indeed more descriptions of SRL than definitions of it. She views self-regulated learners as ones "*who have the capacity to exert control over different dimensions of the learning process… and to allocate resources to the different aspects of the learning process…*" [15, p. 102]. From a goal-oriented perspective, professor Efklides describes SRL as a learning path in which learners set attainable goals corresponding to their "*cognition, metacognition, motivation, affect and volition*" [16, p. 1]. Professor Winne [17, p. 533] sees SRL as a metacognitively guided, intrinsically motivated, and strategic form of learning. The apparent definition of SRL so far is from Professor Zimmerman. Professor Zimmerman defines that "*Self-regulated learning involves metacognitive, motivational, and behavioral processes that are personally initiated to acquire knowledge and skill, such as goal setting, planning, learning strategies, self-reinforcement, self-recording, and self-instruction*" [18, p. 541].

SRL is like a way of learning comprised of components and processes in which the components interact under conscious supervision and regulation.

2.1.2. SRL models

The outstanding achievements of research on SRL are the SRL models, each of which describes the operation of SRL at individual learners from certain specific viewpoints. Panadero [19] described, analyzed, and compared several popular SRL models to the extent of their underlying theories, processes, and empirical evidence about the application and associated measuring tools. In the following, six popular SRL models are reviewed to analyze the viewpoints from which the models are constructed.

2.1.2.1. Winne's model of SRL

Professor Phillip H. Winne's research on SRL provides a view of SRL from a metacognitive perspective demonstrated in his model [20]. Winne and Hadwin's model [20], introduced in 1996, emphasized the role of metacognition in the self-regulation of cognitive tactics and strategies. As shown in Figure 2.1, this model demonstrates a 2-phase SRL process to accomplish a learning task. The first phase is planning, and the second phase is executing the plan, monitoring the progress, and making the adaptation. Though sharing the same SRL patterns, individual learners' SRL ability differs in five points; they are *(i)* domain knowledge that the individual has accumulated from their educational background and history, *(ii)* knowledge of tactics and strategies, which is a reservoir of learning methods and techniques, *(iii)* performance of tactics and strategies which are the proficiency of applying learning techniques, *(iv)* regulation of tactics and strategies that monitor how well ones learn and make appropriate adaptations, and *(v)* global dispositions which are pathways to learn.

How do learners address a task according to the SRL model? The first phase starts with a learner receiving a task to be addressed. The learner uses cues for the task, recalls domain knowledge related to the task, assembles strategy knowledge required to perform the task, and motivates herself to address the task. The outcome of the first phase could be a set of goals with their profiles and standards, which clarify the results from doing the tasks and against which the learner judges the quality of the results. When the plan is ready, the second phase begins to operate. In this phase, the learner applies cognitive strategies and tactics to act on her plan. There would be several outcomes or products generated. There would also be indicators of the current state of the task in comparison to the plan. The learner monitors these pieces of information frequently and uses them as feedback on her plan and current goals and activities to make appropriate changes.

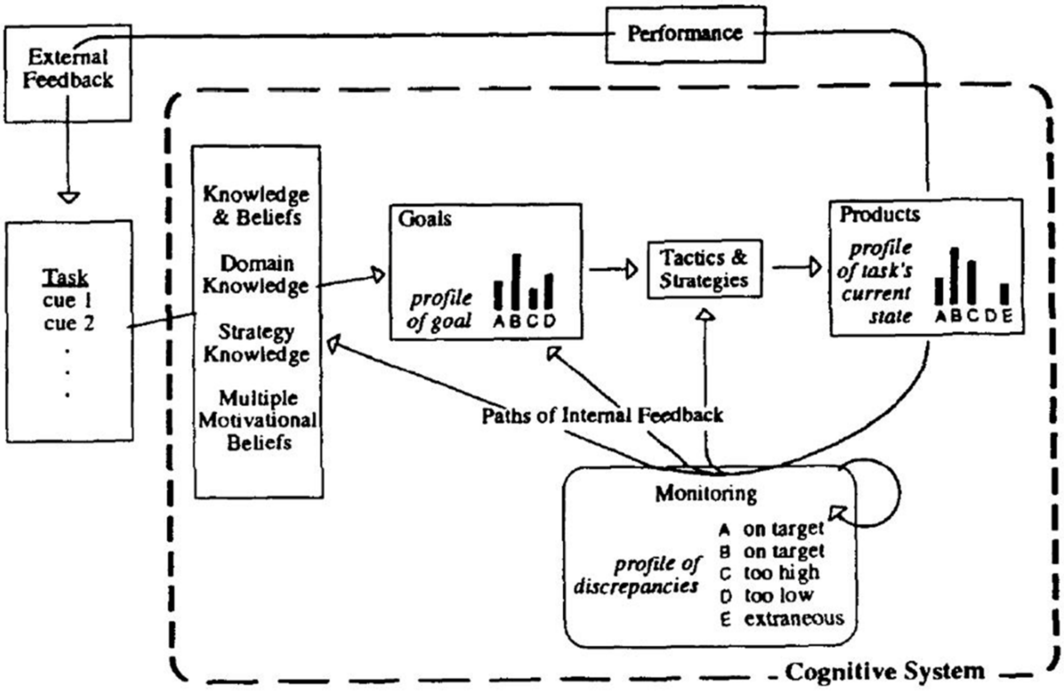


Figure 2.1. Winne's SRL model [20, Fig. 1]. The model contains 2 phases. The first phase is planning; it starts with a learner receiving a task, then using the task's cues and reviewing her knowledge, learning strategies, motivation, and beliefs to form a plan and goals. The second phase is executing the plan, monitoring the progress, and making the adaptation by comparing the achieved products and the planned.

2.1.2.2. Boekaerts' dual-processing model and six-component model

Professor Boekaerts' research on SRL mainly investigates the role of goals of different types in SRL [21]. In 1996, Boekaerts introduced two SRL models, the six-component model of SRL and the dual processing model.

In her dual-processing model (as shown in Figure 2.2), the SRL pattern is determined by a learner's selection of goals; there are two main pathways of goals: the growth of knowledge, skills, and the well-being of self-esteem. Depending on the level between those pathways, learners will gather, align resources, and self-regulate their learning to balance learning performance and self-esteem [22].

How does a learner self-regulate according to the dual-processing model described in Figure 2.2? Having a learning task in front of her, a learner has a mix of two types of goals: gaining knowledge and skills, and protection of ego. The former goal strives to increase understanding; the latter goal preserves self-esteem. The learner tends to apply learning strategies to master the learning task when aiming to the goal of knowledge gain while she might use coping strategies to learn or deal with the task so that it could not harm her self-esteem, e.g., not to have low grades compared to peers. Depending on the balance between two types of goals, learners express different SRL patterns.

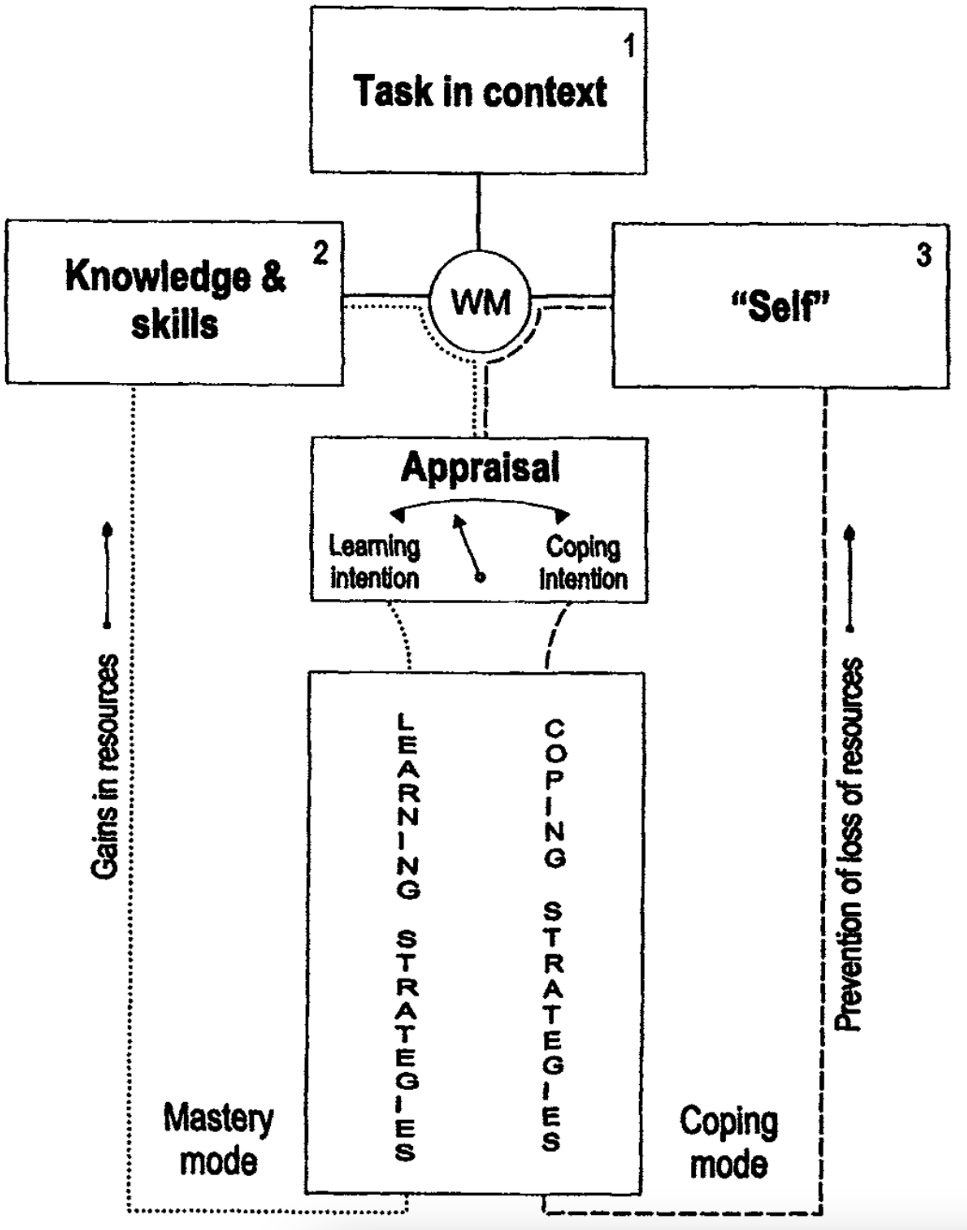


Figure 2.2. Boekaerts' dual processing model [21, Fig. 1]. The model demonstrates 2 types of goals from which a learner will choose when learning a task. The knowledge & skills goal is to gain knowledge, which activates the mastery mode of learning. The "self" goal is to protect self-esteem, which activates the coping mode of learning.

Boekaerts' six-component model of SRL views SRL as the interoperation of cognition and motivation throughout the aspects of goal setting, strategy use, and domain knowledge [15]. In this model, cognition and motivation function simultaneously when self-regulated learners set goals, prepare cognitive and motivational strategies, and recall prior knowledge to learn new knowledge effectively.

How does a learner self-regulate her learning according to the six-component model shown in Figure 2.3? A learner regulates her cognition and motivation in learning, and these two faculties manifest in three areas in learning: (accumulating) domain-specific knowledge, (applying) strategies, and (setting) goals. The combination of the two faculties and the three areas reveals in the six components, as demonstrated in Figure 2.3. According to the model, learners possess 2 types of knowledge: domain knowledge about their expertise (block 1, 2, and 3 accordingly in the model) and metacognitive knowledge about their beliefs and motivation (block 4, 5, and 6 accordingly in the model). Besides the knowledge, learners have abilities and skills to apply it to tasks to a certain extent and regulate the level of application of cognitive strategies according to goals, the learners' intention, and obstacles they face. Overall, the level of regulation corresponds to the balance of two types of goals presented in the dual processing model above.

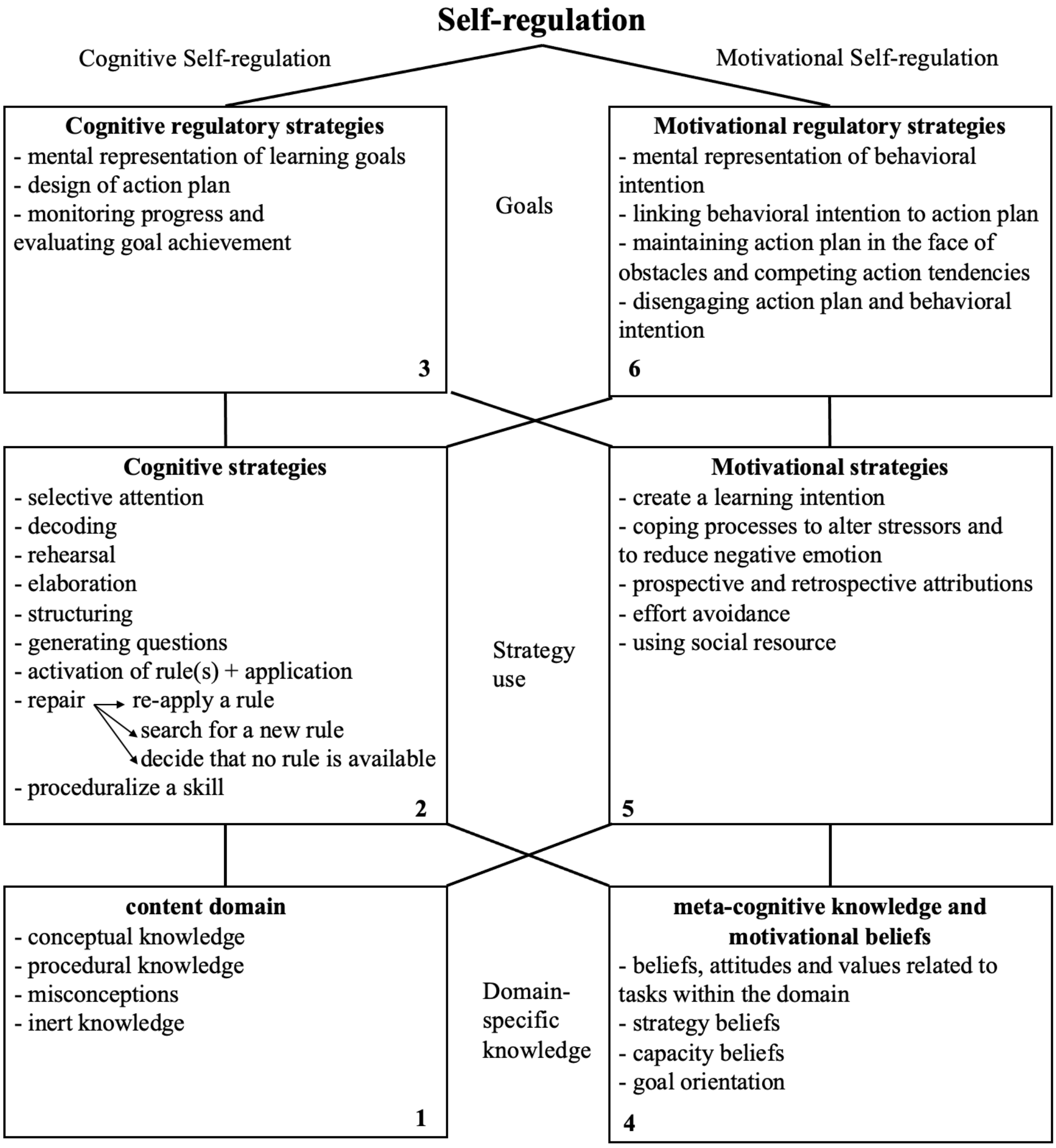


Figure 2.3. Redrawn from Boekaerts' six-component model of SRL [15, Fig. 1]. A learner regulates her cognition and motivation in learning, and these two faculties manifest in three areas in learning: (accumulating) domain-specific knowledge, (applying) strategies, and (setting) goals. The six components reveal the combination of the two faculties and the three areas.

2.1.2.3. Pintrich's framework of phases and areas for SRL

Professor Pintrich's research concerns the role of goal and its effects on motivation which then choreographs cognitive processes. His research also focuses on metacognitive awareness as the principle for motivation. These main points are described in his framework of phases and areas for SRL in 2000 [23]. The framework presents the common attributes shared by all existing SRL models, which are the following:

- Self-regulated learners are active in terms of setting learning goals, reviewing prior knowledge, preparing cognitive strategies and learning environment for their learning process,
- Self-regulated learners have the potential to monitor, control, and regulate internal and external factors of the learning process,
- All SRL models have criteria against which self-regulated learners reflect their progress in order to adjust their learning progress,
- Self-regulatory activities are the means that self-regulated learners apply to reach their learning goals [23].

The framework shown in Figure 2.4 comprises two dimensions. First is a system of four SRL phases: forethought planning and activation, monitoring, control, reaction, and reflection. And second is regulation activities in each phase in the areas of learners' cognition, motivation, behavior, and context. Running throughout the framework and joining self-regulatory activities are learning goals and motivations [23]. According to the framework, when learners self-regulatedly learn or perform a task, they generally follow a 4-phase SRL process chronologically though it is not strictly so. In the Forethought planning and activation phase, the learners apply Ease of Learning Judgments (EOLs) to evaluate the difficulty of the task to be learned, prepare a context, prior knowledge, motivation for learning, target goals, and an activation plan. Then in the monitoring and controlling phases, the learners act the plan, aware of knowledge absorption via Feeling of Knowings (FOKs), monitor the learning progress via Judgments of Learnings (JOLs), and make appropriate adaptions or changes to learning methods, e.g., adjusting time, seeking help, alternating techniques. Finally, in the Reaction and reflection phase, the learners judge their performance, review the effectiveness of their cognitive work, identify attributions to the achievements, and evaluate the process.

As mentioned above, Pintrich's research focuses on metacognitive awareness. It is illustrated in the Feelings of Knowings (FOKs), Judgments of Learnings (JOLs), and Ease of Learning Judgments (EOLs). FOKs refer to learners' confidence in recalling knowledge, although they do not remember it at the moment [23, p. 459]. JOLs are activities the learners do to help them understand what they do not understand at the moment [23, p. 459]. EOLs are activities the learners perform to clear tasks before they learn [23, p.462]. The better the learners can do FOKs, JOLs, and EOLs, the higher they are motivated to learn.

Phases	Areas for regulation			
	Cognition	**Motivation/affect**	**Behavior**	**Context**
1. Forethought, planning, and activation	Target goal setting Prior content knowledge activation Metacognitive knowledge activation	Goal orientation adoption Efficacy judgments Ease of learning judgments (EOLs); perceptions of task difficulty Task value activation Interest activation	[Time and effort planning] [Planning for self-observations of behavior]	[Perceptions of task] [Perceptions of context]
2. Monitoring	Metacognitive awareness and monitoring of cognition (FOKs, JOLs)	Awareness and monitoring of motivation and affect	Awareness and monitoring of effort, time use, and need for help	Monitoring changing task and context conditions
3. Control	Selection and adaptation of cognitive strategies for learning, thinking	Selection and adaptation of strategies for managing motivation and affect	Self-observation of behavior Increase / decrease effort Persist, give up Help-seeking behavior	Change or renegotiate task Change or leave the context
4. Reaction and reflection	Cognitive judgments Attributions	Affective reactions Attributions	Choice behavior	Evaluation of task Evaluation of context

Figure 2.4. Redrawn from Pintrich's Phases and Areas of SRL [23, Table 1]. Two dimensions of the framework are the SRL phases and the areas for regulation. First is a system of four SRL phases: forethought planning and activation, monitoring, control, reaction, and reflection. And second is regulation activities in each phase in the areas of learners' cognition, motivation, behavior, and context. Running throughout the framework and joining self-regulatory activities are learning goals and motivations.

2.1.2.4. Zimmerman's cyclical phase model

Professor Zimmerman is one of the pioneer SRL researchers and mainly bases his SRL models on professor Albert Bandura's well-known socio-cognitive theory [24].

Viewing self-regulation as a result of the intertwinement of an individual's consciousness, behaviors, and the environment where they are working on a particular task, The cyclical phase model emphasizes the process aspect of SRL. It illustrates the interaction paths between learners, learning tasks, and the learning environment in a specific context defined by learning contents and environment settings.

Among Zimmerman's SRL models, the most popular is the cyclical phase model, first introduced in 1998 and then added with detailed subprocesses for each phase in 2000, as shown in Figure 2.5. The model demonstrates that individuals self-regulate their learning via a 3-phase process [25]. Detailed descriptions of the subprocesses of each phase are found in [26]. The process starts with the forethought phase, in which learners begin their learning journey by analyzing learning tasks, setting learning goals, planning cognitive strategies, and motivating themselves to learn. Then, the learners proceed to the performance phase, where they put their learning plan into action with conscious self-control over how they learn and a self-observation of how well they have been learning. Finally, the learners wrap up their learning with the self-reflection phase, in which they judge the learning journey by comparing the learning performance against the goals set in the first phase, analyzing factors that contribute to learning achievements, and in which they seek adjustments and alternative approaches to help them learn more effectively and productively.

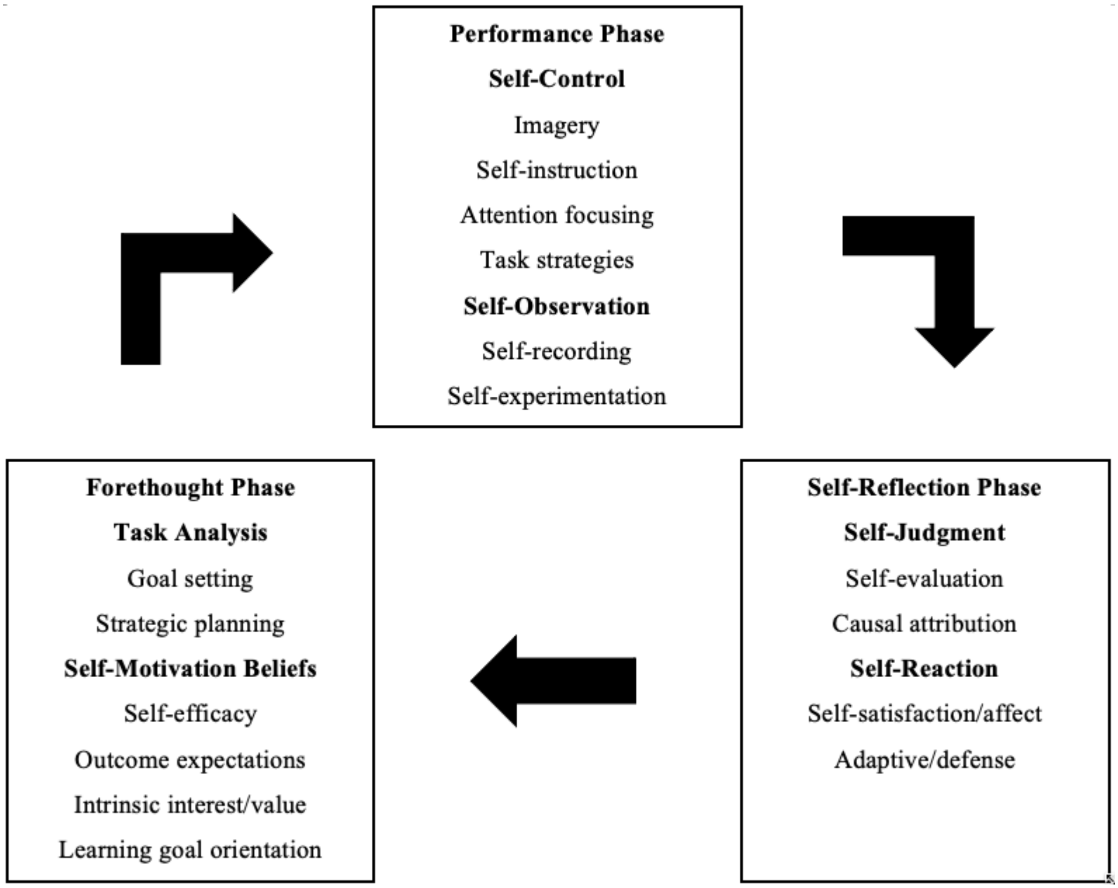


Figure 2.5. Redrawn from Zimmerman's SRL cyclical model with phases and subprocesses [26, Fig. 1]. The cycle arrows indicate the process does not run one time only but repeats as learners progress their learning.

2.1.2.5. Efklides' Metacognitive and Affective model of SRL

One of the latest SRL models is professor Efklides' Metacognitive and Affective model SRL (MASRL), introduced in 2011. Professor Efklides views SRL as a composition of 3 main components: metacognition, motivation, and affect [16], and this viewpoint is illustrated in the MASRL model. The model presents the interaction of metacognition, motivation, and affect in the SRL process when an individual learns specific tasks. The interaction operates within a learner and between the learner and the tasks.

Figure 2.6 demonstrates the MASRL model. Efklides [16] demonstrates that an individual's SRL manifests at two levels. One is the Person level, which is a general SRL level or about SRL characteristics of an individual regardless of learning contents or context. And the other is Task x Person level, which is about the ability of an individual to apply specific SRL behaviors within a particular learning task. When individuals following the MASRL model engage in a learning task, their Person level sets learning goals and establishes top-down self-regulation based on their metacognitive knowledge, metacognitive experiences, and metacognitive skills. Those metacognitive strategies have been accumulated and built into the learners' SRL traits. In the Task x Person level, their cognitive strategies are regulated in a bottom-up self-regulation manner to meet the task requirements and reorganize the Person level.

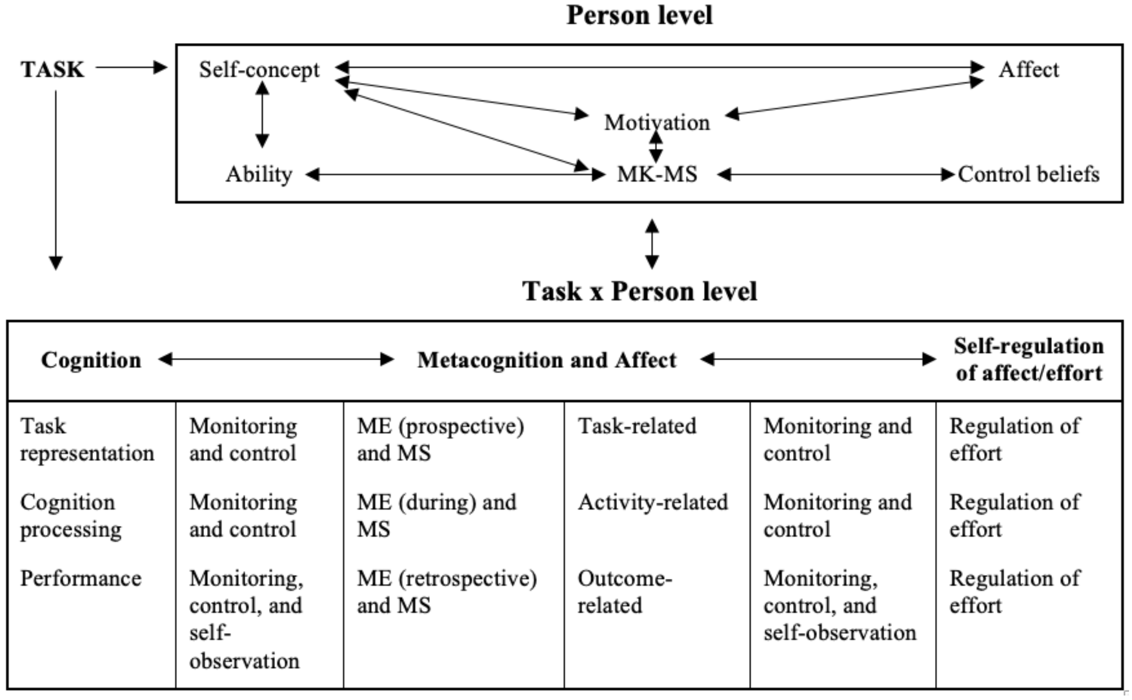


Figure 2.6. Redrawn from Efklides' MASRL model [16, Fig. 1]. When addressing a task, a learner starts a cognitive and metacognitive preparation at the Person level, which represents her general characteristics and ability of SRL independent of the task. Then, she regulates her SRL in response to the specific requirements of the task in the Task x Person level.

How does a learner self-regulate her learning according to MASRL? When addressing a task, a learner prepares herself at the Person level first. At this level, the learner leverages her metacognitive knowledge (MK) and metacognitive experiences (ME) to establish the ability and beliefs to work on the task. She also leverages motivation and affect and the ability to establish a self-concept, which is a '*representation of one's competence in various domains*' [16, p. 6] so that she is motivated and feels confident in doing the task. Then, the learner starts working on the task, transitting SRL to Task x Person level. At this level, the learner's cognition processes the tasks while her metacognition and affect monitor the progress of the cognition so that she can self-regulate her effort on the task and her affect to stay motivated. At this level, task, activity, and outcome-related events are considered by metacognition and affect so that proper self-regulation of affect or effect can be made. Task level and Task x Person level also interact to produce appropriate adjustments according to the task progress. For instance, a task more difficult than expected might make the learner cease doing to prepare extra cognitive strategies needed for fulfilling the task. Or, if the work progresses more than expected, the learner would feel more motivated and confident.

2.1.3. Models comparison

2.1.3.1. SRL Phases

From the phase division perspective, despite containing different processes, the models mentioned above share three common phases: planning, performance, and reflection. Standing in front of knowledge, we go through these phases naturally to obtain knowledge. The more sophisticated the knowledge we are approaching, the more conscious we are of our application of these phases. We simply cannot grasp sophisticated knowledge all at once because of limitations of our time, prior knowledge, ability, maturity, etc.

To embrace new subject-matter of sophisticated knowledge, firstly, we usually observe it from different perspectives, wonder about the breadth and depth of each perspective of the subject-matter we desire to explore and relate it to our prior knowledge to make a connection by distinguishing and similarizing it to what we have known, determine in mind or clearly on paper a certain degree of outcomes, objectives about the subject-matter to strive for and sketch in mind or clearly on paper a course of activities to review, study, experiment the subject-matter. In other words, firstly we plan.

Next comes the performance phase, where we review, study, experiment, and analyze the subject-matter to grasp its knowledge and accomplish the outcomes and objectives that we have established for the subject-matter. It is in the performance phase that we actually discover the breadth and depth of the subject-matter, that we contact the real condition of the subject-matter as it is rather than as we think it is. For example, we might have experienced that we think we can write a good paragraph about a certain idea, but it turns out that the paragraph we write is not as cohesive as an outline of a paragraph that we have in mind. When we produce outcomes in the performance phase, the reflection phase begins.

The reflection phase compares the performed outcomes and the planned ones in order that modifications and adjustments might be made. The items to be compared are various such as objectives, goals, learning tactics and strategies, and levels of growth of knowledge to be digested.

Although described separately and successively in order, the three SRL phases interweave in operation. We rarely delay the performance phase until we finish the planning phase, and we usually reflect as soon as we have the outcomes from the performance phase and the planned outcomes from the planning phase. It seems that there are multiple planning-performance-reflection cycles running when a self-regulated learner learns a subject-matter from the starting point until the finishing point; therefore, we argue that the three SRL phases interoperate throughout a learning process with each SRL phase presenting a dominant density in a particular period. Hence, a timeframe in the learning history of a learner is probable to tell his or her SRL characteristics and ability.

2.1.3.2. SRL elements

From the models mentioned earlier, each of them, on the one hand, describes common overall phases and, on the other hand, presses SRL on specific angles from process orientation to components orientation, from cognition to metacognition, from goal to motivation, affect, emotion, or ego.

Knowing the meanings of these main components is necessary to understand SRL. Table 2.1 summarizes the main components and their associations that each of the six models above focuses on. We extract six main components that repeatedly occur in the six models above. The components are cognition, metacognition, goal, ego, motivation, and affect.

Table 2.1. Remarked associations between components in current SRL models.

Association	Cognition	Meta-cognition	Goal	Ego	Motivation	Affect
Cognition		1, 2	2, 3, 4	2	3, 4, 5	4, 5
Meta-cognition			2, 3, 4	2, 3	6	6
Goal				2	3	
Ego					2	
Motivation						4, 6
Affect						

1: Winne's SRL model [20]; **2:** Boekaerts' dual-processing model [21];

3: Boekaerts' six-component model [15]; **4:** Pintrich's framework of phases and areas for SRL [23];

5: Zimmerman's cyclical phase model [26]; **6:** Efklides's MASRL [16].

Cognition

Winne presented cognition in terms of cognitive strategies and tactics. They are a set of how-to-do and a course of action in certain contexts [20, p. 328]. Winne modeled cognition in a compact, elegant manner as IF contexts THEN how-to-do. Boekaerts referred to cognition as '*cognitive processes and behavior that students use during actual learning experiences to accomplish a goal implied by the academic task*' [15, p. 105]. Pintrich described cognition as the use of content knowledge and strategic knowledge for learning or working on a task [23, p. 455]. Zimmerman expresses cognition in the performance phase of his SRL cyclical model as methods of visual description, attention, self-instruction, and task strategies to gain knowledge [26, p. 68]. According to Efklides, cognition manifests as "*capabilities (ability, knowledge, skills) or competencies*" of learners [16, p. 6].

Metacognition

Winne followed the 2-level model of metacognition, describing metacognition as monitoring and controlling cognitive behaviors [20, p. 329]. A learning process involves two levels: meta-level and object-level. The object-level relates to information about the status, learning progress, the effectiveness of learning strategies applied, and so forth. The meta-level involves the observation on the object-level so that proper adjustments can be made. Researchers have also shared a similar understanding and definition of metacognition, namely, metacognition as one's awareness of and knowledge about one's cognitive thought, strategies, behaviors, and so forth [16, p. 6], [23, p. 462], [26, p. 65].

Motivation

Elfklides views motivation as an orientation toward the achievement of goals [16, p. 6]. Boekaerts expresses a holistic view of motivation. According to Boekaerts, motivation comprises "*beliefs, judgments, and values related to one's capacity*" in response to a learning task and *"strategies to do what is necessary to achieve mastery, complete a learning task, or accomplish a learning goal"* [15, p. 108].

Affect

Elfkides views affect as a combination of attitudes and emotions [16, p. 6], while Zimmerman relates positive affect with self-satisfaction corresponding to learning performance. Pintrich views affect and motivation going together, generated by self-efficacy [23].

Goal and Ego

Besides the general meanings of goals we accept in common sense, according to Pintrich [23], goals are criteria and standards against which performance on a task is compared. Boekearts [22] categorizes goals into two types: growth of knowledge and skills and preservation of ego. Choosing the goal of knowledge growth leads learners to mastery learning mode while choosing the other type of goals makes the learners cope with learning tasks to defend their self-esteem.

Let us review all of the models above. All the models share a pattern that SRLers perform learning activities with a certain level of awareness over the learning process and with a vague to concrete target, and the learning process does not end without thought or desire of improvement or learning better in the future. The models demonstrate how individual learners self-regulate their learning but have not fundamentally explained why such an SRL process can lead to learning efficiency. Furthermore, starting from a specific perspective, the models might not provide a comprehensive ground on which SRL ability is evaluated. It requires a recognition of the principles on which SRL stands and operates. Such principles would be a fundamental reference point for learners to understand their SRL, recognize the strengths and weaknesses of the SRL characteristics, and improve their SRL ability. The quest for the SRL principles and the development of the SRL recognition and improvement framework will be presented in chapter three.

Having reviewed current research on SRL, we would like to shift your attention to the research on online learning area. Online learning are diverse to the extent of content delivery modes, resource variety, data history, data analysis approaches, etc. Our research objective concerns the expression of SRL via learning activity data in OLEs; therefore, we intent to scope our research through the lens of three aspects: content delivery mode, data analysis approach, and data history. Online learning is currently delivered in two primary modes, asynchronous and synchronous. Online learning data that are generated by learners are tremendous and enable productive analysis. Among various types of data history, the most popular is clickstream. Recently, online learning data analysis has advanced so rapidly that it has formed research fields, namely, learning analytics and educational data mining. Let us view each of the three aspects in the following sections.

2.2. Asynchronous mode and synchronous mode of online learning

Nowadays, OLEs are capable of both the synchronous mode of learning, in which teaching, learning, and communication activities are carried out in real-time and the asynchronous mode, in which learners use learning materials at their preferrable time. In her Ph.D. dissertation, Rockinson-Szaphiw [27] suggests the effectiveness of these online learning modes be estimated in the light of three types of presence: social presence, and cognitive presence, teaching presence. Social presence concerns the degree of the existence of community and communication that an online learning mode can give the participants. Cognitive presence concerns the degree of the existence of knowledge obtainment that an online learning mode supports learners. Teaching presence concerns the degree of existence of facilitation, instruction, and guidance enabled by an online learning mode [27, pp. 5–6].

In current OLEs, these three types of presence are offered in both asynchronous and synchronous modes varying from one OLE to the others in favor of one mode to the other.

Examples of online learning resources used in both synchronous and asynchronous modes would be forums and messages to support social presence. Resources mainly offered in the asynchronous mode would be online learning materials such as ebooks and videos to support cognitive presence. For the synchronous mode, livestream lectures are offered to support teaching presence. And for the asynchronous mode, feedback is usually offered to support teaching presence. Though having impacts on all three types of presence, the synchronous and asynchronous modes of online learning tend to favor social presence and cognitive presence, respectively. The synchronous mode enables real-time communication; thus, it supports learning contexts that encourage group work, teamwork, and peer interactions [28]. The asynchronous mode does not constrain learners to a fixed time but gives them autonomy on resource usage; thus, it enables them to self-regulate learning according to their learning conditions. We think asynchronous learning history data reflect self-regulated learning; hence, an online resource use model generated from such data can help learners recall their learning patterns profoundly. Therefore, in this research, we will focus on developing learner models out of learning history data of asynchronous type.

2.3. Learning analytics and Educational data mining

Online learning, with its massive amount of learning data generated, calls for not merely activities or techniques to process it but dedicated fields of learning analytics (LA) or educational data mining (EDM) to make learning data informative, useful, and beneficial for learning processes. LA and EDM are often used interchangeably in terms of methods and techniques for exploring data from such data sources as OLEs or LMS, or MOOCs for insights into learning, teaching, and educational environments and then using such insights for making adjustments or augmentations, or enhancements of learning and teaching. There is no clear distinction between LA and EDM; if the boundary between the two is considered, it is considered differently from study to study [29].

LA is defined as "*the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs*" [30, p. 3]. These activities on data have been and still are focal research subjects for LA. Romeo and Ventura [31] describe EDM as processing available educational data to discover patterns, rules, and knowledge hidden in the data in order to develop applications. The borderlines between LA and EDM are often drawn from these perspectives.

By their very names, LA analyzes learning-related data to investigate learning-related matters while EDM analyzes education-related data to investigate education-related matters such as teaching, learning, programs, policies, etc. In LA, the learning-related data might include learners' learning history and demographics, course content and materials. The learning-related matters might be learning progress, assessment scores, material usage, learners' interactions with peers and teachers and OLEs. In EDM, the education-related data might include a wide range of data about students, academic programs and curriculum, assessment rubrics, examinations and quizzes, and so on [32].

By their definitions mentioned above, EDM studies methods for seeking patterns, knowledge and rules underneath data and while LA develops methods, techniques, and models to synthesize, describe, measure, and present insight from data. Regardless of different definitions, according to the latest survey by Romero and Ventura [32], there are overlaps between EDM and LA in their objectives, methods, and data.

To draw a distinct border between EDM and LA, we refer to their definitions. LA focuses on supporting the learning process of learners while EDM has the purpose of exploring education-wide data to support various parties in the education system. Because this research aims to assist learners in understanding their learning strengths and weaknesses to make adjustments for improvement, we scope this research into the learning analytics field.

2.4. Analysis of online learning data

2.4.1. Clickstream data

Clickstreams are probably the most popular data type that users of online services leave behind as online traces when using online services. The analysis of clickstream data has been performed in various areas [33]–[35]. Online learning clickstream data mainly contain mouse clicks on learning resources, who click the resources and timestamps when they happen. In online learning, clickstream data have attracted tremendous studies and presented interesting insights into online learning behaviors.

A large number of current research focus on clickstream data to predict students' performance in online learning contexts. The research [36]–[38] applies different approaches to the same open dataset [12] to predict students at risk of dropping out of online courses. To predict whether a student faces a dropout risk, research [36] uses a series of a student's daily number of clicks on resources in a course as a feature list to train machine learning prediction models; research [38] uses an accumulative sum of clicks per quarter for similar objectives while research [37] develops and uses an activity-time-assessment graph describing a student performing certain activities on certain dates given certain assessments ahead. Analyses of clickstream data to describe self-regulated learning are on a growing path. Through a review of four recent studies about clickstream and SRL, Baker et al. [10] illustrate what information from clickstream data helps unravel how students manage time or procrastinate in order to predict their performance. Li et al [8] describe how timestamp differences among clickstream data can measure a student's regulation and time management effectiveness.

2.4.2. Currents applications from clickstream data analysis

Clickstream data analysis for learning performance prediction has gained the most attention and earned significant outcomes. Being abundant and varied, learning clickstream data enables the generation of learning-related attributes and features to develop learning performance prediction models with high accuracy. Table 2.2 summarizes recent research in clickstream data analysis methods for learning performance prediction and their corresponding results measured in prediction accuracy.

Table 2.2. Feature selection methods and prediction performance from clickstream data of dataset OULAD.

Authors	Feature selection methods	Prediction targets	Performance (Accuracy)
Haiyang et al. in 2018 [36]	Series of total daily clicks of each resource type done by each student in a course.	Dropout	0.950
Qiu et al. in 2022 [39]	Classification of resource types used by each student in a course into groups.	Pass / Fail	0.974
Jha et al. in 2019 [40]	Demographics, assessment scores, and interactions with OLEs.	Dropout / No dropout Pass / Fail	0.930
He et al. in 2020 [41]	Demographics, assessment scores, and click data.	Pass / Fail	0.80
Alshabandar et al. in 2020 [42]	Number of resource types and number of interactions with resources a student uses during a course.	Pass / Fail / Withdrawn	0.86
Hao et al. in 2022 [43]	Sum of clicks on resources and average assessment scores.	Pass / Fail	0.93
Drousiotis et al. in 2021 [44]	Demographics, clickstream on resources before courses start, first assignment score, and previous attempts.	Distinction / Pass / Fail / Withdrawn	0.80

2.4.3. Learning behavior patterns

Bringing semantics into online learning activities has been a concern. Not to stop at striving for high accuracy prediction of learning results, current research aims to establish learning behavior models from clickstream data. Such models help to unravel learners' original learning activities, therefore, somehow tell how the learners have used resources for learning and corresponding results. Table 2.3 summarizes several recent studies on learning behavior modeling methods from clickstream data and their research outcomes.

Table 2.3. Learning behavior modeling methods from clickstream data and achievements.

Authors	Modeling methods	Achievements
Nitta et al. in 2021 [37]	Relationships between a student's access to resources before assessments or repeatedly in use in a course.	Accesses to certain resources make patterns indicating completion and dropout cases.
Park et al. in 2017 [45]	Observation of daily access to resources and detection of changes in resource access.	Categorizing students into three groups of increased, decreased, and no change of activities access.
Yu et al. in 2018 [46]	Representation of student's resource navigation pathways.	Predicting the next resources to be used.
Li et al. in 2020 [47]	Model of time management and effort regulation using average access to resources at different time steps before deadlines.	Trying to establish a correlation between clickstream data and SRL self-report.
Kizilcec et al. in 2017 [48]	Model of time allocation for resources of each type and resource revisitation.	Trying to link resource use patterns with SRL characteristics in self-reports survey provided to students.
Cicchinelli et al. in 2018 [49]	Use of course organization and resource type to identify SRL phased related activities and The correlation between students' performance and resource use density on OLEs.	Categorizing students into 4 groups: the inactive, the continuously active, the procrastinator, and the prober.

Authors	Modeling methods	Achievements
Wong et al. in 2019 [50]	Analysis of the correlation between frequency of students' clickstream data and the support of SRL prompting videos.	Identifying resource use sequences.
Geigle et al. in 2017 [51]	Applying a two-layer hidden Markov model on resource access sequences of student groups of different performances to obtain hidden states and the transitions from one hidden state to another hidden state.	Capturing learning behavior patterns of general, high-performance, and low-performance students.
Qiao et al. in 2021 [52]	Applying a hidden Markov model with a predefined set of SRL stages and observable access to learning resources to identify the SRL learning process of mastery learners or performance learners.	Capturing different resource use models of mastery learners, goal-oriented learners, and general learners.

2.4.4. Limitations

Clickstream data are evidence and traces of learning behaviors, but they are not equal to students' cognitive and metacognitive activities. Students' thoughts drive certain clickstream patterns; however, it is uncertain to tell what the original thoughts are by only looking at clickstream data in the first place. Therefore clickstream data analysis is challenging. Two main difficulties in making clickstream data understandable are noises or data redundancy generated by unintentional actions and lack of context in which data are generated [8], [10], [53]. Predictions produced from clickstream data are signs of supporting students; however, to give helpful support, it is necessary to know the causes of such clickstream data. That is to know students' intentions and patterns of using learning resources to some extent.

2.5. Learning tactics and learning strategies

"*A learning tactic is a simple or a very short sequence of operations a learner applies to information*" [54, p. 698], and "*learning tactics serve as building blocks for multitactic learning strategies*" [54, p. 700]. Learners apply learning tactics to learning materials to achieve particular outcomes. Examples of learning tactics are skimming a research article to judge its appropriateness for a research topic, scanning a book chapter for writing a summary, and pronouncing aloud a list of vocabulary repeatedly for memorizing. In OLEs, where learning activities have simple appearances, learners' learning tactics are expressed in such traces as accessing a page of course content for grasping the course schedule and requirements, opening an ebook concerning a particular lecture for reinforcing knowledge learned in class, mouse-clicking on quiz questions for assessment or retaining knowledge. The goals of learning tactics done in OLEs might not be clearly understood by the others but learners who have performed the learning tactics. Therefore, presenting traces of learning tactics to learners helps to remind them of the effectiveness of the learning tactics in correspondence to their learning intentions. Such a reminder encourages learners to reflect on strengths and weaknesses in their learning tactics, which opens doors for adjustments and improvements.

As mentioned above that learning tactics are elements of a learning strategy, it is not that learning strategies only contain a group of learning tactics and nothing more. McKeachie described learning strategies as "*alternative mode of learning, which can be chosen when appropriate for a task*" [55, p. 8]. A learning strategy comprises a repertoire of learning tactics and judgments on conditions to select suitable learning tactics to apply. A learner might have learning tactics but no learning strategies unless he is capable of considering, estimating, and judging the conditions of a task in order to choose appropriate learning tactics to apply. In OLEs, it is very difficult for the outsider to recognize the learning strategies of a learner by analyzing his or her learning history data. But the learners themselves might be able to recognize their learning strategies by looking at their learning history data in a particular organization and presentation. They can even begin building new learning strategies by reflecting on their learning history. Such an organization and presentation of learning history data is one of the outcomes that we aim at in this research.

2.6. Conclusion

The studies and research reviewed above express a desire to understand learners' minds when they learn and have achieved significant research outcomes. The achievements have revealed and categorized learning behavior patterns, leading clickstream data closer to learners' original learning activities.

Reviewing the current outcomes, we find the approaches to modeling learning behaviors and the achievement promising and practical to press further. And the further milestone is to enable individual learners to know somehow their particular learning behavior patterns rather than those of a group, a class, or a category. A class of learning behavior patterns is an informative reference. Still, individual learners would know their way of learning or SRL and be compelled to make changes if they can see their own learning behavior patterns.

In the next two chapters, we will describe the method to model learning patterns and a framework to explain the model so that learners can make sense of their learning history data and recognize the strengths and weaknesses of their SRL capacities.

Chapter 3. SRL Recognition and Improvement Framework

Various discoveries and production of SRL models, ways of measurement and intervention call for the fundamentals of SRL operations. It is natural to seek answers to why SRL operates in such a way. No matter how many aspects and components SRL expresses, there are two subjects involved when a learner is self-regulated learning. First is the subject matter the learner is learning. The second is the learning process that the learner is performing. The former concerns cognition, and the latter involves metacognition. Our publication demonstrates observable interoperation between cognition and metacognition of a learner when studying in a certain online learning environment [7]. We believe that such a relationship, though apparent, is fundamental to awareness, recognition, and assessment of one's SRL ability not only in online learning contexts but also in one's learning process in general. In this section, we will present and justify these fundamentals of SRL and introduce the SRL recognition and improvement framework, the SRL framework for short.

This section is a slightly modified version of the authors' article [13] published in 2021 and has been reproduced here with the copyright holder's permission.

3.1. Principles of the mind

To be generic, reliable, and time-withstanding, the SRL framework must be laid on principles of the mind. The mind has two faculties (see Figure 3.1): the intellect, whose functionality is to understand knowledge, and the will, whose functionality is to drive the intellect and to choose to achieve knowledge [56]. The intellect operates as we cognize the world and its knowledge via what we usually call cognition. The activities that signify the operation of the intellect are analyzing, judging, and abstracting certain target knowledge. Specific behaviors of the intellect can be recognized via Bloom's taxonomies [57]. The will operates as we are aware of our learning process.

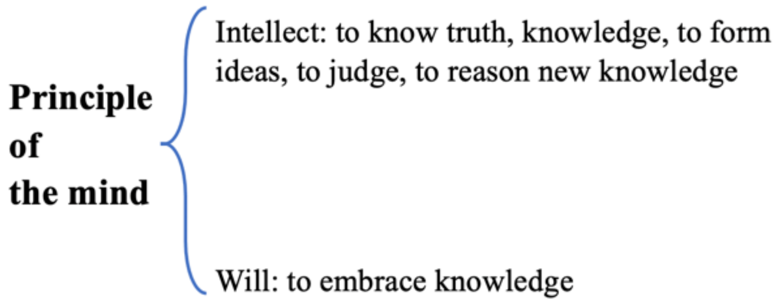


Figure 3.1. Two faculties of the mind. When we learn new knowledge, it is not difficult to notice that we not only try to understand the knowledge but also are excited by the challenges and elegance of the knowledge. The intellect enables us to understand, and the will excites us.

This statement gets clear when comparing the unconscious way a child learns with the conscious way a graduate learns. Both absorb knowledge; however, a child does not recognize their in-progress growth of knowledge while an adult does recognize it. A sign of recognizing the learning process is that adults doubt, reason over the new knowledge, and adjust their learning approach, while children tend to assent to new knowledge and follow instructions. To obtain intricate knowledge, one needs to be aware of one's learning process in order to control cognitive activities. In order words, the stronger one is aware of one's will and uses it, the more fulfillment one has towards knowledge. The will manifests itself via metacognition.

3.2. The philosophical habit of the mind

Whether we have noticed, our mind has a habit of desiring to know. The more we know the world, the more we realize that the extension of knowledge is beyond our current understanding and the more we desire to know. This routine is, as Saint John Henry Neuman [58] puts it, the philosophical habit of the mind (see Figure 3.2). Thanks to this habit, we know more about the world, assimilate knowledge, and apply it for evaluation and creation of various fields of science, art, literature, and so forth. The philosophical habit of the mind manifests in our learning process, and most clearly when we are the regulator of our own learning process, which is self-regulated learning.

SRL is a conscious learning approach by which one plan, manage, and reflect on their learning process. Looking at its characteristics, we can see that SRL operates on the inter-operation of cognition and metacognition, which follows the principles of the mind.

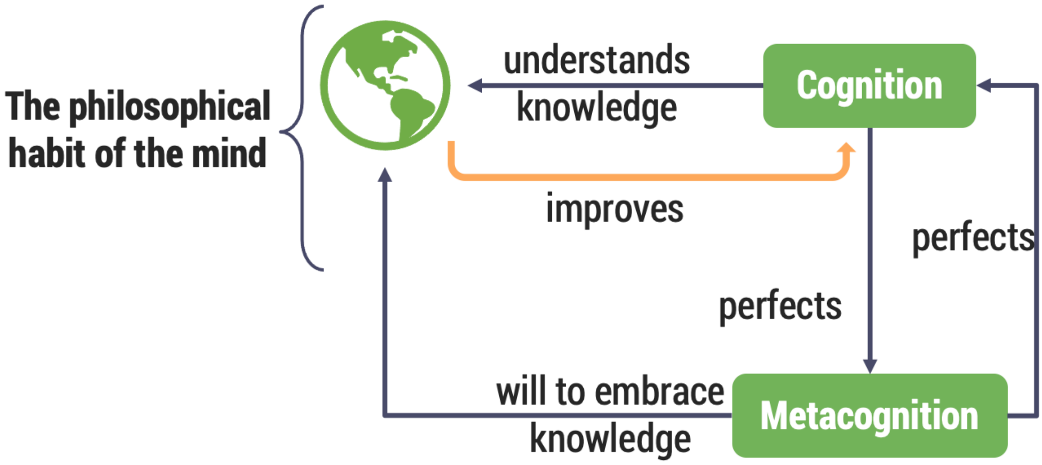


Figure 3.2. The philosophical habit of the mind. The mind with its two faculties - cognition and metacognition - tries to understand knowledge (represented as the globe) and, in the meantime, perceive the learning process. These two activities repeat and form the habit of the mind. This habit makes us know more about knowledge and ourselves.

3.3. Causes of SRL

SRL is a learning pattern that operates on the principles of the mind. Why does it exist? Everything must have reasons for its existence; otherwise, it has no use and cannot be recognized or improved. How can we recognize and evaluate our SRL? What causes SRL to exist? It is recognized based on two types of causes (see Figure 3.3): intrinsic causes, which construct the essence of SRL, and extrinsic causes, which explain the sources of SRL and the end goals where SRL leads us [59]. The intrinsic causes contain the formal cause that defines SRL structure and the material cause that personalizes the individual's SRL quality. The extrinsic causes comprise the efficient cause that explains where SRL comes from and the final cause that shows the goals of SRL and how SRL grows to its end goals.

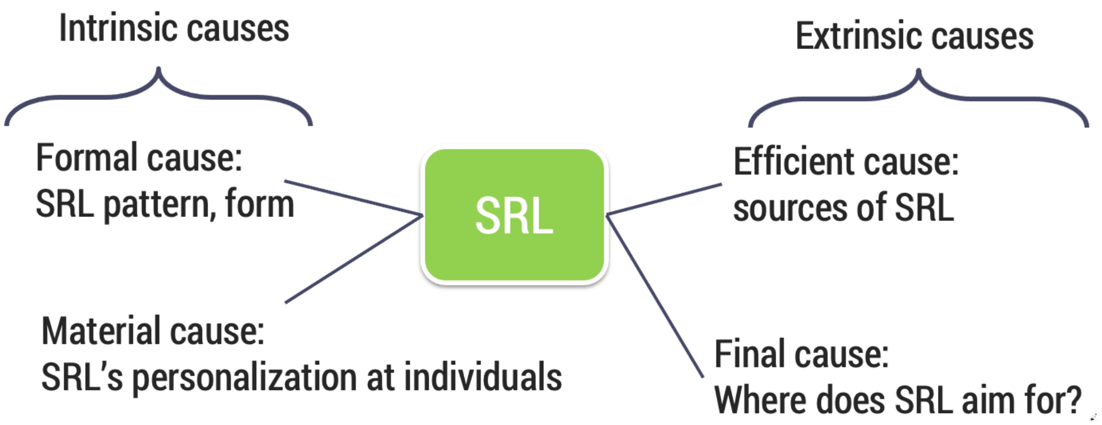


Figure 3.3. The four causes of SRL. Intrinsic causes explain what SRL is made up of and personalized from the inside, and extrinsic causes explain how SRL is modified and shaped by the outside.

When one determines and realizes these four causes of SRL, one knows how to improve SRL ability and fully benefits from SRL.

3.4. Principles of SRL

Starting from 2 faculties of the mind, their inter-operation, which molds into SRL learning pattern, we can form the principles of SRL (see Figure 3.4). As stated in a sentence, SRL is grounded in the operation of the mind, grows with the development of the mind, has a nature designed to reach the goal of understanding, and personalizes to each learner.

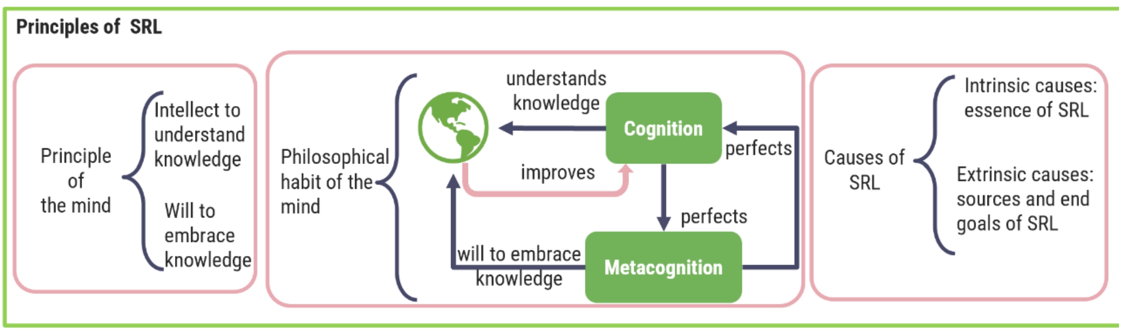


Figure 3.4. Principles of SRL. The principles enable answers to the questions of what SRL is made up of, how it operates, where it starts, and what its goals are. Such questions enhance the reasons for the existence of SRL.

3.5. SRL recognition and improvement process

Over the last two decades, there has been a wide range of research on SRL measurement and intervention for improving SRL. There are two SRL measurement approaches: SRL trait and SRL context-based skills. SRL trait describes the SRL character of a learner in general. SRL context-based skills illustrate a learner's ability to apply specific SRL skills in particular learning tasks.

SRL has been measured traditionally by data from self-reports, interviews, and questionnaires, which are usually known as the offline form of measure, and in recent years by data from learning behavior observation, which is known as the online form of measure. S. F. E. Rovers et al., in their review of SRL measurement methods, show that the offline form tends to give insight into the learner's overall level of SRL while the online form evaluates specific SRL strategies [60]. Though often analyzed separately, these two forms of SRL measure are related to each other. The offline form describes a learner's SRL character, while the online form illustrates the learner's ability to apply specific SRL skills in particular learning tasks. For the SRL measurement to provide accurate and meaningful data for SRL intervention and improvement purposes, there is a need for a firm theoretical model, grounding, or framework of SRL strategies so that the nature of SRL can be understood at the principle level and the SRL intervention can be offered to learners to support them from that fundamental basis [60], [61].

Interventing learners' learning process to improve their SRL ability is the purpose of all the SRL measurement activities. SRL intervention has been conducted via two approaches. One is that teachers help learners with specific learning tasks, and the other is that teachers provide learners with metacognitive feedback, and then the learners use the feedback for reflection and make adaptations to their learning process [61]. In the former, the assistance the learners receive is personal and related to concrete learning tasks. In the latter, the assistance is a kind of reminder and tips about learning methods. Relating to the SRL measurement approaches mentioned in the previous section, the former intervention is performed after the data collected from the online form of intervention, while the latter intervention uses the data from the offline form of intervention. The former approach is usually applied in traditional school settings. In

OLEs, the latter approach is provided with the support of educational data mining and learning analytics tools [61].

One's SRL ability is recognized by one's SRL character, which comprises one's SRL characteristics and habits of regulating his or her learning. Derived from the principles of SRL, the SRL character is fivefold: *(i)* ***wisdom,*** which is the ability to see the start and the end, *(ii)* ***knowledge,*** which is the ability to use prerequisite knowledge to acquire new knowledge, *(iii)* ***understanding*** which is the ability to apply cognitive strategies, *(iv)* ***counsel*** which is the ability to seek helps and reflect, and *(v)* ***fortitude***, which is the ability to persevere during hard times. The more consistency the SRL character demonstrates, the more maturity the SRL ability is. The development of SRL character is constructed via SRL habits, which are the habits of applying cognitive and metacognitive strategies, tactics, and skills to the learning process. For that reason, the improvement of the SRL ability begins with habituating learning strategies, both cognitively and metacognitively (see Figure 3.5).

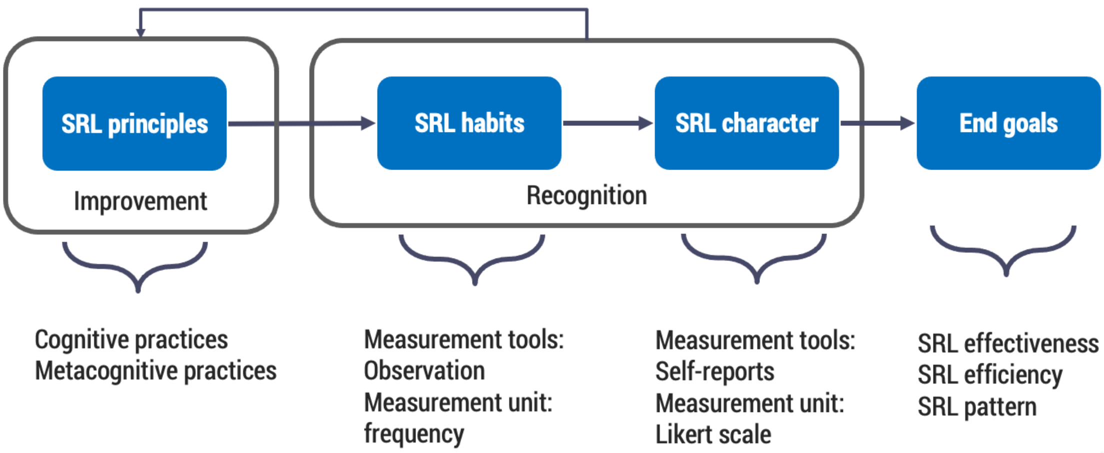


Figure 3.5. SRL Recognition and Improvement Process. The process begins with a learner recognizing and measuring SRL habits and character using appropriate measuring tools and units. With recognition, the learner refers to SRL principles to understand the strengths and weaknesses of his or her SRL habits and character to make proper adjustments according to the end goals of his or her learning.

3.6. SRL recognition and improvement framework

To establish a stable foundation for the SRL framework, we have traced the existence of SRL from the basic principle of the mind and its operation. We have walked through the reasons for the existence and development of SRL. And we have demonstrated the process by which an individual's SRL can be qualitatively and quantitively recognized and improved. Setting the SRL recognition and improvement process on the principles of SRL, we introduce the SRL Recognition and Improvement Framework (or SRL framework for short), as shown in Figure 3.6.

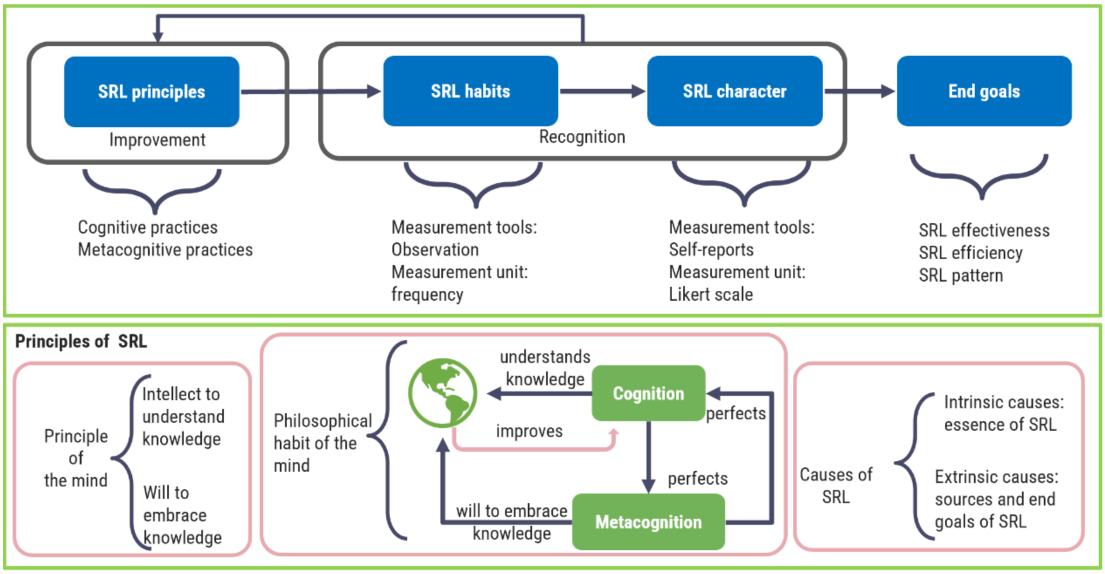


Figure 3.6. SRL Recognition and Improvement Framework.

As we stated, the purpose of this SRL framework is not to replace the existing SRL models, which play a crucial role in guiding and shaping SRL from an idea into concrete components and processes. This SRL framework provides a reference point to argue the appropriate scope where the SRL models can apply.

To demonstrate this purpose, let us briefly review the above SRL models from this SRL framework viewpoint. Reflecting on the principles of the mind, all SRL models above shows the interoperation of cognitive and metacognitive activities though some SRL models pay more attention to metacognition or motivation while the others focus on cognition. Checked against causes of SRL, some SRL models illustrate the SRL as processes and components; the other shows SRL elements to personalize SRL toward individual learners. All SRL models somehow describe the intrinsic causes of SRL, but they have not discussed extrinsic causes of SRL, which play a directive role in the SRL improvement approaches. Viewed from different perspectives and unified within this SRL framework viewpoint, applying these SRL models following a particular arrangement will help learners comprehend their SRL ability cognitively and metacognitively and show them the quality of their SRL character and the frequency of their SRL habits.

Figure 3.7 illustrates how four causes of SRL manifest in each of the SRL models above. Though the current SRL models comprise various components, as summarized in Figure 3.7 and described in the models, these components can be grouped to the causes of SRL in our proposed SRL framework. Cognition and metacognition are common components sharing similar descriptions among the current SRL models. Other components, though repetitively appearing in SRL models, are described in different perspectives with different meanings from model to model, as we mentioned in section 2.1.3 Models comparison.

Association	**Cognition**	**Meta-cognition**	**Goal**	**Ego**	**Motivation**	**Affect**
Cognition		1, 2	2, 3, 4	2	3, 4, 5	4, 5
Meta-cognition			2, 3, 4	2, 3	6	6
Goal				2	3	
Ego					2	
Motivation						4, 6
Affect						

1: Winne's SRL model [20]; **2:** Boekaerts' dual-processing model [21];

3: Boekaerts' six-component model [15]; **4:** Pintrich's framework of phases and areas for SRL [23];

5: Zimmerman's cyclical phase model [26]; **6:** Efklides's MASRL [16]

Formal cause	**Material cause**	**Efficient cause**	**Final cause**

Figure 3.7. SRL Models in reference to the SRL framework. The current SRL models partially illustrate the principles of SRL from the SRL framework. When using the existing SRL models, learners can refer to this figure to identify which aspect of SRL the current SRL models address so that they can combine the SRL models to understand their SRL.

The SRL framework is beneficial for use as a reference point to assess the validation of SRL models and design procedures, methods, and exercises for supporting individuals to evaluate their SRL ability and improve it. Since this framework is developed via arguments, future work must involve applying the framework to design empirical SRL recognition and improvement tools, programs, and exercises. Such empirical evidence will demonstrate the validity of the framework.

3.7. An empirical study on the relationship between cognition and metacognition of learners in OLEs

Because of being built on the principle of the mind of human nature, the SRL framework is applicable to SRL in various learning contexts. To demonstrate the framework's validity, we conducted an experiment as a pilot study. The full paper on this experiment is published in the article [7]. In this section, we would like to present a brief description of the experiment and the result to illustrate the explainability of the SRL framework for SRL in online learning contexts.

Let us start with the experiment background. As we recall, the SRL framework states that the SRL works according to the principle of the mind. It is the interoperation of the two faculties of the mind: intellect and will – cognition and metacognition. Therefore, a relationship between cognition and metacognition is revealed when one regulates one's learning process. We hypothesized that such a relationship between cognition and metacognition is a positive correlation. To test the hypothesis, we developed a hypothetical model about the relationship [62] and then conducted an experiment [7] in which 20 postgraduate students worked on a complex task delivered via an instance of OLEs – the learning management system Moodle [63] – and their cognitive and metacognitive scored are computed to measure the relationship. The hypothetical model demonstrates that learners' cognition and metacognition are positively correlated when the learners self-regulatedly learn complex tasks. The complexity of the tasks follows revised Bloom's taxonomy [64]. After the learners have finished the task, their cognition can be measured by accumulative scores, which we call cognitive scores, which they earn by passing assessments and exams provided by the tasks. And their metacognition can be measured by indicators of preparation, planning and reflection activities that the learners perform to accomplish the learning tasks. We called accumulative indicators the metacognition scores.

Let us describe the structure of the complex task. The complex task was a machine learning problem that required students to modify pre-defined source code to produce a high-accuracy prediction model on the MNIST dataset [65]. There were quizzes related to knowledge and techniques during the task to assist students in working on the task. The students are free to choose to do quizzes. After completing the task, students were prompted to answer a Likert-scale metacognitive questionnaire. The scores students earned by their accomplishments on the task and the quizzes would be cognitive scores, and the scores on the metacognitive questionnaire would be metacognitive scores.

The study showed a moderate positive correlation between cognitive scores and metacognitive scores, as shown in the following graph (see Figure 3.8) from [7].

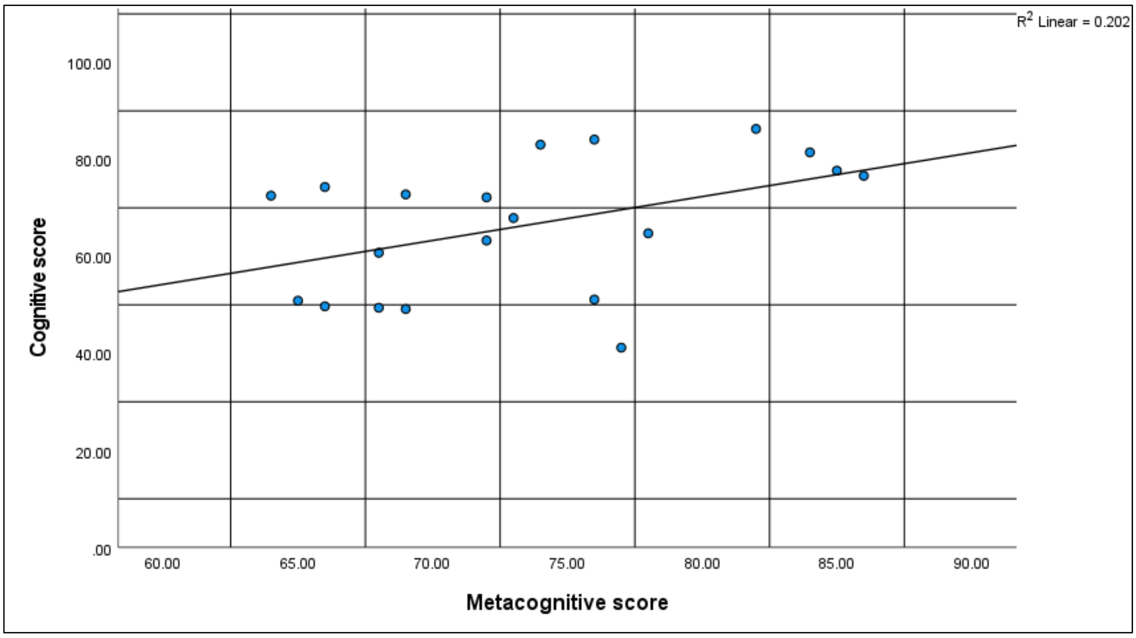


Figure 3.8. Positive correlation between cognitive scores and metacognitive scores [7, Fig. 6]. This positive correlation indicates SRL ability. The high the correlation is, the higher the SRL ability.

Further, such correlation helps categorize students into groups of SRL ability, as shown in the following graph (see Figure 3.9) from [7]. Each of these groups demonstrated a distinctive SRL profile. Following a Barnard-Brak et al.'s research about measuring SRL [66], we assigned the groups to the following 5 SRL profiles:

- Super self-regulators = active SRLers (group (a) in Figure 3.9)
- Competent self-regulators = active and adaptive SRLers (group (a) in Figure 3.9)

- Performance / reflection-endorsing self-regulator = performance-favored SRLers (group (b) in Figure 3.9)
- Forethought -endorsing self-regulators = planning-favored SRLers (group (c) in Figure 3.9)
- Non-self-regulators = non-SRLers (group (d) in Figure 3.9).

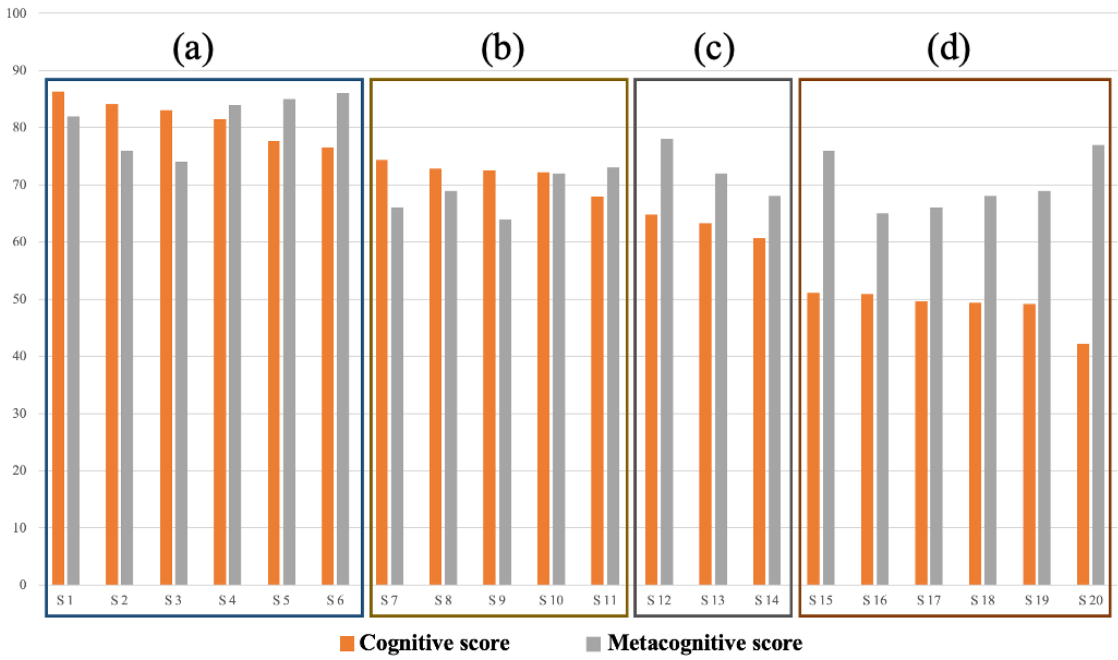


Figure 3.9. Groups of students according to the correlation between cognitive and metacognitive scores [7, Fig. 7]. Groups (a) and (b) have cognitive and metacognitive scores corresponding to one another; group (c) has slightly high metacognitive scores corresponding to cognitive scores, while group (d) has high metacognitive scores contrasting with low cognitive scores. The correspondence between the cognitive and metacognitive scores indicates learners' maturity of SRL recognition. The closer the similarity between the cognitive and metacognitive scores is, the better the learners recognize their SRL ability.

3.8. Conclusion

The result from the empirical study followed the SRL framework to a certain degree. The principle from the SRL framework explains the SRL ability, and the measurement approach from the framework provides classifiable results to some extent. There remain issues that need solutions. One issue is the separation between cognitive scores and metacognitive scores. SRL concerns a union of cognition and metacognition; therefore, there is a need for a measurement representing such a union from a learner's learning history. Another issue is the description of the SRL profile. It would be informative for learners to see their SRL profile so that they know the effectiveness of their SRL patterns in terms of activities they have performed and resources they have used for learning to make proper modifications. In the next chapter, we propose a method to address those two challenges.

Chapter 4. Formulation of SRL related problems in online learning environments

4.1. Problem formulation

Learning in an OLE manifests as a sequence of access to resources provided by the environment. Learners access resources by viewing, reading, and keyboard typing, but the most recognizable type of access is mouse clicking on the resources and the inner content. By this most popular means of resource access, learning can be represented as an observable process of clicking from one resource to another driven by a learner's invisible states of mind. OLEs usually present several resources at once, and a learner can freely access the resources in various orders. When learning the same courses, learners also differ from their peers in accessing resources. From these observations, a learning process in OLEs can be seen as a stochastic process with resources as its states; as the process runs through time, it changes its states randomly. Although there are various sequences of using resources, learners tend to access a certain resource with the intention to approach what they are learning. Therefore, the states of the learning process are dependent. Specifically, in a learning process, a state is dependent on the previous ones, or a state in the next step of the process somehow depends on the current state. Based on this attribute, we consider using the Markov chain to model the learning process in online learning.

Next, we briefly describe Markov chain principles and formulate the learning process as a Markov chain.

4.2. Review on Markov chains (MC)

A Markov chain [67] is a sequence of states appearing through time with a special property in which the values of a state only depend on a state right before it and are independent of other previous states. This property is called the Markov property [67]. The time in which states appear can be discrete-time whose values are in the range of natural numbers or continuous-time whose values are in the range of real numbers. In OLE problems, both types of Markov chains are applicable. In this research, we analyze daily access to resources; hence, we apply the discrete-time Markov chain [67].

A Markov chain is usually illustrated as the following graph shown in Figure 4.1.

S_t: the state value at time t

Figure 4.1. An illustration of a Markov chain.

To represent processes with Markov property, we generally use Markov models. A Markov model is formulated as the following.

A Markov model for a process running through time X_t consists of

- A set $S = \{s_1, \ldots, s_n\}$ of n states
- An $n \times n$ transition probability matrix $A = \{a_{ij}\}$ in which
  - element a_{ij} is the probability of state i at time X_t transitioning to state j *at time* X_{t+1},
  - a_{ij} is nonnegative, and $\sum_{j=1}^{n} a_{ij} = 1$,
  - $a_{ij} = P(X_{t+1} = j \mid X_t = i) = P(X_{t+1} = j \mid X_t = i, X_{t-1} = k, \ldots, X_0 = m)$ (the Markov property or Markov assumption),
- A set $\pi = <\pi_1, \ldots, \pi_n>$ of initial probability distribution over S to show how probable a state would start the process, and $\sum_{i=1}^{n} \pi_i = 1$.

Markov chains are used to model sequences of weather change from one day to another in which tomorrow's weather can be predicted by looking at the weather condition of today, or model the sequence of web page visitation in which the probability of a user visiting the next page is computed by the current page he is staying [68][69].

Next, let us think about how parameters of a Markov chain are calculated.

For a Markov chain, two parameters needed for calculation are the initial probability distribution π and the transition probability matrix A. One of the approaches is to count and divide [68]. As π is about the frequency of a state starting a process, π is computed by first counting the number of each state starting a process out of processes and then dividing those numbers by the number of processes. And since A is about the frequency of changing from one state to another state in the very next step, A is computed by first counting the number of occurrences of a transition between these two states in the same order, then counting the number of occurrences of the former state, and finally divided the former calculation by the latter. The parameter estimation [68] is generalized as the following.

Given a phenomenon whose states transition in *N* observable processes;

a_{oi} the number times state *i* starts a process;

p_i the number of times state *i* occurs in a process (*);

p_{ij} the number of times state *i* transition to state *j* in a process;

a_{ij} the probability of a transition from state *i* to state *j;*

π_i the probability of state *i* starting a process

$$\pi_i = {a_{0i}}/{N} \quad (4.1)$$

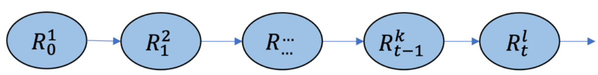
$$a_{ij} = {p_{ij}}/{p_i} \quad (4.2)$$

(*) If state i is the end of a process, this occurrence of the state i will not be counted for the calculation of a_{ij} *because there is no transition from state i in this condition.*

4.3. Markov model of a learning process

A learner joins an online course for a period of time and studies by accessing resources of various kinds provided by the course. The time step for marking resource use is one day. The learner uses several resources for study in a time step. From this description, the learner's learning process can be seen as a Markov chain for one or many observable sequences of transition from one resource use to another with certain relations between resource choice between 2 time steps.

A learning process is illustrated as the following Markov chain in Figure 4.2.



R_t^i : Resource *i* is being used at time step *t*

Figure 4.2. A learning process illustrated as a Markov chain.

Suppose a learner has a record of resource use as following

homepage, content, resource, content, resource, content

content, quiz, content, resource, quiz, content

homepage, resource, homepage, content, homepage

The learner's learning record expresses itself in 3 Markov chains, as shown in Figure 4.3 below.

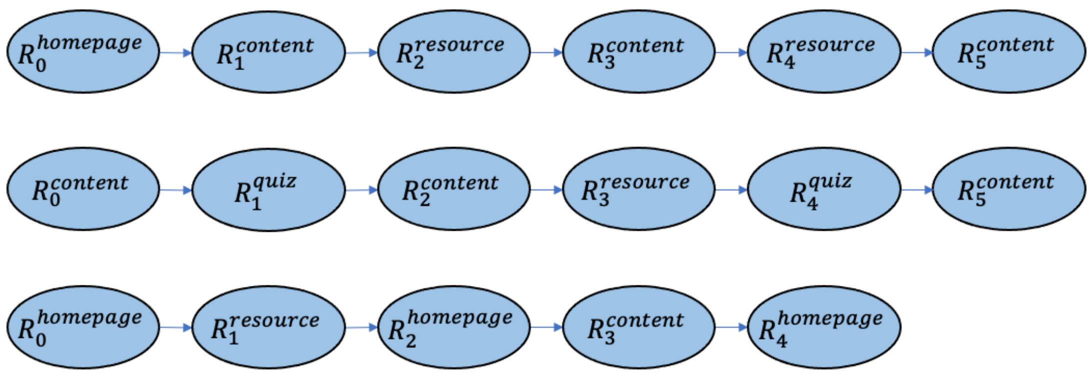


Figure 4.3. Learning records expressed as Markov chains. There are 3 sequences of resource use in this figure, in which R_t^i indicate resource named *i* is used in time step *t.*

All resources, which are identified in the sequences, form a set of possible states, which is

$S = \{homepage, content, resource, quiz\}$

Let us calculate the parameters for the Markov model

The number of states is the number of resources having been used.

$N = |\ \{ homepage, content, resource, quiz \}\ | = 4$

The number of times each resource starts a process

$a_{0homepage} = 2,\ a_{0content} = 1,\ a_{0resource} = 0,\ a_{0quiz} = 0$

The number of times each resource occurs in the processes when there is a transition from the resource to another resource.

$p_{homepage} = 3,\ p_{content} = 5,\ p_{resource} = 4,\ p_{quiz} = 2$

Table 4.1 shows the computation of the number of times there is a transition between 2 resources. Table 4.2 demonstrates the computation of the transition probability matrix.

Table 4.1. Computing numbers of transitions between two learning resources

p_{ij}	homepage	content	resource	quiz
homepage	0	2	1	0
content	1	0	3	1
resource	1	2	0	1
quiz	0	2	0	0

Table 4.2. Computing a transition probability matrix

a_{ij}	homepage	content	resource	quiz
homepage	$p_{homepage\ hompage}$ / $p_{homepage}$ = 0/3 = 0	$p_{homepage\ content}$ / $p_{homepage}$ = 2/3	$p_{homepage\ resource}$ / $p_{homepage}$ = 1/3	$p_{homepage\ quiz}$ / $p_{homepage}$ = 0/3 = 0
content	1/5	0	3/5	1/5
resource	1/4	2/4	0	1/4
quiz	0	1	0	0

4.4. Insights from Markov chains

In a Markov model, the probability distribution over states approaches stability as state transition happens a large number of times. This condition is known as steady-state or a steady-state behavior of a Markov model [68], [70]. In a steady state, the probability distribution over states is independent of the initial probability distribution of the states; therefore, a steady state is like a pattern, routine, or habit that a Markov model represents [71].

4.5. Online learning habits as Markov Models

4.5.1. Problem statement

From a learner's record of resource use in a course in an OLE, what are the learner's habits of using resources?

Putting the problem in specific phases of a learning process, we restate the problem this way. What are a learner's planning habits and reflection habits from his or her record of resource use?

In the learning context that we have been talking about so far, habits are routines and patterns of learners accessing resources. So planning habits and reflection habits are resource access patterns that learners apply in the planning and reflection phases of a learning process. To identify such habits, next, we present how the Markov model helps analyze records of learning resources and develop models to represent habits.

4.5.2. Arrangement of resource use history

To build Markov models out of learning history, let us first look at what kinds of data are contained in them. There are several OLEs being used, different from each other in the kind of data they are storing, but they share similar semantic structure and information about resource use records, namely, what resources learners access or click in a certain time unit. Following the shared structure, we can read from a resource use history of learners about their learning pattern comprising of what resources have been used, sequences of resource access, and resource use before and after a certain milestone such as an exam or a lecture.

Let us look at the problem above about recognizing a learner's planning and reflecting habits. The next step is to organize resource use history data by using the shared structure so that the data, when being turned into Markov models, will reveal patterns for further analysis and help us recognize the learner's habits.

Planning is about preparation activities that learners often do before an event, and when the event has finished, some learners review their performance on the event by reflecting. In an online course, milestone information is often available or easily obtainable, so it is possible to divide data of resource use into those before a certain event and the other after the event, the data related to the planning phase and the other of the reflecting phase. Figure 4.4 below demonstrates such organization of a history of resource use in the form of a resource clickstream through a period of time into sequences of resource use for a particular objective, namely, planning, performance, and reflection.

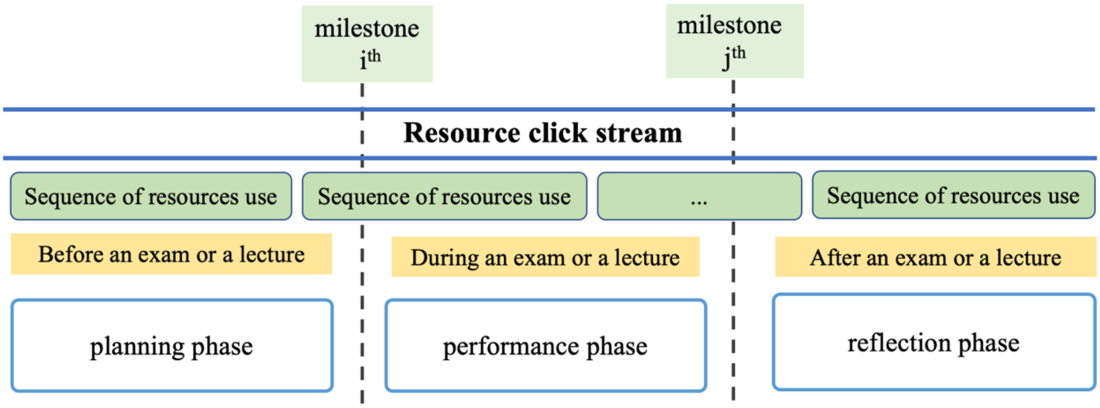


Figure 4.4. Arrangement of resource use into planning, performance, and reflection phases. Using timestamp information as a milestone for a particular learning event, we can divide resource click stream into activities of different SRL phases. For example, planning phase contains activities that learners often do before a milestone, performance phase contains activities happening during from one milestone to another, and reflection phase contains activities done after a milestone.

4.6. Open university learning analysis dataset

The previous section describes a possibility for building Markov models out of records of resource use to represent learning habits, specifically habits of planning and reflection.

Let us go a step further to actually build such Markov models. As we have mentioned that resource records from different OLEs share a similar structure, thanks to this fact, a procedure for building Markov models out of data from one system is possibly applicable to data of other systems. Thus, we think of developing Markov models from a dataset that is open and large so as to justify the procedure. And we start with the Online University Learning Analytics Dataset, which is also known as OULAD [12].

OULAD is provided by the Open University (the university's website is https://www.open.ac.uk) – one of the largest universities for distance learning worldwide. OULAD contains 10,655,280 data rows about online resource use from 32,593 students in 7 courses with 22 course offerings in 2013 and 2014. The dataset also includes specific resource types used in each course offering and students' performance to the extent of exam scores and final results. Courses at Open University are offered via the Moodle learning management system (LMS). Working on the OULAD, we will use the terms *learners* and *students* interchangeably because the academic degree seeking is not required for analysis in this research context.

OULAD contains seven .csv files, each containing the following information (shown in Figure 4.5). The structures of each .csv are displayed in Appendix 1.

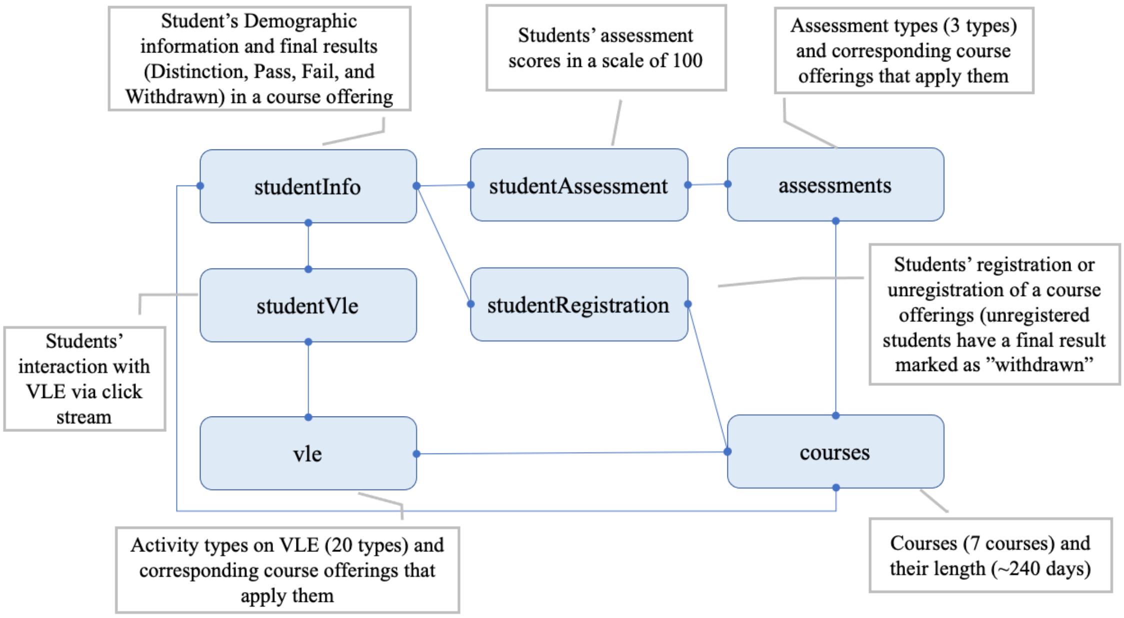


Figure 4.5. Redrawn OULAD structure [12, Fig. 4]. These are the OULAD files with a brief description.

4.6.1. Data pre-processing

We temporarily ignore the demographic information and only focus on academic information.

Applying the shared structure for record of resource use, the following is the information necessitating recognition of a learner's resource use habits and the effects of the habits:

- Resources and dates when the learner use the resources n a course;
- The number of times that learners access a resource. Resource access is about whether a learner uses a resource in a certain. It is not about how many clicks the learner performs on a resource;
- When are exam dates;
- How much score the learner earns in the exams.

Joining appropriate OULAD files, we generate the following simple files for building the Markov model.

The first file (see Figure 4.6) is about students' interaction with OLEs, each row presenting which resource a student uses on which date in a course. The second file (see Figure 4.7) is about examinations of the course, each row presents which exam contributes how much weight to the overall score and is due at which date in the course. The third file (see Figure 4.8) is about students' exam scores.

code_module	code_presentation	id_student	date	activity_type	num_use	sum_click	final_result
BBB	2013J	586516	60	resource	130	141	Withdrawn
BBB	2013J	2464683	65	resource	78	85	Pass
BBB	2013J	1541133	152	resource	38	43	Fail
BBB	2013J	52899	7	resource	36	45	Pass
BBB	2013J	1541133	32	resource	35	42	Fail
BBB	2013J	591558	219	resource	35	53	Pass
BBB	2013J	2229783	3	resource	32	41	Fail
BBB	2013J	585693	232	resource	28	33	Pass
BBB	2013J	2464683	18	resource	27	28	Pass
BBB	2013J	2463213	201	resource	26	27	Pass
BBB	2013J	558983	219	resource	25	29	Fail
BBB	2013J	2464683	81	resource	25	29	Pass
BBB	2013J	574212	241	resource	25	34	Distinction

Figure 4.6. Sample of synthesized data about students' interaction with OLEs. The data include who (learners) use which resources at which date in a course.

id_assessment	assessment_type	date	weight	assess_week
14996	TMA	19	5	3
14997	TMA	47	18	7
15003	CMA	54	1	8
15004	CMA	96	1	14
14998	TMA	96	18	14
15005	CMA	131	1	19
14999	TMA	131	18	19
15006	CMA	166	1	24
15000	TMA	166	18	24
15007	CMA	208	1	30
15001	TMA	208	18	30
15002	Exam	NA	100	NA

Figure 4.7. A sample of an assessment data. The data include assessments with types, dates, weights which are the percentage of contribution to overall assessment, and the week when the assessment is due.

id_assessment	id_student	date_submitted	score	assessment_type	date	weight	assess_week
14996	23798	18	90	TMA	19	5	3
14996	27759	18	61	TMA	19	5	3
14996	30091	24	80	TMA	19	5	3
14996	31014	4	85	TMA	19	5	3
14996	31849	15	81	TMA	19	5	3
14996	37622	19	70	TMA	19	5	3
14996	38234	17	70	TMA	19	5	3
14996	47855	-1	41	TMA	19	5	3
14996	48040	18	60	TMA	19	5	3
14996	48503	17	76	TMA	19	5	3
14996	51301	17	93	TMA	19	5	3
14996	52797	16	87	TMA	19	5	3

Figure 4.8. A sample of students' scores data. The data include assessments with partial scores and weights that students have earned.

4.6.2. Parameter calculation

4.6.2.1. Resource use sequences

In OULAD, all available resources, which are called activity types, are considered as the states in a Markov model; and the time step unit is a day or date (see Figure 4.9). A learner often accesses plenty of resources in a day in an unknown order. So, we do not know resource use sequences in a day; however, it is possible to obtain sequences of resources a learner accesses from day to day successively (as shown in Figure 4.10). A resource use record of a learner, when being observed in this manner, generates a large number of sequences (as shown in Figure 4.11), which is effective in producing a Markov model relevant to the learners' patterns of resource use.

id_student	date	activity_type	num_use	sum_click	final_result
All	All	All	All	All	All
23798	-4	forumng	1	1	Distinction
23798	-4	homepage	1	3	Distinction
23798	-4	resource	1	1	Distinction
23798	-4	subpage	1	1	Distinction
23798	6	homepage	1	2	Distinction
23798	6	subpage	1	1	Distinction
23798	6	url	1	1	Distinction
23798	7	forumng	9	25	Distinction
23798	7	homepage	1	3	Distinction
23798	7	oucollaborate	1	1	Distinction
23798	7	subpage	2	3	Distinction
23798	7	url	1	3	Distinction

Figure 4.9. Resource use history.

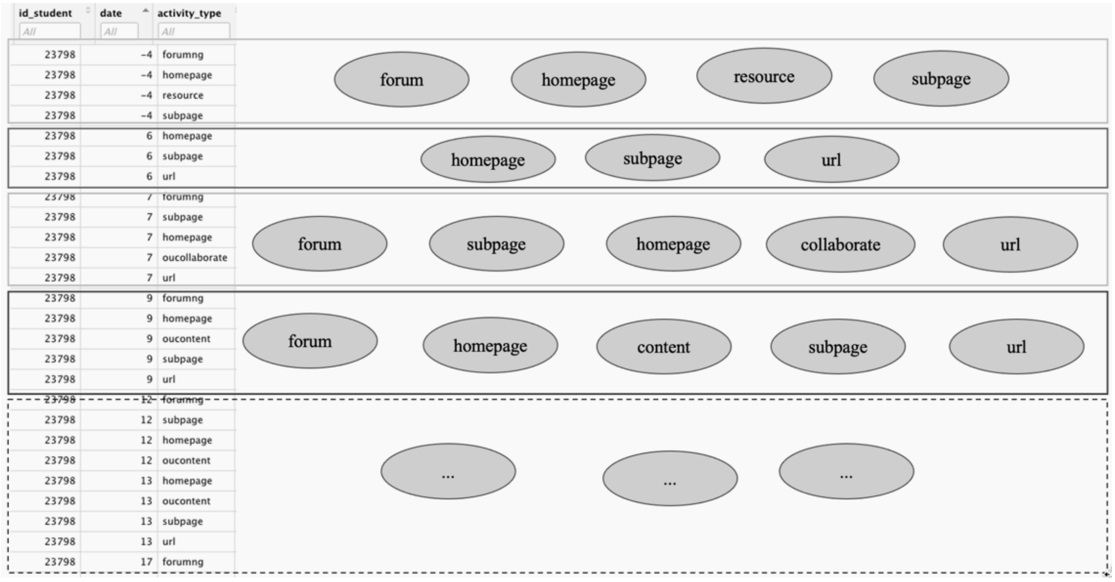


Figure 4.10. Viewing resource use in a successive order. The resource use history data (left) shows the resources or activity types that a learner accesses on a certain day. The sequences of resource use (right) become clearer when we rearrange the data in an ascending order of day or date and group resources used by day.

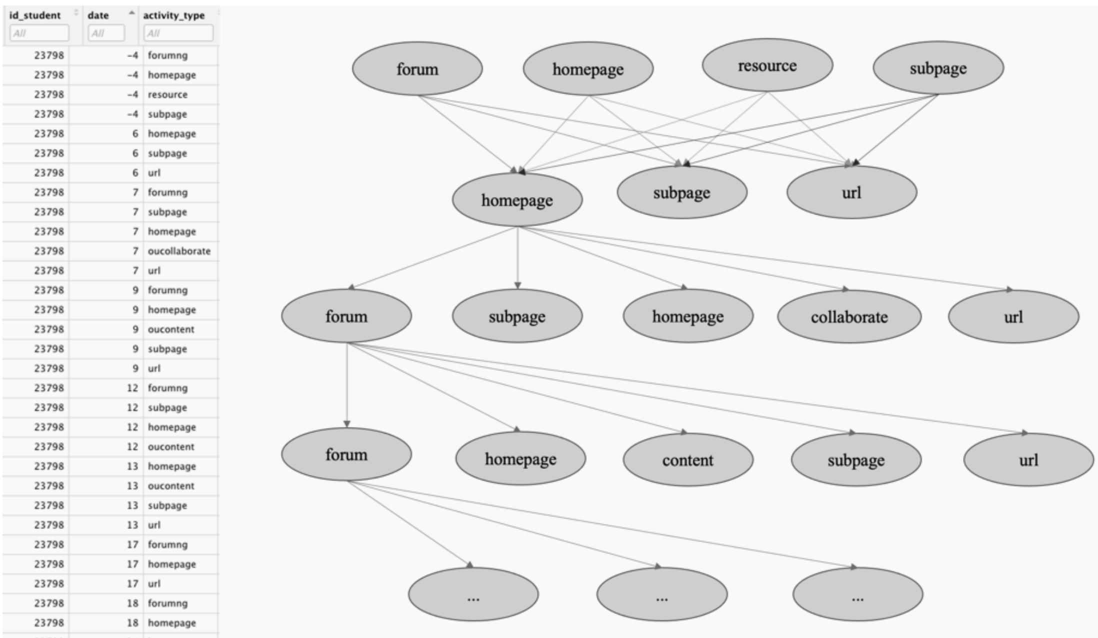


Figure 4.11. Forming resource use sequences. A resource use sequence is formed when we traverse from a resource used on a certain day to other resources used on the next day (as the arrows indicate) and continue this routine to the resources used on the end day in the resource use history. Followed this approach, the learning history of a learner (left) can generate a large number of resource use sequences needed for the learner Markov model to reach its steady state.

4.6.2.2. **Transition probability matrix**

The calculation follows the formula *(4.2*) above. Following the sequence formation above, we develop an algorithm to compute a transition matrix from a student's resource use history, described in Table 4.3.

Table 4.3. Algorithm to compute a transition probability matrix.

function transProbMatrix(*resource use history of a student* ***R***) **returns** *a transition probability matrix*
states ← extract activity_type of the resources from ***R***
A ← initialize a square matrix ***A***, size n = | ***states*** |,
value of each element = 0, ***A*** $= \{ a_{ij} = 0 \}$,
each row and column is named after each activity_type from ***states***
For each row ***i*** in ***R***
next ← query a list of records ***j*** of next date of ***i*** so that ***j****\$date* = ***i****\$date + 1*
For each row ***j*** in ***next***
A[i*\$activity_type,* ***j****\$activity_type]* ← ***A[i****\$activity_type,* ***j****\$activity_type] + 1*
End For
End For
For each row ***i*** in ***A***
sum ← add all element values of row ***i***
For each element ***j*** in row ***i***
A[i*\$activity_type,* ***j****\$activity_type]* ← ***A****[i\$activity_type,* ***j****\$activity_type] /* ***sum***
End For
End For
Return A
End function

4.6.2.3. **Initial probability distribution**

Before we think of how to compute the initial probability distribution, let us recall that a Markov process goes into a steady state in which the probability distribution over states after a large number of time steps will converge to a stable number and be independent of the initial probability. In the case of the learning process extracted from OULAD, the number of learning sequences for each learner in one course is approximately more than a hundred. This number is large enough for the steady state condition to occur. So, the initial probability distribution does not need an exact calculation. In a general sense, we may not know exactly starting points of a person's habits when seeing he or she is performing his or her habits. But we can guess that one's habits come from the repetition of an activity. Applying this approximate calculation, we compute the initial probability distribution as the frequency of use of each resource in a course. The calculation is expressed in formula (4.3) above. We develop an algorithm to compute the initial probability distribution, described in Table 4.4

$$\pi_i = {}^{number\ of\ use\ of\ resource\ i}\!/_{sum\ of\ all\ resource\ use} \qquad (4.3)$$

Table 4.4. Algorithm to compute the initial probability distribution over resource use of a learner.

function InitProbDistribution(*resource use history of a learner **R***) **returns** *a probability distribution vector*

states ← extract activity_type of the resources from ***R***

π ← initialize a vector ***π***, size ***n*** = |***states***|, π_k is the initial probability of resource ***k***

value of each element of ***π*** is 0

sum ← count number of use of all resources in R

For each row ***i*** in ***R***

increment $\boldsymbol{\pi_k}$ of resource ***k*** which appears in row ***i*** by 1

End For

For each element ***k*** in ***π***

$\boldsymbol{\pi_k = \pi_k / sum}$

End For

Return π

End function

4.6.2.4. Learner Markov model

We used the markovchain package [72] for R to build learner Markov models from OULAD data. The package provides functions to build a Markov chain from a transition probability matrix and states. Figure 4.12 is a sample learner Markov model.

Transition probability matrix

```
23798-Distinction
 A  9 - dimensional discrete Markov Chain defined by the following states:
 forumng, glossary, homepage, oucollaborate, oucontent, quiz, resource, subpage, url
 The transition matrix  (by rows)  is defined as follows:
                 forumng glossary  homepage oucollaborate  oucontent       quiz   resource    subpage       url
forumng       0.04761905        0 0.3809524    0.00000000 0.04761905 0.04761905 0.04761905 0.19047619 0.2380952
glossary      0.00000000        0 0.5000000    0.00000000 0.00000000 0.50000000 0.00000000 0.00000000 0.0000000
homepage      0.10666667        0 0.4133333    0.01333333 0.02666667 0.09333333 0.01333333 0.09333333 0.2400000
oucollaborate 0.00000000        0 0.5000000    0.00000000 0.00000000 0.50000000 0.00000000 0.00000000 0.0000000
oucontent     0.00000000        0 0.4000000    0.00000000 0.20000000 0.00000000 0.00000000 0.20000000 0.2000000
quiz          0.15000000        0 0.4000000    0.00000000 0.05000000 0.20000000 0.00000000 0.05000000 0.1500000
resource      0.10000000        0 0.4000000    0.00000000 0.00000000 0.10000000 0.10000000 0.20000000 0.1000000
subpage       0.10000000        0 0.3500000    0.05000000 0.05000000 0.05000000 0.05000000 0.20000000 0.1500000
url           0.12820513        0 0.4358974    0.02564103 0.00000000 0.07692308 0.00000000 0.02564103 0.3076923
```

Steady states

```
   forumng glossary homepage oucollaborate  oucontent       quiz   resource    subpage       url
 0.1037081        0 0.408481    0.01624984 0.03180307 0.09439094 0.01697388 0.09783198 0.2305612
```

Figure 4.12. A sample learner Markov model. The transition probability matrix is computed from resource use sequences of a learner. Each row demonstrates the likelihood of transition in the learner's resource use from a resource indicated in the row header to the following resource indicated in the column header. The higher the cell value, the higher the transition likelihood. The steady state shows the likelihood of each resource starting a resource use sequence. For example, this steady state shows that the learner is likely to start learning with access to the homepage.

4.7. Self-regulated learner profile (SRL profile)

Let us be reminded that the objective of this research is to support learners' recognition of their SRL capability and ability via their learning history data. In the sections above, we see that clickstream data contain learning behaviors and performance and SRL patterns. Applying existing libraries, we have generated learner Markov models from learners' clickstream data.

In this section, we present how learners can recognize their SRL capability and ability. It is achievable in an SRL profile.

SRL profile shows the frequency and density of resource use in a course, before or after an examination or a lecture, and the effectiveness of resource use in such a pattern. Frequency refers to a probability distribution over resource use and probability transition from one resource to another. Density refers to the number of access to resources.

Looking at an SRL profile, one sees the relationship between one's resource use indicators and one's performance in a course. We propose a measurement for scoring this relationship. The measurement is named the SRL index.

4.7.1. Self-regulated learning index (SRL index)

The SRL index indicates how effective a resource use pattern is in corresponding to its learning performance and is expressed in the form

$$SRL\ index = c\_score \sum_{i=1}^{n} r_i^{ss_i} \qquad (4.4)$$

where

SRL index : an index representing how well a learner self-regulates his or her learning process in a course. The higher the SRL index is, the higher a learner performs in SRL;

c_score : accumulative performance score a learner has earned in the course;

r_i : density of resource i in the course; density refers to the number of resource access or number of clicks;

ss_i : the steady state probability of resource r_i;

n : the number of resources available in the course.

SRL index tells the effectiveness a resource use pattern of a learner has on their learning performance. The score demonstrates a correlation between a resource use pattern of a learner and their performance result. The term $r_i^{ss_i}$ illustrates a steady pattern of use of the resource i. $\sum_{i=0}^{n} r_i^{ss_i}$ is a summary of the resource use pattern of a learner in a course. An effective SRL pattern earns high cognitive performance. That expresses in the term $c\_score \sum_{i=0}^{n} r_i^{ss_i}$.

To retain data normality, log transformation is applied to the SRL index.

$$Log\ SRL\ index = log_2(c\_score \sum_{i=1}^{n} r_i^{ss_i}) \qquad (4.5)$$

The formation of the SRL index not only demonstrates the relationship between cognition and metacognition of a learner when they perform SRL in OLEs but also measures such relationship quantitatively and expresses it semantically.

4.7.2. Explaining the SRL profile of a learner by SRL Framework

According to the SRL framework, the formal cause bases SRL on two faculties of the mind of a learner: cognition and metacognition. These two faculties interoperate and have a positive correlation with the learning performance of the learner [7]. Depending on current SRL models by which SRL is analyzed, the correlation between cognition and metacognition manifests differently. For instance, Zimmerman's cyclical phase model [26] presents the cognition and metacognition relationship in SRL as a three-phase process, while Boekaert's dual-processing model views that relationship balancing process between learners' goals and resources.

The form of SRL is then personalized to individual learners in the quality of the relationship between cognition and metacognition. The quality is about, but not limited to, the pattern of resource use and its effectiveness, the steady state distribution over resource use. Such quality is quantitatively measurable via the SRL index.

Recognizing SRL status might enable changes to improve ways of learning and improvement learning performance. Such changes are of the efficient cause. Learners can refer to their peers to adjust learning styles, or seek counsel with instructors or supervisors to quit bad learning habits (such as being absorbed by certain resources), adopt good learning habits (viewing course contents to outline the learning process, posting explicit opinions on forums), or begin to use learning resources they have not used before to strengthen knowledge retainment. Such activities, in return, perfect their SRL ability gradually.

The final cause of SRL is the most difficult cause to identify. A learner wants to perfect his or her SRL ability to seek knowledge to a greater extent. Since SRL comprises cognition and metacognition, perfecting SRL is to perfect cognition and metacognition. Manifestation of perfecting self-regulated learners is that they find learning resources provided in a course useful for learning; therefore, they use the provided resources diligently, and their use of resources demonstrates accumulative contributions to their achievement in learning. Learners with such perfecting SRL ability are reaching the perfection of the mind. The fruit of such a mind is the joy of learning, discovering, and knowing.

4.8. Conclusion

The procedure for data transformation and modeling brings out meanings from learners' learning history data and presents sequences of resource use in such a manner that learners can see their learning patterns. The application of the SRL framework to explaining SRL profiles sheds light on the learners' understanding of their SRL habits. In the next chapter, we present the application of the procedure on the OULAD dataset and evaluate the performance of our proposed methods with that of existing approaches.

Chapter 5. Method evaluation and Discussions

Chapter 4 has elaborated the process of building a learner Markov model from online learning behaviors in the appearance of clickstream data and introduced the SRL profile for illustrating SRL capability and ability. The steady states attribute of the learner Markov model demonstrates learning patterns, which can be seen as learners' resource use habits to achieve certain learning goals. The proposed SRL index in the SRL profile is a single number unifying a learner's resource use pattern and learning performance; therefore, it helps give a valid rank of learners' SRL ability.

In this chapter, we present the generation of the learner Markov model, the SRL index, and the SRL profiles for learners from OULAD (refer to section 0for a description of this open dataset). We show what an SRL profile looks like, how accurately the SRL index classifies learning performances and also evaluate the proposed method by comparing its effectiveness with existing approaches.

5.1. Visualization of a self-regulated learner profile

An SRL profile of a learner represents his or her SRL capacity and ability. Looking at one's SRL profile, one should be able to tell one's SRL level and pattern of learning activities. From the understanding of the SRL profile, a learner should see their SRL strengths and weaknesses to make proper changes.

An SRL profile shows the SRL index in the performance group to which the SRL profile belongs, and in comparison with other SRL indices in a course, the density of resource use, which is measured in the number of accesses to the resource, and the steady states or a steady probability distribution of resource use over resources in a course.

By visualization, an SRL profile is presented in a combination of 4 graphs in Figure 5.1:

- (a) a boxplot of SRL indices separated by learning performance groups (for example, in OULAD, learning performance groups are the distinction, the pass, the fail, and the withdrawn);
- (b) a scatterplot of correlation between SRL indices and cognitive (or performance) scores of learners;
- (c) a scatterplot of the number of accesses to each resource in a course;

- (d) a scatterplot of steady states of resource use.

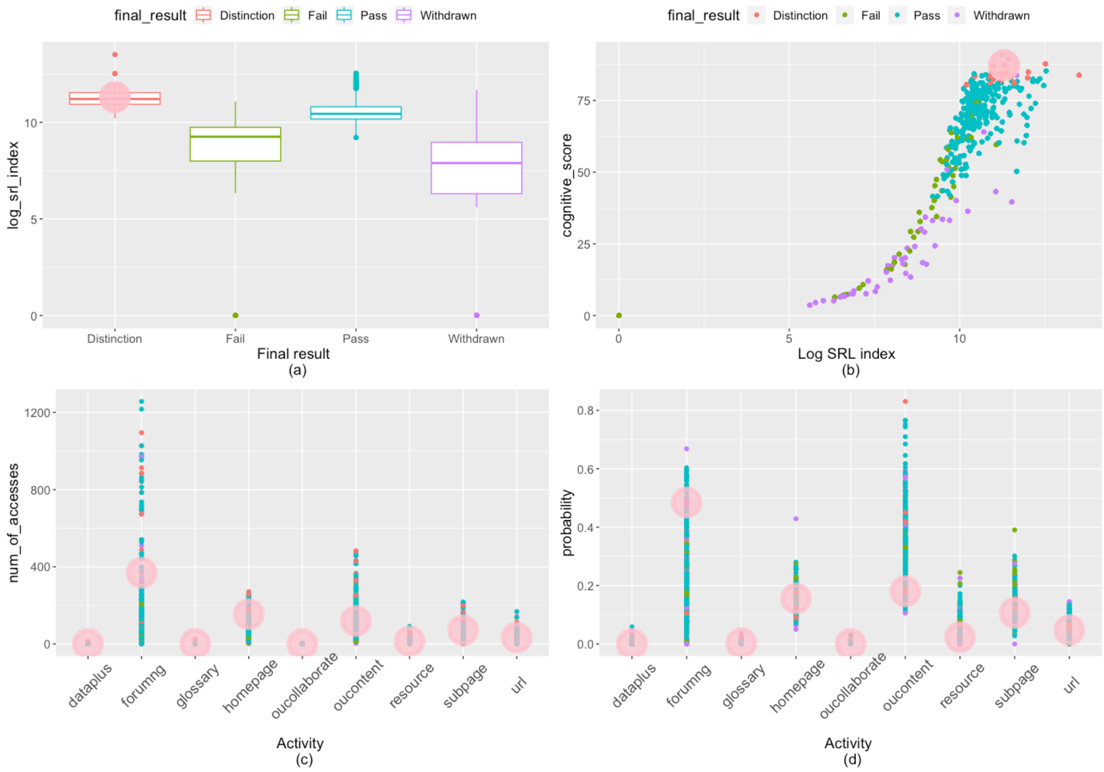


Figure 5.1. Visualization of an SRL Profile in a course that contains 4 learning performance groups indicated by final results. The groups are the Distinction, the Fail, the Pass, and the Withdrawn. The big red circle shows the learning patterns of a learner in comparison with other learners in the course. (a) shows the learner's log SRL index. (b) shows the correlation between the learner's log SRL index and her cognitive score. (c) shows the learner's resource use density. (d) shows resource steady state to indicate learning habits.

Let us look at an example of an SRL profile of a learner in Figure 5.1. In this SRL profile, all four graphs comprise a big picture of learning patterns and SRL abilities of all learners grouped by their learning performance in a course. And the big red circles indicate the current state of the learner who owns this SRL profile. The upper-left boxplot tells that the learner is in the distinction group with an SRL index of about 11 points. The upper-right scatterplot shows that the learner has earned a high cognitive score of more than 90 points in relation to the SRL index of about 11 points. The lower-left and lower-right scatterplots demonstrate that learners might often refer to the course outline provided in the resources named *homepage* and *outline*, and use learning materials provided by the resources named *resource*, *subpage*, and *url* in a reasonable density and frequency.

Section 4.5.2 above describes how resource use history can be divided into the SRL planning phase and reflection phase using assessment milestones. Applying this approach, we can produce SRL profiles for different SRL phases, such as an SRL profile for the planning phase, another SRL profile for the reflection phase, and another SRL Profile for the overall course.

5.2. SRL Profiles of each learning performance group

There are significant differences in SRL profiles of different performance groups. The first apparent difference is the SRL index or log SRL index. Figure 5.2 shows SRL indices of the distinction, the fail, the pass, and the withdrawn groups. The distinction and the pass are clearly separated from the fail and the withdrawn. Between the fail and the withdrawn, the mean log SRL index can also be used for classifying these two groups.

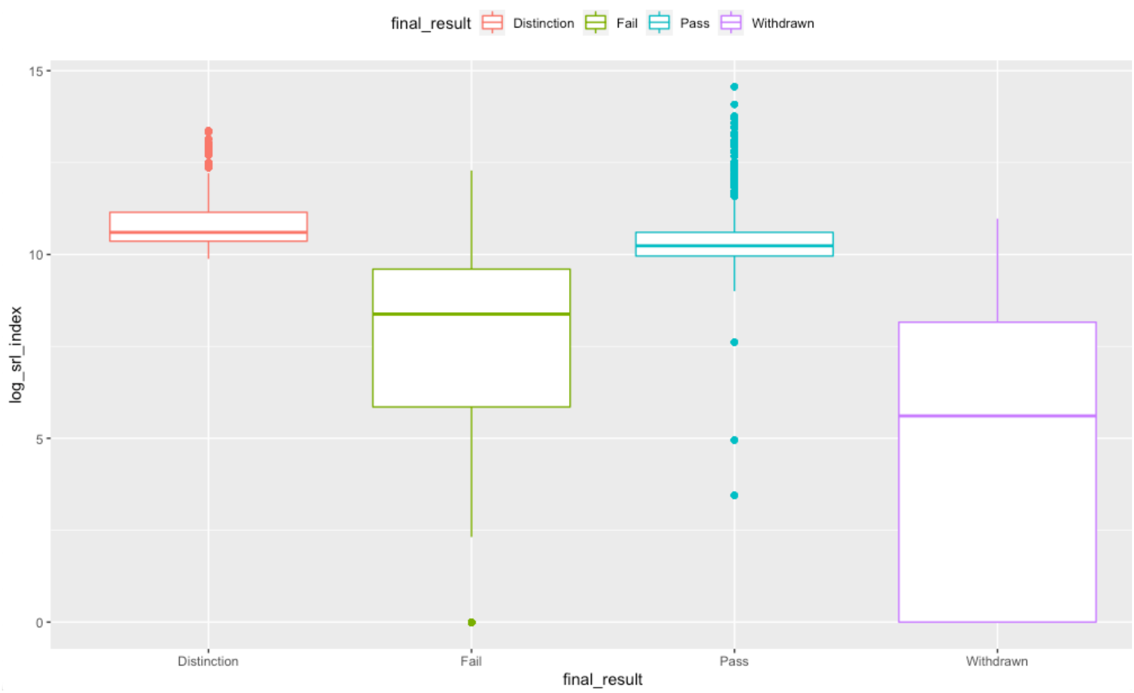


Figure 5.2. Comparison of SRL Indices among performance groups. Except the outliers data, the pass students (of Distinction and Pass groups) are well-separated from the fail students (of Fail and Withdrawn groups).

The second difference between the groups is the correlation between the log SRL index and cognitive score. As reported in Figure 5.3, the relationship between the log SRL index and cognitive score helps recognize learners' performance. Visually speaking, from the graph, the boundaries among the performance groups are clearly defined.

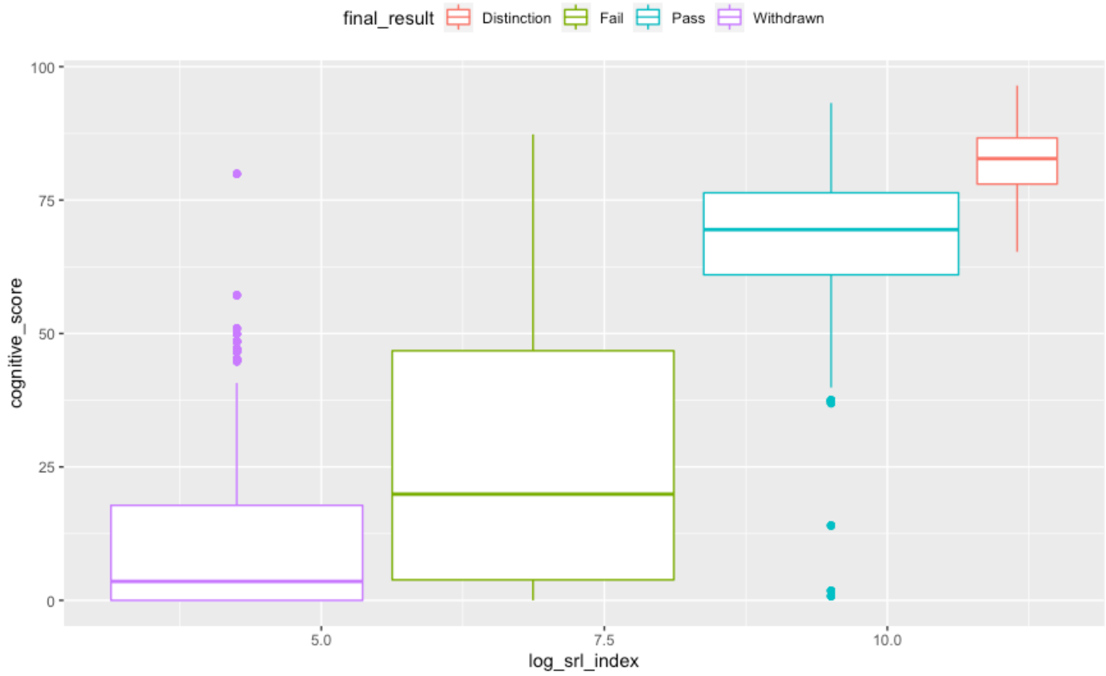


Figure 5.3. Comparison of correlation between Log SRL index and cognitive score among performance groups. Except for the outliers data, the pass learners (of Distinction and Pass groups) have higher log srl indices and cognitive scores than the fail learners (of Fail and Withdrawn groups).

There are also differences in density and probability of resource use among the performance groups. Figure 5.4 compares the density of resource use among the performance groups. The distinction and the pass tend to use learning resources more frequently than the fail and the withdrawn. Other research has also mentioned so. It is subtle to notice that the distinction and the pass access to resources that help them plan their learning journey, e.g., homepage or content, more frequently than their peers of the fail or the withdrawn. Combined with resource use density, steady states of resource use represent habits. Generally, the steady states of resource use do not vary greatly among the performance groups. However, there is a sign in the steady states to distinguish performance groups. In Figure 5.5, the distinction and the pass look like having a habit of using the forum and quiz resources. It can be implied that the pass and the distinction tend to be active and self-regulating.

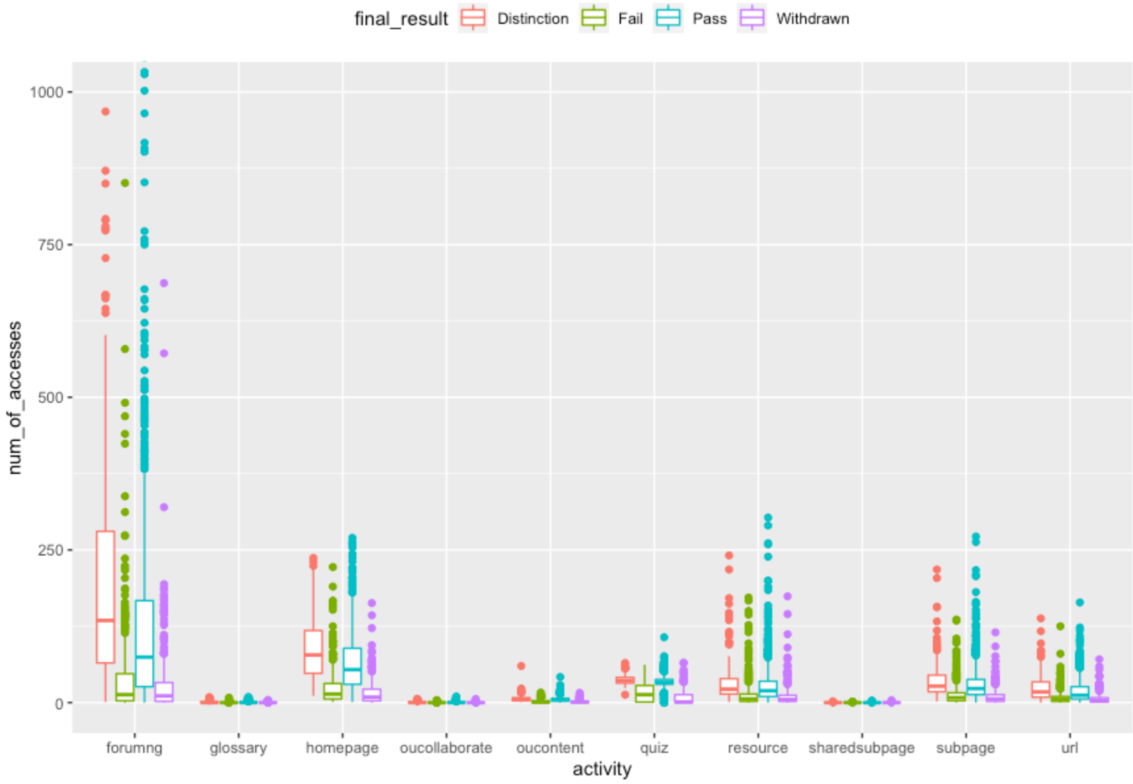


Figure 5.4. Comparison of the number of accesses to resources among performance groups. The pass learners (of Distinction and Pass groups) have higher resource use density than the fail learners (of Fail and Withdrawn groups). It is noticeable that the pass learners access the *homepage* more than the fail learners. The *homepage* usually contains a course outline, description, learning materials, and assignments. It can be implied from the density of access to the *homepage* that the pass learners are more active, better managed, and better prepared than the fail learners.

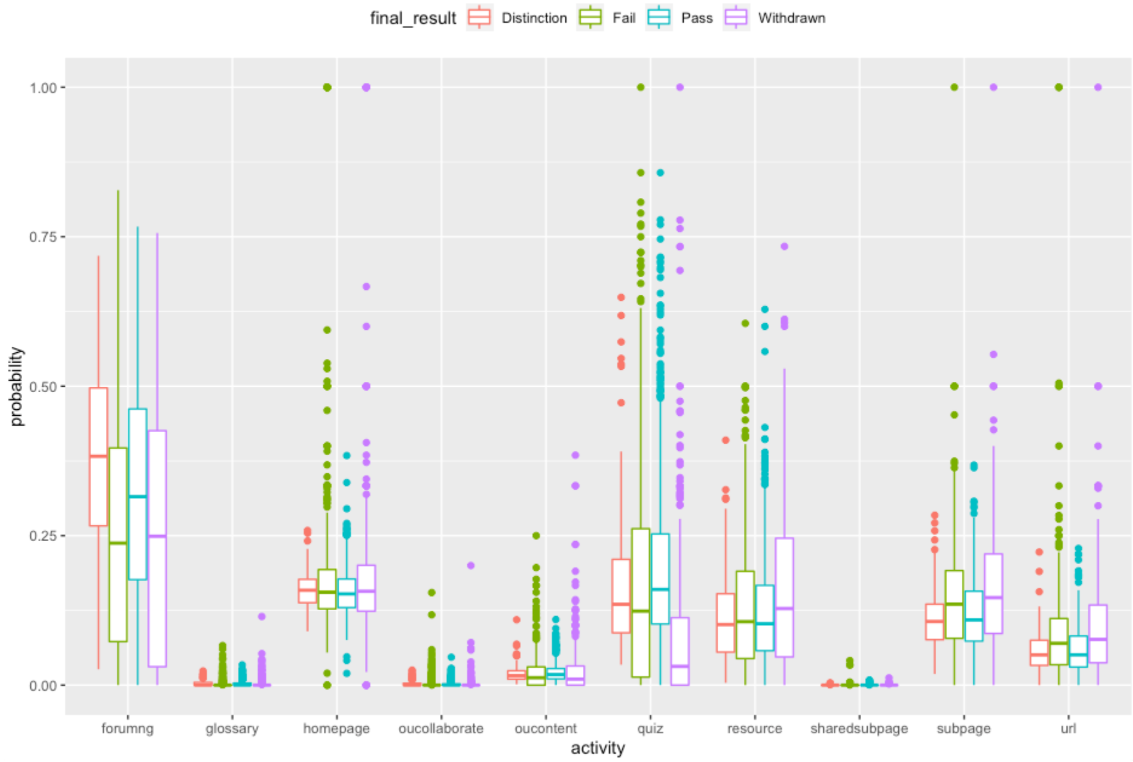


Figure 5.5. Comparison of steady states of resource use among the performance groups. There is not much difference in the steady state among the student groups.

5.3. Prediction

We come to a point to test the performance of prediction models built out of SRL profiles on OULAD data. We choose the overall SRL profile, computed from resource access throughout a whole course, to build the prediction model. The prediction model is trained on only one course presentation and then predicts learners' learning performance in other courses.

5.3.1. Datasets

Let's briefly review the OULAD courses, which are going through the training and prediction evaluation. OULAD contains 7 courses named AAA, BBB, CCC, DDD, EEE, FFF, and GGG; and was offered multiple times as a course offering. A course offering is denoted with a year the course was offered and a letter J or B representing whether the course offering is in the autumn or spring. The author of OULAD [12] recommends analyzing course offerings of J type separately from B type due to their structure difference; however, we opt to consider all of the course offerings equally. The reason for such consideration is this. Regardless of resource content, the resource types are similar from one course offering to another. And the resource uses are measured by resource access only, and such access is synthesized and abstracted into the SRL index, which, as described above, can represent SRL ability in only learning given resource use and its effectiveness.

Among the courses, the course GGG was not used due to its lack of score information; therefore, the SRL index could not be synthesized. Table 5.1 shows the course offerings for training and testing the prediction models.

Table 5.1. Course offerings in OULAD

No	Course module	Code presentation	Length (days)	Number of learners
1	AAA	2013J	268	383
2	AAA	2014J	269	365
3*	BBB	2013J	268	2237
4	BBB	2014J	262	2292
5	BBB	2013B	240	1767
6	BBB	2014B	234	1613
7	CCC	2014J	269	2498
8	CCC	2014B	241	1936
9	DDD	2013J	261	1938
10	DDD	2014J	262	1803
11	DDD	2013B	240	1303
12	DDD	2014B	241	1228
13	EEE	2013J	268	1052
14	EEE	2014J	269	1188
15	EEE	2014B	241	694
16	FFF	2013J	268	2283
17	FFF	2014J	269	2365
18	FFF	2013B	240	1614
19	FFF	2014B	241	1500

*Course offering BBB 2013J is used for training the prediction model

Among the course offerings, the BBB 2013J was used for training prediction models. BBB 2013J was selected for training because it has an adequate sample size and balances the number of learners of final result types. The trained models were then applied to the other course offerings to predict the final results. The predicted final results were then compared with the ground truth final results for the model evaluation.

5.3.2. Generation of datasets for model training and testing

Datasets used for model training and testing should contain data points in a context of a course, comprising 2 features and 1 target label. The features are the log SRL index representing a learner's SRL ability and the sum of cognitive scores the learner earns. The target label is one of the final results, such as pass, fail, withdrawn, or dropout.

To generate such datasets, we propose the following procedure:

1. Produce a data table of resource use in which the columns are the resources, the rows are the resource use history of a learner, and the cells contain the number of accesses to each resource. The resource use data table can contain resource use for a whole course or parts of a course, such as a period before or after exams. The structure of a resource use data table is shown in Table 5.2.
2. Produce a data table of cognitive scores in which the columns are the assessments, each row is about a learner's learning performance, and the cells contain the scores the learner earns in corresponding assessments. The cognitive score data table contains an extra column about a learner's final result, e.g., pass, fail, withdrawn. The structure of a cognitive score data table is shown in Table 5.3, and a sample table is shown in Figure 5.7.
3. For each learner, compute a transition probability matrix (a sample matrix is shown in Figure 5.8).
4. Use the transition probability matrices to make a Markov chain for each matrix.
5. Generate vectors of steady states (a sample table of steady states is shown in Figure 5.9) from each Markov chain.

6. Compute the SRL index table (a sample table is shown in Figure 5.10) by using the resource use data table, cognitive score data table, and vectors of steady states. The structure of an SRL index data table is shown in Table 5.4.

Table 5.2. The structure of resource use data table.

Learner	**Resource 1**	**Resource 2**	**…**	**Resource n-1**	**Resource n**
ID	Number of access	Number of access	…	Number of access	Number of access

Table 5.3. The structure of cognitive scores data table.

Learner	**Assessment 1**	**Assessment 2**	**…**	**Assessment k-1**	**Assessment k**
ID	Number of access	Number of access	…	Number of access	Number of access

Table 5.4. The structure of the SRL index data table.

Learner	**SRL index**	**Log SRL index**	**Cognitive Score**	**Final results**
ID	Number	Number	Number	Pass / Fail / Withdrawn

In this research, we implemented the procedure above using the R language on OULAD. The structure of each OULAD file is provided in Appendix 1. Structures of OULAD file. The implemented source code is provided in Appendix 2. Source code in R for Dataset generation.

From OULAD, these data tables are generated by applying the procedure and synthesizing the following data.

To generate the **resource use data table** (a sample table is shown in Figure 5.6), data from the studentVle.csv, vle.csv, and assesments.csv are synthesized. To generate the **cognitive score data table** (a sample table is shown in Figure 5.7), data from studentInfo.csv, assessments.csv, studentAssessment.csv. To generate the **transition probability matrices** (a sample table is shown in Figure 5.8), data from the studentInfo.csv, studentVle.csv, vle.csv, and assessments.csv are synthesized and computed using the algorithm in Table 4.3. The transition probability matrix is the input for producing Markov models. And the **steady states data table** is generated from the

Markov models, as shown in Figure 5.9. The **SRL index data table** is computed using equations *(4.4)* and *(4.5)* above.

id_student	forumng	glossary	homepage	oucollaborate	oucontent	quiz	resource	sharedsubpage	subpage	url
23798	76	1	77	3	6	48	16	0	33	46
27759	47	2	46	0	6	32	17	0	24	8
30091	120	0	64	0	8	36	33	0	34	48
31014	99	0	47	1	1	14	6	0	5	3
31849	194	0	80	0	8	33	14	0	26	10
37622	62	0	39	0	2	28	16	0	21	12
38234	6	0	4	1	1	18	2	0	7	2
47855	0	0	6	0	0	0	0	0	0	2
48040	225	2	158	0	6	35	184	0	161	60
48503	61	0	36	1	4	40	7	0	14	5
51301	154	1	47	1	3	0	21	0	20	13
52797	174	4	60	3	8	37	16	0	26	12
52899	140	1	120	0	14	32	102	1	70	27
53360	2	1	3	0	1	0	0	0	1	3
54388	3	0	1	0	1	0	0	0	1	1
55026	23	0	26	1	2	25	27	0	10	7
56340	1	0	1	0	0	0	0	0	0	0
56789	20	0	33	1	4	30	29	0	26	8
57079	82	2	61	0	8	47	53	0	65	28

Figure 5.6. Resource use data table. Column names are the resources available in a course; each row contains the number of access for each resource performed by individual learners.

id_student	14996	14997	14998	14999	15000	15001	15003	15004	15005	15006	15007	total_score	final_result
23798	4.50	16.20	16.02	15.84	15.66	16.02	1.0	1.0	1.0	1.0	1.0	89.24	Distinction
27759	3.05	11.70	10.98	8.64	12.06	9.90	0.8	1.0	1.0	1.0	1.0	61.13	Fail
30091	4.00	12.96	13.50	0.00	14.04	13.50	1.0	1.0	1.0	1.0	1.0	63.00	Pass
31014	4.25	14.94	14.76	0.00	0.00	0.00	1.0	1.0	0.0	0.0	0.0	35.95	Withdrawn
31849	4.05	12.60	11.16	11.34	11.70	13.86	1.0	1.0	0.8	0.6	1.0	69.11	Pass
37622	3.50	13.50	12.60	11.16	13.86	11.70	0.0	1.0	1.0	1.0	1.0	70.32	Pass
38234	3.50	13.86	0.00	13.14	11.52	0.00	0.8	0.0	0.8	1.0	0.0	44.62	Fail
47855	2.05	7.92	7.56	9.54	8.10	7.20	0.6	0.8	0.4	0.4	0.2	44.77	Withdrawn
48040	3.00	10.08	12.06	11.52	10.62	11.52	1.0	1.0	0.8	1.0	1.0	63.60	Pass
48503	3.80	12.78	13.50	0.00	12.96	13.32	1.0	0.8	0.6	0.6	1.0	60.36	Pass
51301	4.65	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	4.65	Withdrawn
52797	4.35	12.60	14.22	12.96	15.30	16.20	1.0	0.8	1.0	1.0	1.0	80.43	Pass
52899	3.65	11.52	11.34	11.88	0.00	0.00	1.0	1.0	1.0	1.0	0.0	42.39	Pass
55026	2.95	11.88	11.52	11.16	10.62	0.00	1.0	0.0	1.0	0.6	0.8	51.53	Pass
56789	4.25	13.86	13.50	7.38	13.14	12.60	0.8	0.8	1.0	1.0	0.6	68.93	Pass
57079	3.70	10.98	11.70	10.08	12.24	0.00	1.0	0.8	1.0	1.0	0.4	52.90	Fail
63044	3.25	11.70	9.00	0.00	0.00	0.00	0.8	0.6	0.0	0.0	0.0	25.35	Fail

Figure 5.7. Cognitive score data table. Column names are the assessments, total scores, and final results; each row contains individual learners’ partial scores, total scores, and final results.

```
23798-Distinction
 A  9 - dimensional discrete Markov Chain defined by the following states:
 forumng, glossary, homepage, oucollaborate, oucontent, quiz, resource, subpage, url
 The transition matrix  (by rows)  is defined as follows:
                 forumng glossary  homepage oucollaborate  oucontent       quiz   resource    subpage       url
forumng       0.04761905        0 0.3809524    0.00000000 0.04761905 0.04761905 0.04761905 0.19047619 0.2380952
glossary      0.00000000        0 0.5000000    0.00000000 0.00000000 0.50000000 0.00000000 0.00000000 0.0000000
homepage      0.10666667        0 0.4133333    0.01333333 0.02666667 0.09333333 0.01333333 0.09333333 0.2400000
oucollaborate 0.00000000        0 0.5000000    0.00000000 0.00000000 0.50000000 0.00000000 0.00000000 0.0000000
oucontent     0.00000000        0 0.4000000    0.00000000 0.20000000 0.00000000 0.00000000 0.20000000 0.2000000
quiz          0.15000000        0 0.4000000    0.00000000 0.05000000 0.20000000 0.00000000 0.05000000 0.1500000
resource      0.10000000        0 0.4000000    0.00000000 0.00000000 0.10000000 0.10000000 0.20000000 0.1000000
subpage       0.10000000        0 0.3500000    0.05000000 0.05000000 0.05000000 0.05000000 0.20000000 0.1500000
url           0.12820513        0 0.4358974    0.02564103 0.00000000 0.07692308 0.00000000 0.02564103 0.3076923
```

Figure 5.8. A transition probability matrix and Markov model.

id_student	forumng	glossary	homepage	oucontent	ouelluminate	quiz	resource	sharedsubpage	subpage	url
23629	0.19071517	0.0000000000	0.18865352	0.0000000000	0.000000000	0.566083728	0.027273789	0.0000000000	0.02727379	0.000000000
25107	0.67680879	0.0043284977	0.15220238	0.0012452194	0.001692736	0.079779282	0.034074955	0.0000000000	0.02510804	0.024760104
27891	0.50094279	0.0000000000	0.13584736	0.0179433183	0.000000000	0.147024822	0.082805065	0.0000000000	0.07493328	0.040503367
29144	0.20379831	0.0057369639	0.09493675	0.0029857416	0.000000000	0.041153962	0.349167299	0.0000000000	0.26462951	0.037591457
31663	0.58043916	0.0023580734	0.11686801	0.0006119822	0.008850918	0.022647122	0.137082795	0.0000000000	0.10582544	0.025316504
33666	0.70863226	0.0000000000	0.16052809	0.0028949354	0.004617000	0.023221643	0.013752680	0.0000000000	0.05390064	0.032452753
34229	0.26416838	0.0149050911	0.11835917	0.0000000000	0.000000000	0.087601641	0.218548216	0.0000000000	0.23260313	0.063814369
34353	0.42014097	0.0000000000	0.16861296	0.0000000000	0.000000000	0.208126105	0.064930078	0.0000000000	0.12676734	0.011422542
34431	0.40139815	0.0000000000	0.15349933	0.0000000000	0.000000000	0.201935766	0.011804895	0.0000000000	0.12615499	0.105206878
34756	0.51302283	0.0000000000	0.11601145	0.0000000000	0.000000000	0.165933033	0.061855505	0.0000000000	0.09177075	0.051406424
34863	0.18947753	0.0000000000	0.13412097	0.0126057234	0.000000000	0.447274340	0.039681265	0.0000000000	0.11093521	0.065904966
35812	0.56603774	0.0000000000	0.20754717	0.0000000000	0.000000000	0.000000000	0.000000000	0.0000000000	0.11320755	0.113207547
39655	0.40009014	0.0000000000	0.11601395	0.0107441361	0.000000000	0.240248147	0.086299363	0.0000000000	0.08474956	0.061854706
50993	0.43429037	0.0000000000	0.13953067	0.0000000000	0.000000000	0.301703616	0.013938939	0.0000000000	0.03282168	0.077714727
52899	0.45681625	0.0000000000	0.10279626	0.0080503704	0.004366090	0.064754533	0.126209402	0.0000000000	0.16256521	0.074441890
58089	0.30743587	0.0000000000	0.11596683	0.0000000000	0.000000000	0.108744701	0.274980868	0.0000000000	0.16493412	0.027937612
62631	0.56671487	0.0000000000	0.16820472	0.0087844887	0.000000000	0.136702785	0.044818595	0.0000000000	0.03999048	0.034784059
70608	0.17196010	0.0019844967	0.13017574	0.0000000000	0.000000000	0.290621555	0.120763660	0.0000000000	0.24470118	0.039793269

Figure 5.9. Steady states data table. Column names are resources available in a course; each row contains the probability values about a resource used by a learner to start a learning sequence.

srl_index	cognitive_score	log_srl_index	final_result
253.53106	16.69	7.986019	Fail
3651.43207	60.51	11.834247	Pass
498.76257	27.81	8.962209	Withdrawn
1163.61911	53.02	10.184403	Fail
4222.62279	63.92	12.043924	Pass
1011.23598	17.94	9.981904	Withdrawn
1045.98268	65.07	10.030643	Pass
1141.39523	65.73	10.156583	Fail
11.50871	0.80	3.524655	Withdrawn
219.43298	13.53	7.777637	Withdrawn
976.63224	63.22	9.931672	Pass
42.64711	3.55	5.414376	Fail
1404.15863	82.98	10.455490	Distinction
346.61324	22.75	8.437183	Fail
80.69862	4.40	6.334472	Withdrawn
62.01520	4.75	5.954550	Fail
1009.63641	43.99	9.979620	Fail
1134.03453	77.95	10.147249	Distinction
0.00000	0.00	0.000000	Fail
2669.53958	75.34	11.382375	Distinction

Figure 5.10. SRL index data table

5.3.3. Pass and Fail prediction

Prepared the datasets for the pass-fail prediction, the final result attribute of data points were relabeled. The *pass* and the *distinction* were relabeled as 'pass'. The *fail* and the *withdrawn* were relabeled as 'fail'.

As demonstrated in Table 5.5, the SRL index distinguishes the pass learners from the fail learners clearly. Pass learners have a high SRL index than fail learners. And learners with a higher SRL index tend to achieve higher cognitive scores than learners with a lower SRL index, as illustrated by the correlation between SRL indices and cognitive scores in Table 5.6.

Table 5.5. Boxplot of log SRL indices of Pass and Fail learners of each course offering.

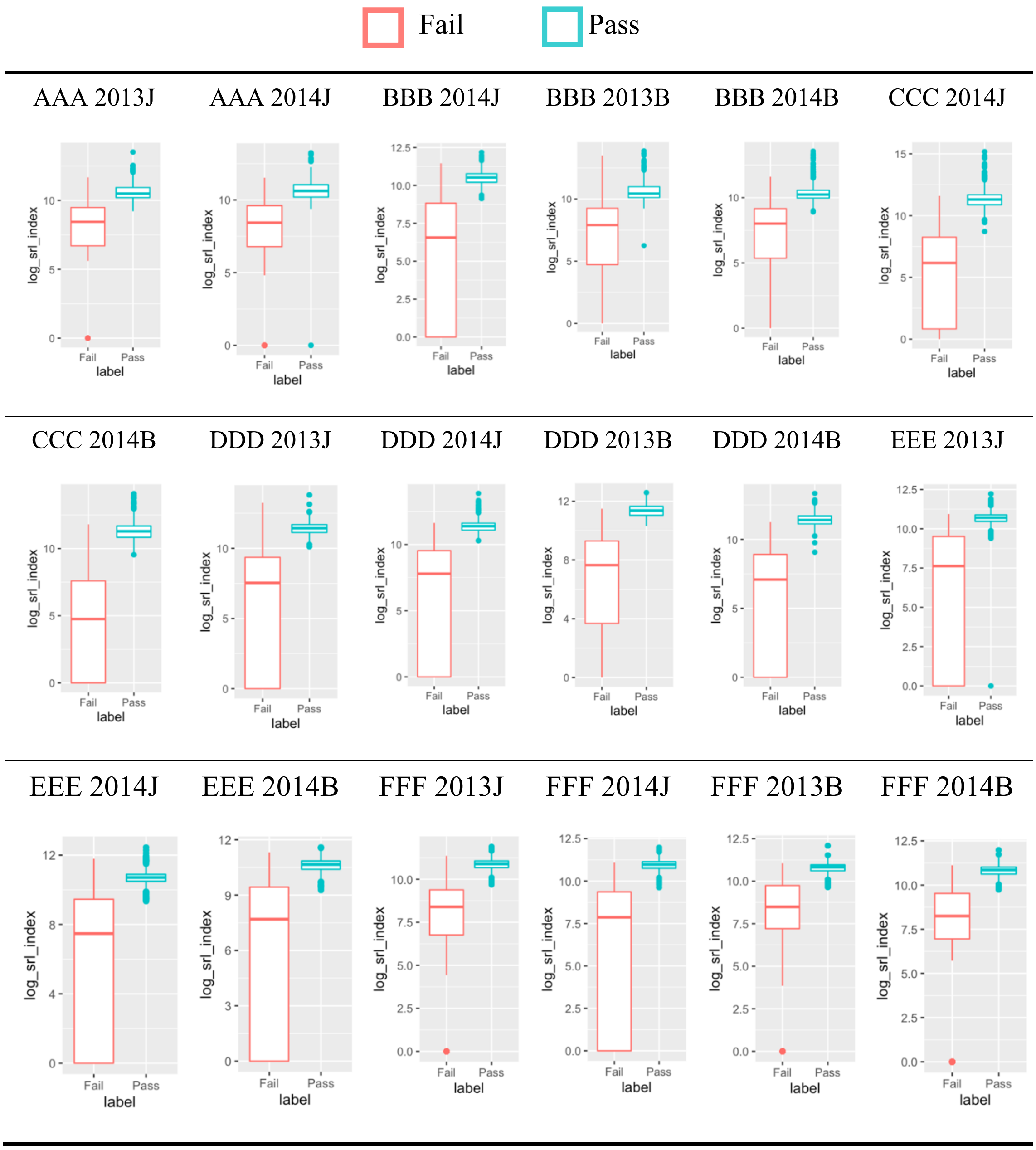
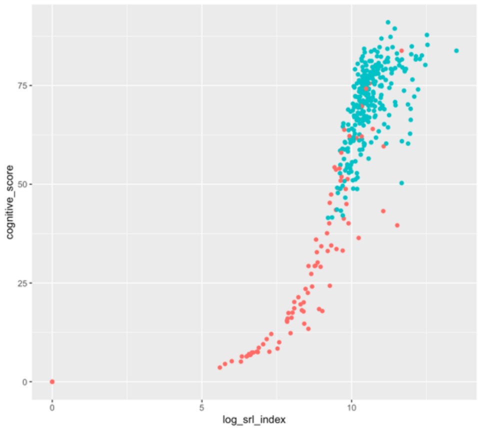


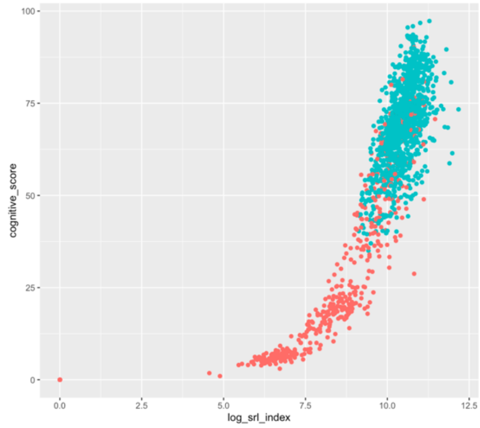
Table 5.6. Scatterplot of correlation between log SRL indices and cognitive scores of learners in several course offerings.

Fail Pass

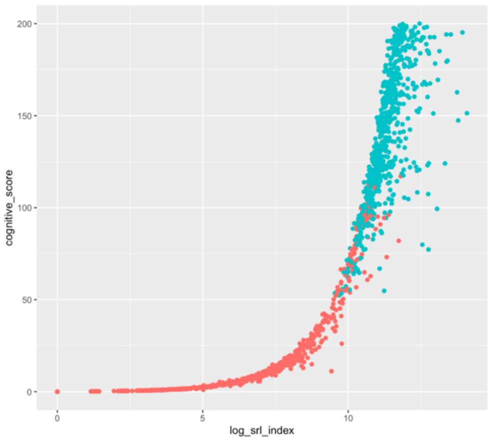
AAA 2013J



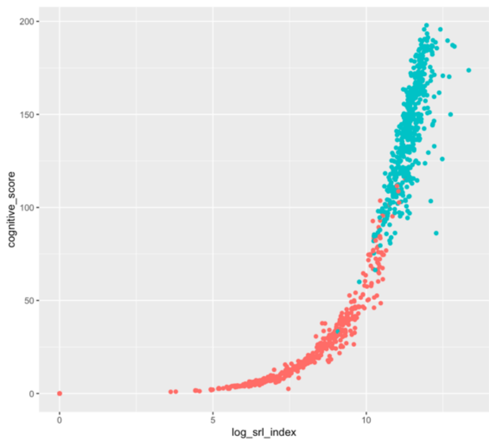
BBB 2014J



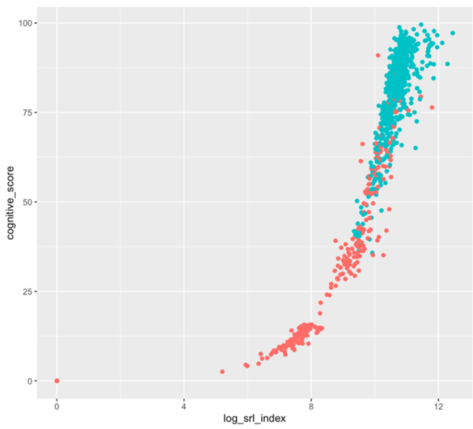
CCC 2014B



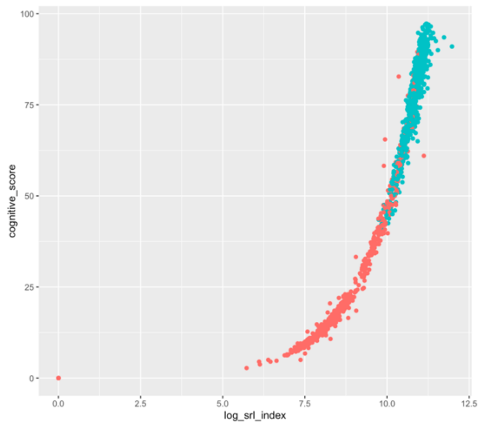
DDD 2014B



EEE 2014J



FFF 2014B



For the prediction performance, the models trained by Support vector classification (SVC) and Support vector machine with the linear kernel (SVM) [73], [74] yielded the highest accuracies from 87% to 93% on each course offering. Using only one course offering for training, the models with such accuracies have stability and generalization for applying to other datasets. Table 5.7 presents the accuracies of the models on each course offerings.

Table 5.7. Accuracy of the pass and fail prediction models on each course offering.

No	Course	Overall accuracy	Labels	Precision	Recall	F1-score	Support
1	AAA 2013J	0.93 (SVC)	Pass	0.92	0.99	0.95	278
			Fail	0.96	0.76	0.85	100
2	AAA 2014J	0.92 (SVC)	Pass	0.91	0.99	0.95	253
			Fail	0.98	0.76	0.85	104
3	BBB 2014J	0.93 (SVC)	Pass	0.91	0.99	0.95	1150
			Fail	0.98	0.85	0.91	771
4	BBB 2013B	0.91 (SVC)	Pass	0.86	0.99	0.92	803
			Fail	0.99	0.83	0.90	734
5	BBB 2014B	0.91 (SVC)	Pass	0.86	0.99	0.92	727
			Fail	0.98	0.80	0.88	567
6	CCC 2014J	0.93 (SVM)	Pass	0.86	1.00	0.93	1014
			Fail	1.00	0.87	0.93	1288
7	CCC 2014B	0.94 (SVM)	Pass	0.87	1.00	0.93	663
			Fail	1.00	0.91	0.95	1018
8	DDD 2013J	0.89 (SVM)	Pass	0.81	1.00	0.90	829
			Fail	1.00	0.80	0.89	939
9	DDD 2014J	0.87 (SVM)	Pass	0.79	1.00	0.88	792
			Fail	1.00	0.76	0.86	855
10	DDD 2013B	0.90 (SVM)	Pass	0.81	1.00	0.90	510
			Fail	1.00	0.83	0.91	704
11	DDD 2014B	0.93 (SVM)	Pass	0.86	1.00	0.93	479
			Fail	1.00	0.88	0.94	637

No	Course	Overall accuracy	Labels	Precision	Recall	F1-score	Support
12	EEE 2013J	0.93 (SVM)	Pass	0.91	0.99	0.94	609
			Fail	0.97	0.83	0.89	355
13	EEE 2014J	0.92 (SVM)	Pass	0.90	0.98	0.94	684
			Fail	0.96	0.82	0.89	413
14	EEE 2014B	0.91 (SVM)	Pass	0.88	0.97	0.92	357
			Fail	0.96	0.82	0.88	267
15	FFF 2013J	0.93 (SVM)	Pass	0.89	0.98	0.93	1095
			Fail	0.98	0.87	0.92	1003
16	FFF 2014J	0.92 (SVM)	Pass	0.89	0.98	0.93	1117
			Fail	0.98	0.86	0.91	1004
17	FFF 2013B	0.89 (SVM)	Pass	0.83	0.98	0.90	782
			Fail	0.97	0.79	0.87	728
18	FFF 2014B	0.91 (SVM)	Pass	0.86	0.98	0.92	654
			Fail	0.98	0.85	0.91	709

5.3.4. Dropout prediction

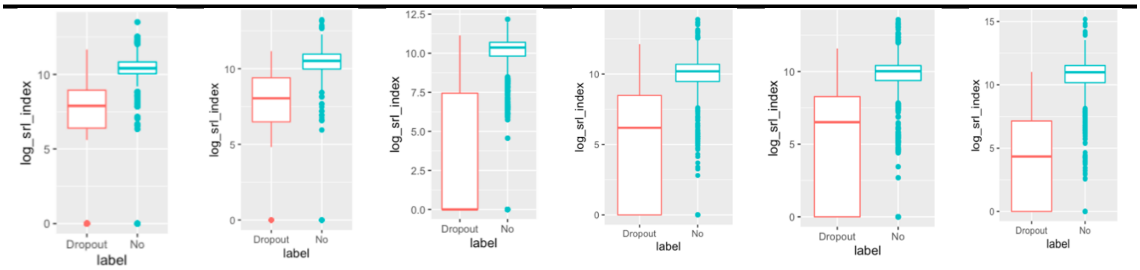
Prepared the datasets for dropout prediction, the final result attribute of data points were relabeled. The *pass*, the *distinction*, and the *fail* were relabeled as ‘no dropout’ (ND). The *withdrawn* were relabeled as ‘dropout’ (D).

As Table 5.8 shows SRL indices of the dropout and the not dropouts in each course offering, it is apparent that the SRL index helps distinguish the dropouts and the not-dropouts effectively.

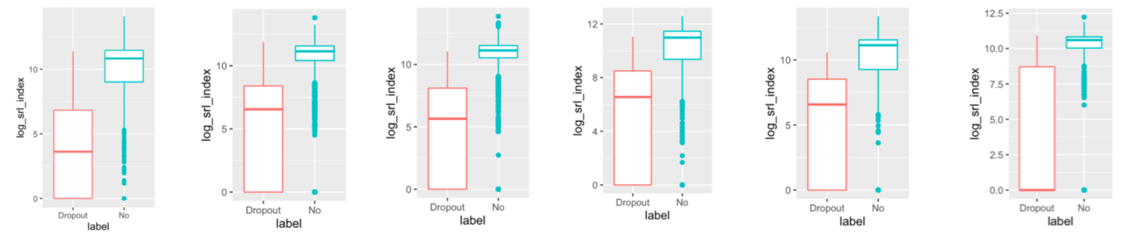
K Nearest Neighbors (Knn) [75] and Gaussian Naïve Bayes (GNB) [76] produced predictions with accuracies from 77% up to 90%, as presented in Table 5.9.

Table 5.8. Boxplot of log SRL indices of Dropout and No dropout learners of each course offering.

Dropout	No dropout				
AAA 2013J	AAA 2014J	BBB 2014J	BBB 2013B	BBB 2014B	CCC 2014J



CCC 2014B	DDD 2013J	DDD 2014J	DDD 2013B	DDD 2014B	EEE 2013J



EEE 2014J	EEE 2014B	FFF 2013J	FFF 2014J	FFF 2013B	FFF 2014B

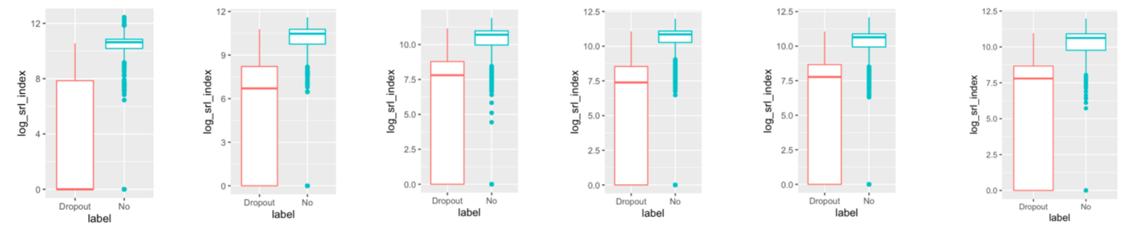


Table 5.9. Accuracy of the dropout prediction models on each course offering.

No	Course	Overall accuracy	Labels	Precision	Recall	F1-score	Support
1	AAA 2013J	0.90 (KNN)	D	0.65	0.64	0.64	55
			ND	0.94	0.94	0.94	323
2	AAA 2014J	0.87 (GNB)	D	0.62	0.45	0.52	58
			ND	0.90	0.95	0.92	299
3	BBB 2014J	0.87 (GNB)	D	0.66	0.77	0.71	403
			ND	0.94	0.89	0.91	1518
4	BBB 2013B	0.82 (GNB)	D	0.55	0.59	0.57	320
			ND	0.89	0.87	0.88	1217
5	BBB 2014B	0.85 (GNB)	N	0.51	0.58	0.54	206
			ND	0.92	0.90	0.91	1888
6	CCC 2014J	0.86 (GNB)	N	0.83	0.81	0.82	890
			ND	0.88	0.89	0.89	1412

No	Course	Overall accuracy	Labels	Precision	Recall	F1-score	Support
7	CCC 2014B	0.82 (GNB)	N	0.73	0.85	0.79	663
			ND	0.89	0.80	0.84	1018
8	DDD 2013J	0.81 (GNB)	N	0.69	0.64	0.67	525
			ND	0.85	0.88	0.87	1243
9	DDD 2014J	0.85 (GNB)	N	0.79	0.67	0.72	497
			ND	0.87	0.93	0.89	1150
10	DDD 2013B	0.79 (GNB)	D	0.64	0.65	0.65	357
			ND	0.85	0.85	0.85	857
11	DDD 2014B	0.77 (GNB)	D	0.69	0.64	0.66	388
			ND	0.82	0.84	0.83	728
12	EEE 2013J	0.86 (GNB)	D	0.59	0.66	0.63	168
			ND	0.93	0.90	0.92	796
13	EEE 2014J	0.88 (GNB)	D	0.71	0.70	0.70	228
			ND	0.92	0.93	0.92	869
14	EEE 2014B	0.83 (GNB)	D	0.53	0.62	0.57	112
			ND	0.91	0.88	0.90	512
15	FFF 2013J	0.81 (GNB)	D	0.61	0.53	0.57	497
			ND	0.86	0.90	0.88	1601
16	FFF 2014J	0.86 (GNB)	D	0.81	0.67	0.73	618
			ND	0.87	0.93	0.90	1503
17	FFF 2013B	0.83 (GNB)	D	0.60	0.52	0.56	316
			ND	0.88	0.91	0.89	1194
18	FFF 2014B	0.79 (GNB)	D	0.57	0.53	0.55	337
			ND	0.85	0.87	0.86	1026

D: Dropout ND: No dropout

5.3.5. Pass / Fail / Withdrawn prediction

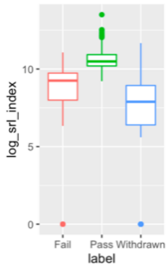
Prepared the datasets for pass/fail/withdrawn prediction, the final result attribute of data points was relabeled. Pass and distinction were relabeled as ‘pass’ (P). The *fail* were relabeled as ‘fail’ F. And the *withdrawn* were relabeled as ‘withdrawn’ (W).

For pass / fail / withdrawn prediction, the SRL index also clearly separates the pass from the fail and the withdrawn, as shown in Table 5.10. The SRL indices of the fail and the withdrawn mingle. That causes the accuracy of the prediction models of this prediction type to be not as high as the two types above. The accuracies gain from 67% to 83% with training algorithms KNN, GNB, SVC, and SVM. It can be seen in Table 5.11 that the cause of not high accuracy is the misclassification between the fail and the withdrawn.

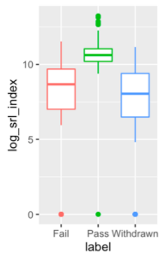
Table 5.10. Boxplot of log SRL indices of Pass / Fail /Withdrawn learners of each course offering.

Fail   Pass   Withdrawn

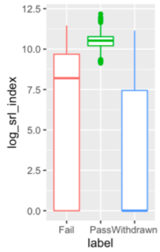
AAA 2013J



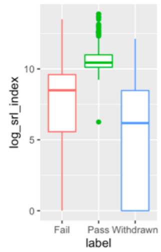
AAA 2014J



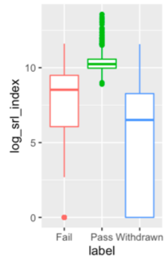
BBB 2014J



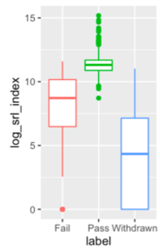
BBB 2013B



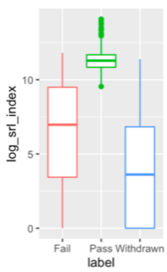
BBB 2014B



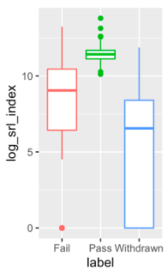
CCC 2014J



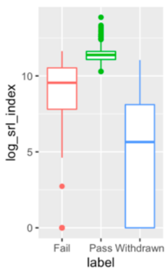
CCC 2014B



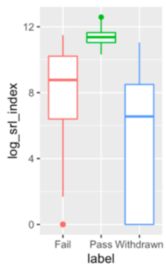
DDD 2013J



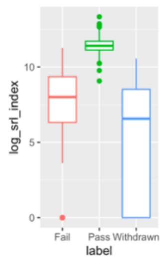
DDD 2014J



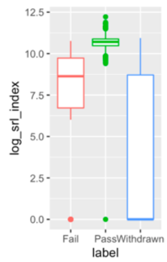
DDD 2013B



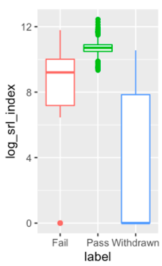
DDD 2014B



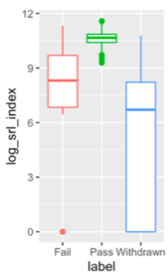
EEE 2013J



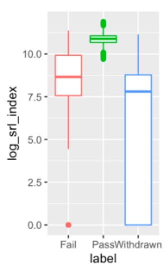
EEE 2014J



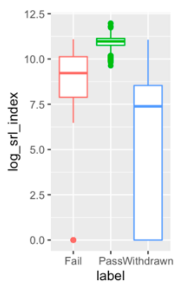
EEE 2014B



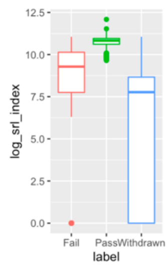
FFF 2013J



FFF 2014J



FFF 2013B



FFF 2014B

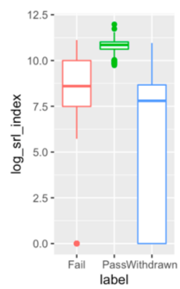


Table 5.11. Accuracy of the pass/fail/withdrawn prediction models on each course offering.

No	Course	Overall accuracy	Labels	Precision	Recall	F1-score	Support
1	AAA 2013J	0.83 (GNB)	P	0.92	0.99	0.95	278
			F	0.35	0.36	0.35	45
			W	0.74	0.42	0.53	55
2	AAA 2014J	0.83 (GNB)	P	0.91	1.00	0.95	253
			F	0.45	0.41	0.43	46
			W	0.63	0.41	0.50	58
3	BBB 2014J	0.81 (GNB)	P	0.91	0.99	0.95	1150
			F	0.52	0.31	0.39	368
			W	0.66	0.74	0.70	403
4	BBB 2013B	0.74 (GNB)	P	0.86	0.98	0.92	803
			F	0.55	0.40	0.46	414
			W	0.55	0.56	0.56	320
5	BBB 2014B	0.76 (GNB)	P	0.87	0.98	0.92	727
			F	0.61	0.44	0.51	361
			W	0.52	0.53	0.53	206
6	CCC 2014J	0.74 (SVC)	P	0.85	1.00	0.92	1014
			F	0.33	0.42	0.37	398
			W	0.88	0.59	0.71	890
7	CCC 2014B	0.72 (SVC)	P	0.87	1.00	0.93	663
			F	0.36	0.37	0.37	355
			W	0.76	0.63	0.69	663
8	DDD 2013J	0.68 (SVM)	P	0.80	1.00	0.89	829
			F	0.37	0.42	0.40	414
			W	0.75	0.38	0.51	525
9	DDD 2014J	0.72 (KNN)	P	0.79	1.00	0.88	792
			F	0.38	0.32	0.35	358
			W	0.82	0.57	0.67	497
10	DDD 2013B	0.67(SVM)	P	0.80	1.00	0.89	510
			F	0.46	0.51	0.48	347
			W	0.71	0.36	0.48	357
11	DDD 2014B	0.71 (KNN)	P	0.87	1.00	0.93	479
			F	0.40	0.46	0.43	249
			W	0.72	0.52	0.60	388
12	EEE 2013J	0.81(SVC)	P	0.89	0.99	0.94	609
			F	0.58	0.46	0.51	187
			W	0.67	0.55	0.61	168
13	EEE 2014J	0.81 (KNN)	P	0.91	0.97	0.94	684
			F	0.48	0.39	0.43	185
			W	0.72	0.68	0.70	228

No	Course	Overall accuracy	Labels	Precision	Recall	F1-score	Support
14	EEE 2014B	0.75 (KNN)	P	0.88	0.96	0.92	357
			F	0.52	0.37	0.44	155
			W	0.52	0.58	0.55	112
15	FFF 2013J	0.72 (KNN)	P	0.88	0.95	0.91	1095
			F	0.45	0.37	0.40	506
			W	0.57	0.58	0.58	497
16	FFF 2014J	0.77 (KNN)	P	0.89	0.96	0.93	1117
			F	0.42	0.39	0.41	386
			W	0.74	0.67	0.70	618
17	FFF 2013B	0.73 (KNN)	P	0.84	0.96	0.89	782
			F	0.55	0.37	0.44	412
			W	0.61	0.65	0.63	316
18	FFF 2014B	0.72 (KNN)	P	0.86	0.96	0.91	654
			F	0.51	0.37	0.43	372
			W	0.59	0.64	0.61	337

5.4. Method evaluation

We justified our proposed learner model by developing learning performance prediction models from the SRL index and assessment scores, then compared the prediction accuracies with current prediction models. Two state-of-the-art works with which we compare the accuracy of our proposed method are these.

- Hao et al. [37] analyzed the sum of clicks on resources and average assessments scores to predict pass/fail final results on OULAD, with 0.8 training : 0.2 testing data ratio on the whole dataset, and gained 93% accuracy;
- and Qiu et al. [33] classified and grouped resource types used by each learner in a course to predict pass/fail final results on OULAD, with 0.7 training : 0.3 testing data ratio on the whole dataset, and gained 97% accuracy.

Applying the method to OULAD, we generated SRL profiles for individual learners from their resource access throughout a whole course, computed the SRL indices, and used the SRL indices and accumulative assessment scores as features to build the machine learning models to predict the final results of pass or fail. Two approaches for splitting training and testing datasets were carried out. The first approach is that the prediction model is trained on only one course presentation and then predicts learners' learning performance in other courses. The second approach is to split each course offering into a 0.5 training : 0.5 testing data ratio.

Table 5.12 presents the statistical values of the accuracy of our prediction models. For the prediction performance, the model was trained by the Support vector classification (SVC) and the Support vector machine with the linear kernel (SVM) (Pedregosa et al., 2011). The first approach yielded accuracy from **87% to 93%** on each course offering, with the **maximum accuracy of 93%**, equivalent to the state-of-the-art work of Hao et al. [37]. Using only one course offering for training, the models with such accuracies have stability and generalization for applying to other datasets. Following the second approach, the prediction model gained very high accuracy from **91% to 97%** on each course offering with the **maximum accuracy of 97%**, equivalent to the state-of-the-art work of Qiu et al. [33]. Detailed accuracies that our prediction models yielded on the whole OULAD are presented in Table 5.13 for the first approach and Table 5.14 for the second approach.

Table 5.12. Mean, maximum, and minimum values with standard deviation (SD) of the accuracy and F1-score of the prediction models developed in two approaches.

Approach	**Accuracy**			**F1-Score**		
	Mean± SD	**Max**	**Min**	**Mean ± SD**	**Max**	**Min**
First approach: prediction model trained by using linear SVM on one course offering and tested on the others	0.92 ± 0.02	**0.94**	0.87	0.90 ± 0.03	0.96	0.87
Second approach: prediction model trained by SVC on 0.5 training : 0.5 testing data ratio for each course offering	0.93 ± 0.02	**0.97**	0.91	0.93 ± 0.03	0.97	0.89
State-of-the-art						
Hao et al. [37]		**0.93**			0.88	
Qiu et al. [33]		**0.97**			0.98	

Table 5.13. Accuracies from Pass/Fail prediction models using the first approach

No	Course	Overall accuracy	Precision (Pass / Fail)	Recall (Pass / Fail)	F1-score (Pass / Fail)	Support (Pass / Fail)
1	AAA 2013J	0.93	0.92 / 0.96	0.99 / 0.76	0.95 / 0.85	278 / 100
2	AAA 2014J	0.92	0.91 / 0.98	0.99 / 0.76	0.95 / 0.85	253 / 104
3	BBB 2014J	0.93	0.91 / 0.98	0.99 / 0.85	0.95 / 0.91	1150 / 771
4	BBB 2013B	0.91	0.86 / 0.99	0.99 / 0.83	0.92 / 0.90	803 / 734
5	BBB 2014B	0.91	0.86 / 0.98	0.99 / 0.80	0.92 / 0.88	727 / 567
6	CCC 2014J	0.93	0.86 / 1.00	1.00 / 0.87	0.93 / 0.93	1014 / 1288
7	CCC 2014B	0.94	0.87 / 1.00	1.00 / 0.91	0.93 / 0.95	663 / 1018
8	DDD 2013J	0.89	0.81 / 1.00	1.00 / 0.80	0.90 / 0.89	829 / 939
9	DDD 2014J	0.87	0.79 / 1.00	1.00 / 0.76	0.88 / 0.86	792 / 855
10	DDD 2013B	0.90	0.81 / 1.00	1.00 / 0.83	0.90 / 0.91	510 / 704
11	DDD 2014B	0.93	0.86 / 1.00	1.00 / 0.88	0.93 / 0.93	479 / 637
12	EEE 2013J	0.93	0.91 / 0.97	0.99 / 0.83	0.94 / 0.89	609 / 355
13	EEE 2014J	0.92	0.90 / 0.96	0.98 / 0.82	0.94 / 0.89	684 / 413
14	EEE 2014B	0.91	0.88 / 0.96	0.97 / 0.82	0.92 / 0.88	357 / 267
15	FFF 2013J	0.93	0.89 / 0.98	0.98 / 0.87	0.93 / 0.92	1095 / 1003
16	FFF 2014J	0.92	0.89 / 0.98	0.98 / 0.86	0.93 / 0.91	1117 / 1004
17	FFF 2013B	0.89	0.83 / 0.97	0.98 / 0.79	0.90 / 0.87	782 / 728
18	FFF 2014B	0.91	0.86 / 0.98	0.98 / 0.85	0.92 / 0.91	654 / 709

Table 5.14. Accuracies of the Pass/Fail prediction models using the second approach

No	Course	Overall accuracy	Precision (Pass / Fail)	Recall (Pass / Fail)	F1-score (Pass / Fail)	Support (Pass / Fail)
1	AAA 2013J	0.94	0.94 / 0.95	0.99 / 0.82	0.96 / 0.88	139 / 50
2	AAA 2014J	0.93	0.92 / 0.95	0.98 / 0.78	0.95 / 0.86	127 / 52
3	BBB 2013J	0.91	0.89 / 0.95	0.96 / 0.85	0.93 / 0.89	536 / 399
4	BBB 2014J	0.93	0.90 / 0.99	0.99 / 0.84	0.95 / 0.91	575 / 386
5	BBB 2013B	0.92	0.89 / 0.96	0.97 / 0.86	0.93 / 0.91	402 / 367

No	Course	Overall accuracy	Precision (Pass / Fail)	Recall (Pass / Fail)	F1-score (Pass / Fail)	Support (Pass / Fail)
6	BBB 2014B	0.91	0.88 / 0.97	0.98 / 0.82	0.93 / 0.89	364 / 283
7	CCC 2014J	0.93	0.89 / 0.96	0.95 / 0.91	0.92 / 0.93	507 / 644
8	CCC 2014B	0.96	0.93 / 0.97	0.96 / 0.95	0.95 / 0.96	332 / 509
9	DDD 2013J	0.95	0.92 / 0.99	0.99 / 0.92	0.95 / 0.95	415 / 469
10	DDD 2014J	0.96	0.94 / 0.98	0.98 / 0.94	0.96 / 0.96	396 / 428
11	DDD 2013B	0.95	0.92 / 0.97	0.96 / 0.94	0.94 / 0.96	255 / 352
12	DDD 2014B	0.97	0.95 / 0.99	0.98 / 0.96	0.97 / 0.97	240 / 318
13	EEE 2013J	0.93	0.92 / 0.96	0.98 / 0.85	0.95 / 0.90	304 / 178
14	EEE 2014J	0.92	0.91 / 0.95	0.97 / 0.84	0.94 / 0.89	342 / 207
15	EEE 2014B	0.91	0.90 / 0.94	0.96 / 0.85	0.93 / 0.89	179 / 133
16	FFF 2013J	0.93	0.90 / 0.97	0.98 / 0.88	0.93 / 0.92	547 / 502
17	FFF 2014J	0.93	0.92 / 0.95	0.95 / 0.91	0.94 / 0.93	559 / 502
18	FFF 2013B	0.91	0.89 / 0.94	0.95 / 0.87	0.91 / 0.90	391 / 364
19	FFF 2014B	0.93	0.89 / 0.97	0.97 / 0.89	0.93 / 0.93	327 / 355

5.4.1. Prediction accuracy

Our proposed method achieves accuracy from 87% to 97% for pass-fail prediction, 77% to 90% for dropout prediction, and 67% to 80% for pass-fail-withdrawn prediction. All accuracies are measured on an analysis unit of a course offering. In pass-fail prediction, our proposed method has researched the accuracy of 97%, equivalent to that of the state-of-the-arts.

In dropout and pass-fail-withdrawn prediction, most of the miss classified cases come from distinguishing between withdrawn and fail due to their similar patterns of resource use.

The better of our proposed method is that it only needs a small proportion of data for training as long as the dataset used for training comprises sufficient classification targets or labels. Trained on only one course offering, the prediction model can be applied to other course offerings with reasonable accuracy, as presented in Table 5.7, Table 5.9, and

Table 5.11 above. These results justify the features used for training, which are the log SRL index (or SRL index) and the cognitive score, which are likely to represent SRL ability. For these features reflect SRL ability, they can be generalized to apply to other course offerings.

5.4.2. SRL ability

The learning performance of learners corresponds very well with the correlation between cognitive score and SRL index as the results described in section 5.3. The correlation justifies the relationship between cognition and metacognition in OLE [7], [62], and provides a measurable indicator of SRL ability.

5.4.3. Generalization

As we have mentioned, the structure of the dataset and cognitive score and SRL index are not constrained by the structure of OULAD. Our proposed method can be generalized to apply to data of any OLEs, provided that the data are synthesized into the proposed structure to compute the SRL index.

5.5. Discussion

The objective of this research is to support learners' understanding of their SRL ability in OLEs. There are two questions we might ask further. How do the SRL framework and the Markov chain relate to one another? How is the objective achievement via the SRL framework and the SRL profile provided?

5.5.1. The relationship between the SRL framework and the Markov chain

The SRL framework and the Markov chain harmonize to describe a learner's SRL pattern from a principle perspective and mathematical perspective. In other words, the SRL pattern is generated by the nature of the mind, and it can be represented by the Markov chain with its characteristics. Markov chains from a learner's sequences of resource use generate a transition probability matrix with its eigenvector and eigenvalue to illustrate the learner's long-term behaviors or learning habits. The eigenvector is the steady state of the transition probability matrix, and the corresponding eigenvalue is 1. The eigenvector describes habit directions, and the eigenvalue describes habit strength. Such habit attributes are a manifestation of the principles of the mind. These attributes demonstrate the density, effectiveness, and appearance of one's use of cognition and metacognition when learning in OLEs.

5.5.2. The relationship between the SRL framework and the SRL profile

According to the SRL framework, cognition and metacognition enable us to self-regulatedly learn; and we learn in order to perfect these two faculties. And we perfect these by firstly recognizing their status, next determining target goals, and next changing properly for improvement. In OLEs, there are available supports for these steps.

In OLEs, our cognition and metacognition leave their traces in learning history data. When the data are arranged in a certain manner, they assist us in recalling our SRL patterns. The SRL profile proposed in Chapter 5 is for the SRL recognition purpose. Looking at our SRL profile might help us recall resources attributing learning effectiveness, resources consuming time but not beneficial, and resources that we miss. The SRL profile also helps us recall the motive by which we access certain resources at a certain time. Such reflection triggers self-modification and adaption to the SRL pattern.

Further, the SRL pattern is a representation of SRL habits. Individual learners have their own habits of learning cognitively and metacognitively (recall the philosophical habit of the mind in the SRL framework). Some SRL habits are effective for learning, and some are not. For instance, learners with a habit of knowing a task requirement before working on a task might access the course outline at first, then access the task and use resources related to the task for preparation; finally, they do the task and earn high scores. Learners without such a habit might access resources without a proper order; hence, they might miss fulfilling the task requirements, which affects scores. When learners need an indicator of SRL ability, the SRL index is such an indicator of a union between their cognition and metacognition - a union between their SRL patterns and the corresponding performance.

5.5.3. Use of the SRL framework to explain SRL ability

Recognizing current SRL ability, we step into improvement. In this research, we did not address the improvement part because of a lack of expertise. However, the SRL framework provides paths for improvement. The paths are derived from the causes of SRL. Except for the formal cause defining what SRL is made up of, the other three causes suggest directions for improvement. The material cause suggests improvement in density and frequency of cognitive and metacognitive activities. For instance, a learner might access resources in proper order but still has not achieved high performance. It might be because he has not read an article diligently enough or practiced with online quizzes well enough. The material cause suggests that the SRL pattern is in good shape but needs strengthening. The next cause that suggests improvement is the efficient cause. It tells the learner where his or her SRL habits come from. If they are good habits, learners need to strengthen them by practice. If they are not, the learners need to omit these and replace them with new ones. For instance, learners might spend much time reading forum messages without posting opinions or sharing ideas. Such patterns of resource use might waste their time, reduce their strength of thought, and weaken their SRL ability. The improve SRL, the learners need to omit the habit of reading posts only and then replace it with reading-and-sharing ideas on discussion forums. The last cause that encourages improvement is the final cause. It is the most challenging yet most effective for perfecting SRL. The final cause tells how ones want one's SRL to become. Do we want SRL to be strong so that we have good SRL habits to study complicated knowledge? Do we want SRL to be perfect so that our minds reach perfection because SRL represents a combination of cognition and metacognition, the intellect and the will?

5.6. Conclusion

In summary, the SRL framework comprises the simple yet stable principles of SRL and a process for recognizing and improving SRL ability. Our proposed method was built on the SRL framework and designed for recognizing SRL in OLEs. The proposed method is not restricted to applying to a specific case; however, the evaluation and discussion above show that our proposed method is the potential for generalizing.

Chapter 6. Conclusion, limitations, and future works

6.1. Conclusion

Learning is a lifelong journey of seeking to know: knowledge, the world, and ourselves. Any of us has once experienced the joy of discovering and understanding certain pieces of knowledge. When we are aware of our learning with its capacity, characteristics and ability, is it not that we are pleased and confident to pursue our learning path? The further we learn, the more we need to understand our own learning habits and the more we need to master our SRL. And we can only understand our habits when we can explain them on sound principles. Those are the SRL framework as the principles and the learner Markov model as a means to understand SRL.

In this thesis, we have presented the topic of self-regulated learning in online learning environments and exhibited the idea of realizing SRL ability from the support provided by such environments. We would like to leave a note that the idea of learning and realizing one's learning ability is not constrained within online learning environments only. While traditional face-to-face classrooms enable instructors and supervisors to interact with learners to support them with SRL skills, online learning environments lack such in-person interaction but have a learning history to open changes for SRL support.

Through this thesis, we have tried to show that SRL is a manifestation of the intellect and the will, the cognition and metacognition; learning activities can be utilized to improve the way of learning. The relationship between cognition and metacognition is laid on the apparent yet stable foundation that is the operation of the mind. SRL has its causes; therefore, SRL has reasons for its existence, and we can understand our SRL capability and improve SRL ability through concrete activities. The SRL framework provides a foundation for a learner or a learner to describe SRL capability, explain SRL ability, and refer to adjusting and improving SRL.

We have also proposed the learner Markov model to turn learning history data into an SRL description and a measurement – SRL index – to unify learning activities, learning patterns, and their effectiveness. The learner Markov model and SRL index are the implementation of the SRL framework for online learning environments.

Our proposed method is not limited to a specific case like the OULAD in this thesis. The simplicity of the method and the straightforward procedure can be generalized and applicable to almost any OLEs.

6.2. Answers to the research questions

Let us return to the research questions. The SRL framework provides the following answers to research questions RQ1, RQ2, and RQ3. The SRL profile provides the following answers to research questions RQ4 and RQ5.

RQ1: *What intrinsic and extrinsic factors construct and differentiate the SRL ability of a learner?*

SRL is an entity that actually exists and operates on the basic functionalities of the mind; therefore, its ability can be explained by specifying the causes of SRL for each learner. In general, the factors constructing SRL ability are the source, form, goal, and pattern of SRL. Each learner has his or her educational background, learning experiences, and personality, therefore, has his or her path or source of SRL development. SRL operates to enable individuals to approach knowledge effectively and efficiently; hence, it possesses a form for achieving that aim. Since SRL does not end for itself but supports the learner to a goal in knowledge achievement, the goals to which SRL is directed also shape the SRL ability. Finally, individuals develop their cognitive and metacognitive strategies differently and shape their SRL habits and character on different paths; thus, the SRL pattern is then personalized to each individual. Thus, SRL converges in its form but varies according to individuals' backgrounds, learning goals, and cognitive and metacognitive habits.

RQ2: How can these SRL factors be identified and measured from a learner's learning history?

SRL ability reveals via a learner's SRL habits and character, which are currently evaluated by learning behavior observation and different types of self-reports. The frequency of behavior application should be the measurement unit for learning behaviors, and for measuring the quality of self-reports, such measurement scales as the Likert scale is reasonable.

RQ3: Under what cognitive or metacognitive conditions are individuals intrinsically/extrinsically motivated to self-regulate their learning?

Although motivation is one of SRL's critical components, a learner may find it unintriguing or unnecessary to develop the SRL ability since the benefits that SRL delivers are vaguely visible. However, by understanding SRL from the principles of the mind, a learner can be motivated to self-regulate his or her learning extrinsically by progress to knowledge and intrinsically by the perfection of the intellect and will, cognition and metacognition.

RQ4: By what signs can learners' learning history data in OLEs manifest SRL patterns?

Learners manifest their SRL patterns in the way they use learning resources in different timestamps in a course. Resource access before assessment milestones might demonstrate the planning phase in SRL. Resource access right after assessment days might signify the reflection phase in SRL. A combination of resource access throughout a course might indicate a complication of SRL activities of learners.

RQ5: How can learning history data from OLEs be synthesized for assessing the SRL ability of a learner?

A learner's SRL ability should be measured by the combination of SRL activities and their effectiveness on learning performance. And SRL index is a reasonable unit for measurement.

6.3. Contributions

The contributions of this research are two. First is the SRL recognition and improvement framework. Second is the learner Markov modeling method to transform learning history data into the SRL profile and SRL index. This combination of these components transforms online learning history data into a meaningful description of SRL for individual learners; thus, our proposed solution can support learners' understanding of their SRL ability in OLEs.

6.4. Limitations and future works

6.4.1. Limitations

The first limitation relates to the justification of the SRL framework. Since this framework is developed via arguments, future work must involve applying the framework to design empirical SRL recognition and improvement tools, programs, and exercises. Such empirical evidence will demonstrate the validity of the framework.

The second limitation involves the application of the SRL framework. Although saying that the SRL framework provides a foundation to describe existing SRL models, we have not used the SRL framework to describe the existing SRL models in detail. Such description is not for critiquing the SRL models but for showing the causes from which the SRL models are designed; thus, an appropriate understanding of SRL is provided. For instance, Zimmerman's cyclical phase model reflects the material cause of SRL, while Pintrich's model describes the characteristics of SRL, which is about the formal cause. Boekaerts' dual-processing model implies the efficient cause of SRL via the description of a learner's self-esteem. And almost all existing models do not clearly indicate the final cause of SRL. Different causes encourage different paths for improvement.

Although we have introduced the SRL recognition and improvement framework, the thesis does not mention concrete approaches, recommendations, or methods for improving SRL. Such contents are difficult to articulate in a general case. We, with a constraint of our research domain, think that our proposed method and framework contribute a way of recognizing SRL. And when one recognizes one's SRL ability, one can figure out how to modify it best for one's specific circumstances.

The third limitation relates to the verification of the learner Markov model. Although we have mentioned and subjectively believed the generalization of the proposed learner Markov model and SRL index, we have only applied the method on OULAD so far. It is important to apply the method to other open datasets and specific learning history data at educational institutions to validate the method.

Fourth is the limitation of the SRL profile. By introducing the SRL profile, we intend to give learners and learners a detailed illustration of their SRL pattern. By saying a detailed illustration of the SRL pattern, it means that learners can see their repetitive resource use, density and frequency of resource access, and effectiveness of resource use on learning performance. Looking at their SRL profile, learners and learners are expected to recognize their SRL habits and be able to make suitable modifications. The current SRL profile somehow shows those kinds of SRL pattern information; however, the current information about resource use patterns does not provide learners and learners with insight or guidelines to make an adjustment because the resources lack semantic contexts. In other words, all resources are currently just names or activity types with very little meaning implied in their names and without concrete purpose or intent implanted by instructors or course designers. When understanding the purpose of the resources provided, learners and learners will use them efficiently and reflect on their resource usage consciously, therefore, be able to make proper changes.

6.4.2. Future works

We have built and tested the learner Markov model from the OULAD dataset, whose learning resources do not provide much meaning except for resource types. Instructors and course designers always have an intention for every resource provided. We encourage to attach online learning resources with the intentions, such as the purpose of use, timeframe, and lecture content that instructors and course designers implant in the resources.

The learner Markov model represents the SRL capability and ability of a learner. To validate the model, learners should see the model to give justification. We encourage future work to model learners' learning history, then present the model to the learners for their feedback and evaluation of the correctness of the model.

SRL models are valuable outcomes from research in SRL. They bring SRL illustrative observable structures so that SRL can be analyzed and understood concretely from various aspects. We encourage further study to reshape the SRL profile from the viewpoint of different SRL models to assist learners and learners in seeing their SRL from various angles. With such concrete observations, students and learners can avoid the overwhelmingness of a vague SRL idea and have specific aspects of SRL to focus their modification and improvement.

Overall, we hope that the proposed SRL framework and the proposed method for modeling and measuring SRL contribute to the body of knowledge of SR, and the application of these propositions will be able to support SRL recognition and improvement; thus, lifelong learning becomes enjoyable and fruitful for each of us.

Publication List

Scholarly Journal

1. Tuan M. Tran, and Shinobu Hasegawa, 2022. "An Empirical Study on the Relationship between Cognition and Metacognition in Technology-Enhanced Self-Regulated Learning" Sustainability 14, no. 7: 3837. https://doi.org/10.3390/su14073837.

2. T. M. Tran, R. Beuran, S. Hasegawa, 2022. “Gamification-based cybersecurity awareness course for self-regulated learning.” International Journal of Information and Education Technology.

Conferences

1. T. M. Tran and S. Hasegawa, “A Hypothetical Model towards Establishing a Relationship between Cognition and Metacognition in Technology-Enhanced Self-Regulated Learning,” in Proceedings of the International Conference on Frontiers in Education: Computer Science & Computer Engineering, 2019, pp.155–158.

2. T. M. Tran., & S. Hasegawa (2021) Self-regulated Learning Recognition and Improvement Framework ISSN: 2186-5892 The Asian Conference on Education 2020: Official Conference Proceedings https://doi.org/10.22492/issn.2186-5892.2021.40.

3. X. Zheng, S. Hasegawa, M. T. Tran, K. Ota, T. Unoki (2021) Estimation of Learners’ Engagement Using Face and Body Features by Transfer Learning. In: Degen H., Ntoa S. (eds) Artificial Intelligence in HCI. HCII 2021. Lecture Notes in Computer Science, vol 12797. Springer, Cham. https://doi.org/10.1007/978-3-030-77772-2_36.

4. T. M. Tran, R. Beuran, S. Hasegawa, “Gamification-based cybersecurity awareness course for self-regulated learning” in 6th International Conference on Education and Distance Learning, Rome, Italy, July 21-23, 2022.

5. Tuan M. Tran and Shinobu Hasegawa, 2022. “Using Markov chain on online learning history data to develop learner model for measuring strength of learning habits” in 19th

International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2022), Lisbon, Portugal, November 8-10, 2022.

Bibliography

[1] N. van Halem, C. van Klaveren, H. Drachsler, M. Schmitz, and I. Cornelisz, "Tracking Patterns in Self-Regulated Learning Using Students' Self-Reports and Online Trace Data," *Frontline Learning Research*, vol. 8, no. 3, pp. 140–163, 2020, doi: 10.14786/flr.v8i3.497.

[2] L. Barnard, W. Y. Lan, Y. M. To, V. O. Paton, and S.-L. Lai, "Measuring self-regulation in online and blended learning environments," *The Internet and Higher Education*, vol. 12, no. 1, pp. 1–6, Jan. 2009, doi: 10.1016/j.iheduc.2008.10.005.

[3] T. J. Cleary, G. L. Callan, and B. J. Zimmerman, "Assessing Self-Regulation as a Cyclical, Context-Specific Phenomenon: Overview and Analysis of SRL Microanalytic Protocols," *Education Research International*, vol. 2012, pp. 1–19, 2012, doi: 10.1155/2012/428639.

[4] M. Tran and S. Hasegawa, "Self-Regulated Learning Recognition and Improvement Framework," 2021. doi: 10.22492/issn.2186-5892.2021.40.

[5] M. L. Bernacki, A. C. Aguilar, and J. P. Byrnes, "Self-Regulated Learning and Technology-Enhanced Learning Environments," in *Fostering Self-Regulated Learning through ICT*, G. Dettori and D. Persico, Eds. IGI Global, 2011. doi: 10.4018/978-1-61692-901-5.ch001.

[6] M. Yamada, Y. Goda, T. Matsuda, Y. Saito, H. Kato, and H. Miyagawa, "Self-Regulator: Preliminary Research of the Effects of Supporting Time Management on Learning Behaviors," in *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, Jul. 2017, pp. 370–372. doi: 10.1109/ICALT.2017.85.

[7] T. M. Tran and S. Hasegawa, "An Empirical Study on the Relationship between Cognition and Metacognition in Technology-Enhanced Self-Regulated Learning," *Sustainability*, vol. 14, no. 7, p. 3837, Mar. 2022, doi: 10.3390/su14073837.

[8] Q. Li, R. Baker, and M. Warschauer, “Using clickstream data to measure, understand, and support self-regulated learning in online courses,” *The Internet and Higher Education*, vol. 45, Apr. 2020, doi: 10.1016/j.iheduc.2020.100727.

[9] O. Viberg, M. Hatakka, O. Bälter, and A. Mavroudi, “The current landscape of learning analytics in higher education,” *Computers in Human Behavior*, vol. 89, Dec. 2018, doi: 10.1016/j.chb.2018.07.027.

[10] R. Baker *et al.*, “The benefits and caveats of using clickstream data to understand student self-regulatory behaviors: opening the black box of learning processes,” *International Journal of Educational Technology in Higher Education*, vol. 17, no. 1, p. 13, Dec. 2020, doi: 10.1186/s41239-020-00187-1.

[11] F. Hlioui, N. Aloui, and F. Gargouri, “A Withdrawal Prediction Model of At-Risk Learners Based on Behavioural Indicators,” *International Journal of Web-Based Learning and Teaching Technologies*, vol. 16, no. 2, pp. 32–53, Mar. 2021, doi: 10.4018/IJWLTT.2021030103.

[12] J. Kuzilek, M. Hlosta, and Z. Zdrahal, “Open University Learning Analytics dataset,” *Scientific Data*, vol. 4, no. 1, Dec. 2017, doi: 10.1038/sdata.2017.171.

[13] M.-T. Tran and S. Hasegawa, “Self-regulated Learning Recognition and Improvement Framework,” Jan. 2021, pp. 449–465. doi: 10.22492/issn.2186-5892.2021.40.

[14] J. Smith, “Self-consciousness,” in *The Stanford Encyclopedia of Philosophy*, Summer 2020., E. N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2020. [Online]. Available: https://plato.stanford.edu/archives/sum2020/entries/self-consciousness/

[15] M. Boekaerts, “Self-regulated Learning at the Junction of Cognition and Motivation,” *European Psychologist*, vol. 1, no. 2, pp. 100–112, 1996, doi: 10.1027/1016-9040.1.2.100.

[16] A. Efklides, “Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model,” *Educational Psychologist*, vol. 46, no. 1, pp. 6–25, 2011, doi: 10.1080/00461520.2011.538645.

[17] P. H. Winne and N. E. Perry, “Measuring Self-Regulated Learning,” in *Handbook of Self-Regulation*, Elsevier, 2000, pp. 531–566. doi: 10.1016/B978-012109890-2/50045-7.

[18] B. J. Zimmerman, *Self-Regulated Learning : Theories , Measures , and Outcomes*, Second Edi., vol. 21. Elsevier, 2015. doi: 10.1016/B978-0-08-097086-8.26060-1.

[19] E. Panadero, “A Review of Self-regulated Learning: Six Models and Four Directions for Research,” *Frontiers in Psychology*, vol. 8, p. 422, 2017, doi: 10.3389/fpsyg.2017.00422.

[20] P. H. Winne, “A metacognitive view of individual differences in self-regulated learning,” *Learning and Individual Differences*, vol. 8, no. 4, pp. 327–353, 1996, doi: 10.1016/S1041-6080(96)90022-9.

[21] M. Boekaerts and M. Niemivirta, “Self-Regulated Learning: Finding a Balance between Learning Goals and Ego-Protective,” in *Handbook of Self-Regulation*, M. Boekaerts, P. R. Pintrich, and M. Zeidner, Eds. Academic Press, 2000, pp. 417–450. [Online]. Available: https://doi.org/10.1016/B978-012109890-2/50042-1

[22] M. Boekaerts and E. Cascallar, “How far have we moved toward the integration of theory and practice in self-regulation?,” *Educational Psychology Review*, vol. 18, no. 3, pp. 199–210, 2006, doi: 10.1007/s10648-006-9013-4.

[23] P. R. Pintrich, “The role of goal orientation in self-regulated learning,” in *Handbook of Self-Regulation*, Academic Press, 2000, pp. 451–502. doi: https://doi.org/10.1016/B978-012109890-2/50043-3.

[24] A. Bandura, “Social Cognitive Theory,” in *Six theories of Children Development*, vol. 6, Greenwich: JAI Press, 1989, pp. 1–60. doi: 10.4324/9780429052675-22.

[25] B. J. Zimmerman, “Chapter 2: Attaining Self-Regulation:A social cognitive perspective,” in *Handbook of Self-Regulation*, M. Boekaerts, P. R. Pintrich, and M. Zeidner, Eds. Academic Press, 2000, pp. 13–39.

[26] B. J. Zimmerman, “Becoming a Self-Regulated Learner: An Overview,” *Theory Into Practice*, vol. 41, no. 2, pp. 64–70, May 2002, doi: 10.1207/s15430421tip4102_2.

[27] A. J. Rockinson-Szapkiw, “The impact of asynchronous and synchronous instruction and discussion on cognitive presence, social presence, teaching presence, and learning,” Ann Arbor, 2009. [Online]. Available: https://www.proquest.com/dissertations-theses/impact-asynchronous-synchronous-instruction/docview/250918726/se-2

[28] M. Oztok, D. Zingaro, C. Brett, and J. Hewitt, “Exploring asynchronous and synchronous tool use in online courses,” *Computers & Education*, vol. 60, no. 1, pp. 87–94, Jan. 2013, doi: 10.1016/j.compedu.2012.08.007.

[29] D. J. Lemay, C. Baek, and T. Doleck, “Comparison of learning analytics and educational data mining: A topic modeling approach,” *Computers and Education: Artificial Intelligence*, vol. 2, p. 100016, 2021, doi: 10.1016/j.caeai.2021.100016.

[30] P. Long, “Message from the LAK 2011 General & Program Chairs ,” 2011.

[31] C. Romero and S. Ventura, “Educational data mining: A survey from 1995 to 2005,” *Expert Systems with Applications*, vol. 33, no. 1, pp. 135–146, 2007, doi: 10.1016/j.eswa.2006.04.005.

[32] C. Romero and S. Ventura, “Educational data mining and learning analytics: An updated survey,” *WIREs Data Mining and Knowledge Discovery*, vol. 10, no. 3, May 2020, doi: 10.1002/widm.1355.

[33] J. Lee and M. Podlaseck, “Visualization and Analysis of Clickstream Data of Online Stores with a Parallel Coordinate System,” in *Electronic Commerce and Web Technologies*, 2000, pp. 145–154. doi: 10.1007/3-540-44463-7_13.

[34] W. W. Moe, “Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream,” *Journal of Consumer Psychology*, vol. 13, no. 1–2, pp. 29–39, 2003, doi: 10.1207/s15327663jcp13-1&2_03.

[35] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, “Characterizing user behavior in online social networks,” in *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference - IMC ’09*, 2009, pp. 49–62. doi: https://doi.org/10.1145/1644893.1644900.

[36] L. Haiyang, Z. Wang, P. Benachour, and P. Tubman, "A Time Series Classification Method for Behaviour-Based Dropout Prediction," in *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, Jul. 2018, pp. 191–195. doi: 10.1109/ICALT.2018.00052.

[37] I. Nitta *et al.*, "Graph-based Massive Open Online Course (MOOC) Dropout Prediction using Clickstream Data in Virtual Learning Environment," in *2021 16th International Conference on Computer Science & Education (ICCSE)*, Aug. 2021, pp. 48–52. doi: 10.1109/ICCSE51940.2021.9569582.

[38] H. Waheed, S.-U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, "Predicting academic performance of students from VLE big data using deep learning models," *Computers in Human Behavior*, vol. 104, Mar. 2020, doi: 10.1016/j.chb.2019.106189.

[39] F. Qiu *et al.*, "Predicting students' performance in e-learning using learning process and behaviour data," *Scientific Reports*, vol. 12, no. 1, p. 453, Jan. 2022, doi: 10.1038/s41598-021-03867-8.

[40] N. Jha, I. Ghergulescu, and A.-N. Moldovan, "OULAD MOOC Dropout and Result Prediction using Ensemble, Deep Learning and Regression Techniques," in *Proceedings of the 11th International Conference on Computer Supported Education*, 2019, pp. 154–164. doi: 10.5220/0007767901540164.

[41] Y. He *et al.*, "Online At-Risk Student Identification using RNN-GRU Joint Neural Networks," *Information*, vol. 11, no. 10, p. 474, Oct. 2020, doi: 10.3390/info11100474.

[42] R. Alshabandar, A. Hussain, R. Keight, and W. Khan, "Students Performance Prediction in Online Courses Using Machine Learning Algorithms," in *2020 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2020, pp. 1–7. doi: 10.1109/IJCNN48605.2020.9207196.

[43] J. Hao, J. Gan, and L. Zhu, "MOOC performance prediction and personal performance improvement via Bayesian network," *Education and Information Technologies*, Feb. 2022, doi: 10.1007/s10639-022-10926-8.

[44] E. Drousiotis, L. Shi, and S. Maskell, "Early Predictor for Student Success Based on Behavioural and Demographical Indicators," 2021, pp. 161–172. doi: 10.1007/978-3-030-80421-3_19.

[45] J. Park, K. Denaro, F. Rodriguez, P. Smyth, and M. Warschauer, "Detecting changes in student behavior from clickstream data," in *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, Mar. 2017, pp. 21–30. doi: 10.1145/3027385.3027430.

[46] R. Yu, D. Jiang, and M. Warschauer, "Representing and predicting student navigational pathways in online college courses," in *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, Jun. 2018, pp. 1–4. doi: 10.1145/3231644.3231702.

[47] Q. Li, R. Baker, and M. Warschauer, "Using clickstream data to measure, understand, and support self-regulated learning in online courses," *The Internet and Higher Education*, vol. 45, Apr. 2020, doi: 10.1016/j.iheduc.2020.100727.

[48] R. F. Kizilcec, M. Pérez-Sanagustín, and J. J. Maldonado, "Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses," *Computers & Education*, vol. 104, pp. 18–33, Jan. 2017, doi: 10.1016/j.compedu.2016.10.001.

[49] A. Cicchinelli *et al.*, "Finding traces of self-regulated learning in activity streams," in *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, Mar. 2018, pp. 191–200. doi: 10.1145/3170358.3170381.

[50] J. Wong, M. Khalil, M. Baars, B. B. de Koning, and F. Paas, "Exploring sequences of learner activities in relation to self-regulated learning in a massive open online course," *Computers & Education*, vol. 140, Oct. 2019, doi: 10.1016/j.compedu.2019.103595.

[51] C. Geigle and C. Zhai, "Modeling MOOC Student Behavior With Two-Layer Hidden Markov Models," in *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*, Apr. 2017, pp. 205–208. doi: 10.1145/3051457.3053986.

[52] L. Qiao, W. Zhao, and X. Xu, “Mining and Analysis of Self-regulated Learning Process Model: Based on Hidden Markov Model,” in *2021 Tenth International Conference of Educational Innovation through Technology (EITT)*, Dec. 2021, pp. 276–281. doi: 10.1109/EITT53287.2021.00061.

[53] G. K. W. K. Chung, L. F. de Vries, A. M. Cheak, R. H. Stevens, and W. L. Bewley, “Cognitive process validation of an online problem solving assessment,” *Computers in Human Behavior*, vol. 18, no. 6, pp. 669–684, Nov. 2002, doi: 10.1016/S0747-5632(02)00023-7.

[54] P. H. Winne and Z. Marzouk, “Learning Strategies and Self-Regulated Learning,” in *The Cambridge Handbook of Cognition and Education*, Cambridge University Press, 2019, pp. 696–715. doi: 10.1017/9781108235631.028.

[55] W. J. McKeachie, “THE NEED FOR STUDY STRATEGY TRAINING,” in *Learning and Study Strategies*, Elsevier, 1988, pp. 3–9. doi: 10.1016/B978-0-12-742460-6.50007-4.

[56] R. McInerny and J. O’Callaghan, “Saint Thomas Aquinas,” in *The Stanford Encyclopedia of Philosophy*, Summer 2018., E. N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2018. [Online]. Available: https://plato.stanford.edu/archives/sum2018/entries/aquinas/

[57] B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, “Taxomony of Educational Objective,” *Taxonomy of educational objectives*, pp. 62–197, 1956, [Online]. Available: https://www.uky.edu/~rsand1/china2018/texts/Bloom et al -Taxonomy of Educational Objectives.pdf

[58] M. K. Tillman, “The Philosophic Habit of Mind: Aristotle and Newman on the End of Liberal Education,” *Philosophical Inquiry in Education*, vol. 3, no. 2, pp. 17–27, 1990.

[59] C. Shields, “Aristotle,” in *The Stanford Encyclopedia of Philosophy*, Fall 2020., Edward N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2020. [Online]. Available: https://plato.stanford.edu/archives/fall2020/entries/aristotle/

[60] S. F. E. Rovers, G. Clarebout, H. H. C. M. Savelberg, A. B. H. de Bruin, and J. J. G. van Merriënboer, “Granularity matters: comparing different ways of measuring self-regulated learning,” *Metacognition and Learning*, vol. 14, no. 1, pp. 1–19, 2019, doi: 10.1007/s11409-019-09188-6.

[61] E. Araka, E. Maina, R. Gitonga, and R. Oboko, “Research trends in measurement and intervention tools for self-regulated learning for e-learning environments—systematic review (2008–2018),” *Research and Practice in Technology Enhanced Learning*, vol. 15, no. 1, 2020, doi: 10.1186/s41039-020-00129-5.

[62] M.-T. Tran and S. Hasegawa, “A Hypothetical Model towards Establishing a Relationship between Cognition and Metacognition in Technology-Enhanced Self-Regulated Learning,” in *Proceedings of the 2019 International Conference on Frontiers in Education: Computer Science & Computer Engineering*, 2019, pp. 155–158.

[63] Moodle, “Solutions for Higher education.” https://moodle.com/solutions/higher-education/ (accessed Jul. 30, 2021).

[64] D. R. Krathwohl, “A Revision of Bloom’s Taxonomy: An Overview,” *Theory Into Practice*, vol. 41, no. 4, pp. 212–218, Nov. 2002, doi: 10.1207/s15430421tip4104_2.

[65] G. Cohen, S. Afshar, J. Tapson, and A. van Schaik, “EMNIST: an extension of MNIST to handwritten letters,” 2017, Accessed: Aug. 13, 2021. [Online]. Available: https://arxiv.org/abs/1702.05373v2

[66] L. Barnard-Brak, W. Y. Lan, and V. O. Paton, “Measuring and Profiling Self-Regulated Learning in the Online Environment,” in *Fostering Self-Regulated Learning through ICT*, IGI Global, 2011. doi: 10.4018/978-1-61692-901-5.ch002.

[67] D. P. Bertsekas and J. N. Tsitsiklis, “Markov chains,” in *Introduction to Probability*, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2000, pp. 2–3.

[68] L. E. Sucar, “Chapter 5. Hidden Markov Models,” in *Probabilistic Graphical Models: Principles and Applications*, Cham: Springer International Publishing, 2020, pp. 74–76. doi: 10.1007/978-3-030-61943-5.

[69] L. Page, “The PageRank Citation Ranking: Bringing Order to the Web.” 1998. Accessed: Mar. 01, 2022. [Online]. Available: http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf

[70] D. P. Bertsekas and J. N. Tsitsiklis, “6. Markov chains,” in *Introduction to Probability*, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2000, pp. 13–13.

[71] Gilbert Strang, “Chapter 6. Eigenvalues and Eigenvectors,” in *Introduction to Linear Algebra*, 5th ed., Wellesley-Cambridge Press, 2016, pp. 288–303.

[72] G. A. Spedicato, “Discrete Time Markov Chains with R,” *The R Journal*, Mar. 2017, [Online]. Available: https://journal.r-project.org/archive/2017/RJ-2017-036/index.html

[73] scikit learn, “sklearn.svm.svc.” https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html (accessed Apr. 29, 2022).

[74] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[75] Scikit learn, “sklearn.neighbors.KNeighborsClassifier.” https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html (accessed Apr. 29, 2022).

[76] Scikit learn, “sklearn.naive_bayes.GaussianNB.” https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html (accessed Apr. 29, 2022).

Appendix 1. Structures of OULAD file

StudentInfo.csv contains 32,593 rows with the columns as shown in the following snapshot.

code_module	code_presentation	id_student	gender	region	highest_education	imd_band	age_band	num_of_prev_	studied_credits	disability	final_result
AAA	2013J	11391	M	East Anglian Region	HE Qualification	90-100%	55<=	0	240	N	Pass
AAA	2013J	28400	F	Scotland	HE Qualification	20-30%	35-55	0	60	N	Pass
AAA	2013J	30268	F	North Western Region	A Level or Equivalent	30-40%	35-55	0	60	Y	Withdrawn
AAA	2013J	31604	F	South East Region	A Level or Equivalent	50-60%	35-55	0	60	N	Pass
AAA	2013J	32885	F	West Midlands Region	Lower Than A Level	50-60%	0-35	0	60	N	Pass
AAA	2013J	38053	M	Wales	A Level or Equivalent	80-90%	35-55	0	60	N	Pass
AAA	2013J	45462	M	Scotland	HE Qualification	30-40%	0-35	0	60	N	Pass
AAA	2013J	45642	F	North Western Region	A Level or Equivalent	90-100%	0-35	0	120	N	Pass
AAA	2013J	52130	F	East Anglian Region	A Level or Equivalent	70-80%	0-35	0	90	N	Pass
AAA	2013J	53025	M	North Region	Post Graduate Qualification		55<=	0	60	N	Pass
AAA	2013J	57506	M	South Region	Lower Than A Level	70-80%	35-55	0	60	N	Pass
AAA	2013J	58873	F	East Anglian Region	A Level or Equivalent	20-30%	0-35	0	60	N	Pass
AAA	2013J	59185	M	East Anglian Region	Lower Than A Level	60-70%	35-55	0	60	N	Pass
AAA	2013J	62155	F	North Western Region	HE Qualification	50-60%	0-35	0	60	N	Pass
AAA	2013J	63400	M	Scotland	Lower Than A Level	40-50%	35-55	0	60	N	Pass
AAA	2013J	65002	F	East Anglian Region	A Level or Equivalent	70-80%	0-35	0	60	N	Withdrawn
AAA	2013J	70464	F	West Midlands Region	A Level or Equivalent	60-70%	35-55	0	60	N	Pass
AAA	2013J	71361	M	Ireland	HE Qualification		35-55	0	60	N	Pass
AAA	2013J	74372	M	East Anglian Region	A Level or Equivalent	10-20	35-55	0	150	N	Fail
AAA	2013J	75091	M	South West Region	A Level or Equivalent	30-40%	35-55	0	60	N	Pass

Courses.csv contains 22 rows with the columns as shown in the following snapshot.

code_module	code_presentation	module_presentation_length
AAA	2013J	268
AAA	2014J	269
BBB	2013J	268
BBB	2014J	262
BBB	2013B	240
BBB	2014B	234
CCC	2014J	269
CCC	2014B	241
DDD	2013J	261
DDD	2014J	262
DDD	2013B	240
DDD	2014B	241
EEE	2013J	268
EEE	2014J	269
EEE	2014B	241
FFF	2013J	268
FFF	2014J	269
FFF	2013B	240
FFF	2014B	241
GGG	2013J	261
GGG	2014J	269
GGG	2014B	241

studentRegistration.csv contains 32,593 rows with the columns as shown in the following snapshot

code_module	code_presentation	id_student	date_registration	date_unregistration
AAA	2013J	11391	-159	
AAA	2013J	28400	-53	
AAA	2013J	30268	-92	12
AAA	2013J	31604	-52	
AAA	2013J	32885	-176	
AAA	2013J	38053	-110	
AAA	2013J	45462	-67	
AAA	2013J	45642	-29	
AAA	2013J	52130	-33	
AAA	2013J	53025	-179	
AAA	2013J	57506	-103	
AAA	2013J	58873	-47	
AAA	2013J	59185	-59	
AAA	2013J	62155	-68	
AAA	2013J	63400	-67	
AAA	2013J	65002	-180	96
AAA	2013J	70464	-95	

Assessments.csv contains 206 rows with the columns as shown in the following snapshot

code_module	code_presentation	id_assessment	assessment_type	date	weight
AAA	2013J	1752	TMA	19	10
AAA	2013J	1753	TMA	54	20
AAA	2013J	1754	TMA	117	20
AAA	2013J	1755	TMA	166	20
AAA	2013J	1756	TMA	215	30
AAA	2013J	1757	Exam		100
AAA	2014J	1758	TMA	19	10
AAA	2014J	1759	TMA	54	20
AAA	2014J	1760	TMA	117	20
AAA	2014J	1761	TMA	166	20
AAA	2014J	1762	TMA	215	30
AAA	2014J	1763	Exam		100
BBB	2013B	14991	CMA	54	1
BBB	2013B	14992	CMA	89	1
BBB	2013B	14993	CMA	124	1
BBB	2013B	14994	CMA	159	1

studentAssessment.csv contains 173,912 rows with the columns as shown in the following snapshot

id_assessment	id_student	date_submitted	is_banked	score
1752	11391	18	0	78
1752	28400	22	0	70
1752	31604	17	0	72
1752	32885	26	0	69
1752	38053	19	0	79
1752	45462	20	0	70
1752	45642	18	0	72
1752	52130	19	0	72
1752	53025	9	0	71
1752	57506	18	0	68
1752	58873	19	0	73

studentVle.csv contains 10,655,280 rows with the columns as shown in the following snapshot

code_module	code_presentation	id_student	id_site	date	sum_click
AAA	2013J	28400	546652	-10	4
AAA	2013J	28400	546652	-10	1
AAA	2013J	28400	546652	-10	1
AAA	2013J	28400	546614	-10	11
AAA	2013J	28400	546714	-10	1
AAA	2013J	28400	546652	-10	8
AAA	2013J	28400	546876	-10	2
AAA	2013J	28400	546688	-10	15
AAA	2013J	28400	546662	-10	17
AAA	2013J	28400	546890	-10	1
AAA	2013J	28400	547011	-10	1
AAA	2013J	28400	547013	-10	1
AAA	2013J	28400	546871	-10	3

vle.csv contains 6,364 rows with the columns as shown in the following snapshot.

id_site	code_module	code_presentation	activity_type	week_from	week_to
773452	DDD	2014B	url	0	0
813933	DDD	2014J	oucontent	0	0
546719	AAA	2013J	oucontent	1	1
546681	AAA	2013J	oucontent	1	1
877045	AAA	2014J	oucontent	1	1
877044	AAA	2014J	oucontent	1	1
704071	BBB	2013J	resource	1	1
703900	BBB	2013J	oucontent	1	1
704217	BBB	2013J	resource	1	1
703943	BBB	2013J	subpage	1	1
704215	BBB	2013J	resource	1	1
768475	BBB	2014B	resource	1	1
768472	BBB	2014B	resource	1	1

Appendix 2. Source code in R for Dataset generation

This function runs the 6 steps to generate a dataset containing the SRL index, the log SRL index, and the cognitive score of each student in a course offering from OULAD.

The parameter `studentInfoData` gets data from the studentInfo.csv file

The parameter `studentVleData` gets data from the studentVle.csv file

The parameter `vleData` gets data from the vle.csv file

The parameter `assessmentData` gets data from the assessments.csv

The parameter `studentAssessmentData` gets data from the studentAssessment.csv

The parameter `codeModule` is the values from the code_module column

The parameter `codePresentation` is the values from the code_presentation column

The parameter `phase` gets the following values: ‘planning’ to represent the SRL plannning phase, ‘reflection’ to represent the SRL reflection phase, and ‘all’ to represent SRL activities in general throughout a course.

The parameter `days` is used when the parameter phase is assigned as ‘plannning’ or ‘reflection’. When `phase` is assigned as ‘planning’, `days` is the number of days before the assessment days in a course. When `phase` is assigned as ‘reflection’, `days` is the number of days after the assessment days in a course.

```
RunSRLScorePipeline <- function(studentInfoData, studentVleData, vleData,
assessmentData, studentAssessmentData, codeModule, codePresentation, phase,
days) {
    print("Step 1. Get VLE use history")
    vleDf <- CreateVleAccessByStudentinCourse(studentVleData =
studentVleData, vleData = vleData, assessmentData = assessmentData,
codeModule = codeModule, codePresentation = codePresentation)
    print("----- Step 1 completed.")


    print("Step 2. Synthesize scores")
```

```
    scoreDf <-GetStudentsTotalScoreInCourse(studentInfoData =
studentInfoData, studentAssessmentData = studentAssessmentData,
assessmentData = assessmentData, codeModule = codeModule, codePresentation =
codePresentation)
    print("----- Step 2 completed.")


    print("Step 3. Compute vle transition matrices")
    transitionMatrices <- BuildTransitionMatrices(studentInfoData =
studentInfoData, studentVleData = studentVleData, vleData = vleData,
assessmentData = assessmentData, codeModule = codeModule, codePresentation =
codePresentation, phase = phase, days = days)
    print("----- Step 3 completed.")


    print("Step 4. Build Markov chains")
    markovChainModels <- BuildMarkovChainModels(transitionMatrices)
    print("----- Step 4 completed.")


    print("Step 5. Generate a steady states dataset")
    steadyStatesDf<-GenerateSteadyStatesDataset(studentInfoData =
studentInfoData, codeModule=codeModule, codePresentation=codePresentation,
markovChainModels=markovChainModels)
    print("----- Step 5 completed.")


    print("Step 6. Compute SRL indices")
    srlScoreDf<-ComputeSRLIndex(steadyStatesData=steadyStatesDf,
vleUseData=vleDf, totalScoreData=scoreDf)
    print("----- Step 6 completed.")
    srlScoreDf
  }
```

Function `CreateVleAccessByStudentinCourse` for step 1 of the procedure

```
CreateVleAccessByStudentinCourse <- function(studentVleData, vleData, assessmentData, codeModule, codePresentation, forExamDays, days) {

    # step 1. getVleUseInCourse
    vleUseDf <- GetVleUseInCourse(studentVleData = studentVleData, vleData = vleData, codeModule = codeModule, codePresentation = codePresentation)

    if(forExamDays != "") {
      assessmentDates <- GetAssessmentDate(assessmentData = assessmentData, codeModule = codeModule, codePresentation = codePresentation)
      df <- vleUseDf[0,]
      if(forExamDays == "before") {
        for(i in assessmentDates) {
          temp <- vleUseDf %>% filter((date<i) & (date >= i - days))
          df <- rbind(df, temp)
        }
      }
      else if (forExamDays == "after") {
        for(i in assessmentDates) {
          temp <- vleUseDf %>% filter((date>i) & (date <= i + days))
          df <- rbind(df, temp)
        }
      }
      vleUseDf <- df
    }

    # step 2. Count vle use by type for each student
    vleUseByTypePerStudent <- vleUseDf %>%
                                 group_by(id_student, activity_type) %>%
                                 summarize(num_use = n())

    # Step 3. Add missing vle type with num_use for each student
```

```
    vleUseByTypePerStudent <- vleUseByTypePerStudent %>%
                                ungroup() %>%
                                  complete(id_student, activity_type,
                                  fill = list(num_use = 0))


    # Step 4. Do pivot_wider the dataset into dataframe with activity_types as columns and students as row
    vleUseByTypePerStudent <- vleUseByTypePerStudent %>%
                                pivot_wider(names_from = activity_type,
                                            values_from = num_use)
    vleUseByTypePerStudent
  }
```

Sub functions `GetVleUseInCourse, GetAssessmentDate` in step 1 of the procedure

```
  GetVleUseInCourse <- function(studentVleData, vleData,
                                  codeModule, codePresentation) {


    temp1 <- studentVleData
    temp2 <- vleData


    if (codeModule != "ALL") { # step 1. filter data to rows of code_module and code_presentation
      temp1 <- temp1 %>% filter(code_module == codeModule)
      temp2 <- vleData %>% filter(code_module == codeModule)
    }


    if (codePresentation != "ALL") { # step 1. filter data to rows of code_module and code_presentation
      temp1 <- filter(temp1, code_presentation == codePresentation)
      temp2 <- filter(temp2, code_presentation == codePresentation)
    }


    # step 2. inner join vle and student vle on id_site to get activity_type name
```

```
    df <- inner_join(temp1, temp2, by = "id_site")


    # step 3. remove redundant columns and rename code_module and code_presentation columns
    df <- df %>%
      select(-code_module.y, -code_presentation.y) %>%
      rename(code_module = code_module.x, code_presentation = code_presentation.x)
  }
```

```
  GetAssessmentDate <- function(assessmentData, codeModule, codePresentation) {

    df <- assessmentData %>% filter(code_module == codeModule,
                                     code_presentation == codePresentation)
    dates <- df %>%
              select(date) %>%
              arrange(date) %>%
              distinct()
    dates <- unlist(dates, use.names = FALSE)
    dates <- dates[!is.na(dates)]
    dates
  }
```

Function `GetStudentsTotalScoreInCourse` for step 2 of the procedure

```
  GetStudentsTotalScoreInCourse <- function(studentInfoData, studentAssessmentData, assessmentData, codeModule, codePresentation) {

    assessDf <- assessmentData %>%
                 filter(code_module == codeModule,
                        code_presentation == codePresentation)
```

```
    studentDf <- studentInfoData %>%
                filter(code_module == codeModule,
                       code_presentation == codePresentation) %>%
                       select(id_student, final_result)


    # step 1. Join assessment with studentAssessment to compute total
scores earned by each student
    df <- studentAssessmentData %>%
                inner_join(assessDf, by = "id_assessment")


    # step 2. Compute partial score by multiply the score by its weight
and fill NA score with 0
    df$partial_score <- df$score * df$weight / 100
    df$partial_score[is.na(df$partial_score)] <- 0


    # step 3. Apply pivot_wider on df with columns as partial score for
each test, fill NA score wit 0
    widerpivotScore <- df %>%
         pivot_wider(names_from = id_assessment,
                     values_from = partial_score, id_cols = id_student)
    widerpivotScore[is.na(widerpivotScore)] <- 0


    # Step 4. Compute total score of each student
    widerpivotScore$total_score <-

  rowSums(widerpivotScore[,2:ncol(widerpivotScore)])
    widerpivotScore


    # Step 5. Join score and final results
    df <- widerpivotScore %>%
                       right_join(studentDf, by = "id_student")
    df
  }
```

Function BuildTransitionMatrices for step 3 of the procedure

```
  BuildTransitionMatrices <- function(studentInfoData, studentVleData,vleData,
assessmentData, codeModule, codePresentation, phase, days) {
    # step 1. filter data to rows having codemodule and codepresentation
    df <- studentInfoData %>%
            filter(code_module==codeModule,
code_presentation==codePresentation)
    vleUseDf <- GetVleUseInCourse(studentVleData = studentVleData,
                                  vleData = vleData,
                                  codeModule = codeModule,
                                  codePresentation = codePresentation)
    assessDates <- GetAssessmentDate(assessmentData = assessmentData,
                                     codeModule = codeModule,
                                  codePresentation = codePresentation)
    # step 2. iterate the filtered dataset, run compute_transition_matrix on each
student id and assign a new transition matrix of each run into a shared list
    list <- list()
    print(paste("Number of students to whom transition matrices will be built:",
nrow(df), sep=""))
    # make sure nrow(df > 0)
    for(i in 1:nrow(df)) {
        sdata <- vleUseDf %>% filter(id_student == df$id_student[i])
        transition_matrix <- ""
        if(nrow(sdata) > 0) {
          if (phase == "planning") {
            transition_matrix <-
              compute_plan_transition_matrix(student_vle_data = sdata,
                                             assessDates, days)
          }
          else if (phase == "reflection") {
            transition_matrix <-
                 compute_reflect_transition_matrix(sdata,
                                                   assessDates, days)
          }
          else {
```

```
        transition_matrix <- compute_transition_matrix(sdata)
      }
      comment(transition_matrix) <- paste(df$id_student[i],
                                          df$final_result[i], sep="-" )
      list[[length(list)+1]] <- transition_matrix
      print(paste(i, "matrix made", sep=" - "))
    }
    else {
      comment(transition_matrix) <- paste(df$id_student[i],
                                          df$final_result[i], sep="-" )
      list[[length(list)+1]] <- transition_matrix
      print(paste(i, "Failed to build a matrix", sep=" - "))
    }
  }
  list
}
```

Subfunctions `compute_transition_matrix`, `compute_plan_transition_matrix`, `compute_reflect_transition_matrix` for step 3 of the procedure

```
compute_transition_matrix <- function(student_vle_data) {

  # Count the number of use of each activity_type
  act <- student_vle_data %>% group_by(activity_type) %>%
                                        summarize(num_use = n())

  # sort data in ascending order of date
  df <- student_vle_data %>% arrange(date)

  # Build a matrix for storing transition probability
  matrix <- matrix(data = 0, nrow = length(act$activity_type),
                             ncol = length(act$activity_type),
                   dimnames = list(act$activity_type, act$activity_type))

  # count the time an act in one date is followed by another act in the next date
```

```
    if(nrow(df) > 0) {
      for (i in 1:nrow(df)) {
        # Consider activities in a day take turn to follow one another, extract
other rows of the same day as i
        same_date <- df %>% filter(date == df$date[i])
        if(nrow(same_date) > 0) {
          for (j in 1:nrow(same_date)) {
            matrix[df$activity_type[i], same_date$activity_type[j]] <-

            matrix[df$activity_type[i], same_date$activity_type[j]] + 1
          }
        }
        # extract rows whose date is the next of i
        next_date <- df %>% filter(date == df$date[i] + 1)
        # add 1 when there is a transition from activity in date i to activity in
date j
        if(nrow(next_date) > 0) {
          for (j in 1:nrow(next_date)) {
            matrix[df$activity_type[i], next_date$activity_type[j]] <-

                    matrix[df$activity_type[i], next_date$activity_type[j]] + 1
          }
        }
      }
    }


    # divide each row of the matrix by the number of use of respective activity_type
    if(nrow(act)>0) {
      for(i in 1:nrow(act)) {
        # count the total number of sequences started with act$activity_type[i]
        total <- sum(matrix[act$activity_type[i], ])
        for( j in 1:ncol(matrix)) {
          matrix[act$activity_type[i],j] <- matrix[act$activity_type[i],j] / total
        }
        # if a row of the matrix is NA, assign matrix[row, row] = 1
        if(is.na(matrix[act$activity_type[i],1])) {
          matrix[act$activity_type[i],] <- 0
          matrix[act$activity_type[i], act$activity_type[i]] <- 1
        }
      }
```

```
    }
    matrix
  }
```

```
  compute_plan_transition_matrix <- function(student_vle_data,
                                             assess_date, days) {

    print("Computing planning transition matrix")
    # sort data in ascending order of date
    df <- student_vle_data %>% arrange(date)

    # Count vle sequences one week before assess
    current_date <- df[0,] # prepare a dataset for the extraction
    # extract rows within num_week before each assess date
    for(i in assess_date) {
      temp <- df %>% filter((date <= i) & (date >= i - days))
      current_date <- rbind(current_date, temp)
    }

    df <- current_date

    # Count the number of use of each activity_type
    act <- df %>% group_by(activity_type) %>% summarize(num_use = n())

    # Build a matrix for storing transition probability
    matrix <- matrix(data = 0, nrow = length(act$activity_type),
                     ncol = length(act$activity_type),
                     dimnames = list(act$activity_type, act$activity_type))

    #... Now, we have the dataset current_date of rows about vle use num_week
before assess dates
    # Let's compute transition matrix
    # count the time an act in one date is followed by another act in the next
date
    if(nrow(df) > 0) {
      for (i in 1:nrow(df)) {
```

```
        # Consider activities in a day take turn to follow one another, extract
other rows of the same day as i
        same_date <- df %>% filter(date == df$date[i])
        if(nrow(same_date) > 0) {
          for (j in 1:nrow(same_date)) {
            matrix[df$activity_type[i], same_date$activity_type[j]] <-
              matrix[df$activity_type[i], same_date$activity_type[j]] + 1
          }
        }
        # extract rows whose date is the next of i
        next_date <- df %>% filter(date == df$date[i] + 1)
        # add 1 when there is a transition from activity in date i to activity
in date j
        if(nrow(next_date) > 0) {
          for (j in 1:nrow(next_date)) {
            matrix[df$activity_type[i], next_date$activity_type[j]] <-
              matrix[df$activity_type[i], next_date$activity_type[j]] + 1
          }
        }
      }
    }

    # divide each row of the matrix by the number of use of respective
activity_type
    if(nrow(act) > 0) {
      for(i in 1:nrow(act)) {
        # count the total number of sequences started with act$activity_type[i]
        total <- sum(matrix[act$activity_type[i], ])
        for( j in 1:ncol(matrice)) {
          matrix[act$activity_type[i],j] <-
                      matrix[act$activity_type[i],j] / total
        }
        # if a row of the matrix is NA, assign matrice[row, row] = 1
        if(is.na(matrix[act$activity_type[i],1])) {
          matrix[act$activity_type[i],] <- 0
          matrix[act$activity_type[i], act$activity_type[i]] <- 1
        }
      }
```

```
    }
    matrice
  }
```

```
  compute_reflect_transition_matrix <- function(student_vle_data,
                                        assess_date, days) {

    print("Computing reflection transition matrix")
    # Count the number of use of each activity_type
    act <- student_vle_data %>%
                  group_by(activity_type) %>% summarize(num_use = n())

    # sort data in ascending order of date
    df <- student_vle_data %>% arrange(date)

    # Build a matrix for storing transition probability
    matrice <- matrix(data = 0, nrow = length(act$activity_type),
                      ncol = length(act$activity_type),
                      dimnames = list(act$activity_type, act$activity_type))

    # Count vle sequences one week before assess
    # prepare a dataset for the extraction
     current_date <- df[0,]
    # extract rows within num_week after each assess date
    for(i in assess_date) {
      temp <- df %>% filter((date >= i) & (date <= i + days))
      current_date <- rbind(current_date, temp)
    }

    df <- current_date
    #... Now, we have the dataset current_date of rows about vle use num_week
before assess dates
    # Let's compute transition matrix
    # count the time an act in one date is followed by another act in the next
date
```

```
    # count the time an act in one date is followed by another act in the next
date
    if(nrow(df) > 0) {
      for (i in 1:nrow(df)) {
        # Consider activities in a day take turn to follow one another, extract
other rows of the same day as i
        same_date <- df %>% filter(date == df$date[i])
        if(nrow(same_date) > 0) {
          for (j in 1:nrow(same_date)) {
            matrix[df$activity_type[i], same_date$activity_type[j]] <-
                  matrix[df$activity_type[i], same_date$activity_type[j]] + 1
          }
        }
        # extract rows whose date is the next of i
        next_date <- df %>% filter(date == df$date[i] + 1)
        # add 1 when there is a transition from activity in date i to activity
in date j
        if(nrow(next_date) > 0) {
          for (j in 1:nrow(next_date)) {
            matrix[df$activity_type[i], next_date$activity_type[j]] <-
                  matrix[df$activity_type[i], next_date$activity_type[j]] + 1
          }
        }
      }
    }

    # divide each row of the matrix by the number of use of respective
activity_type
    if(nrow(act) > 0) {
      for(i in 1:nrow(act)) {
        # count the total number of sequences started with act$activity_type[i]
        total <- sum(matrix[act$activity_type[i], ])
        for( j in 1:ncol(matrice)) {
          matrix[act$activity_type[i],j] <-
                              matrix[act$activity_type[i],j] / total
        }
        # if a row of the matrix is NA, assign matrice[row, row] = 1
        if(is.na(matrix[act$activity_type[i],1])) {
```

```
          matrix[act$activity_type[i],] <- 0
          matrix[act$activity_type[i], act$activity_type[i]] <- 1
        }
      }
    }
    matrix
  }
```

Function `BuildMarkovChainModels` for step 4 of the procedure

```
BuildMarkovChainModels <- function(transitionMatrices) {

  list <- list()
  for(i in transitionMatrices) {
    print(paste("Building markov chain for ", comment(i), sep="-"))
    mc <- generate_markovchain(i)
    if(is.null(mc)) {
      print(paste("Failed to build markov chain for",
                  comment(i), sep = "-"))
      list[[length(list) + 1]] <-
                       (paste("Failed to build markov chain for",
                              comment(i), sep = " - "))
    }
    else {
      print(paste("Succeeded at building markov chain for",
                                  comment(i), sep = " - "))
      list[[length(list) + 1]] <- mc
    }
  }
  list
}
```

Function `GenerateSteadyStatesDataset` for step 5 of the procedure

```
  GenerateSteadyStatesDataset   <-   function(studentInfoData,   codeModule,
codePresentation, markovChainModels) {

    studentInfo <- studentInfoData %>%
            filter(code_module == codeModule,
                   code_presentation == codePresentation) %>%
            select(id_student)
    df <- matrix(nrow = 0, ncol = 3)
    df <- data.frame(df)
    colnames(df) <- c("id_student", "activity_type", "probability")
    for(i in 1:length(markovChainModels)) {
      if(!is.null(dim(markovChainModels[[i]]))) {
         # id[[1]][1] == id_student, id[[1]][2] == final_result
        id <- strsplit(name(markovChainModels[[i]]), "-")
        ss <- steadyStates(markovChainModels[[i]])
        print(paste("Generating steady states for student ",
                                                  id, sep=" "))
        for(j in 1:length(ss)) {
          ssrow <- data.frame(as.double(id[[1]][1]),
                                   dimnames(ss)[[2]][j], ss[[j]])
          names(ssrow) <- colnames(df)
          df <- rbind(df, ssrow)
        }
      }
    }
    # fill  dataframe with students unabled to generate ss
    df <- studentInfo %>% left_join(df, by = "id_student")

    # remove rows with NA
    df <- na.omit(df)

    # fill missing activity type for each student
```

```
    df <- df %>% complete(id_student, activity_type, fill = list(probability
= 0))


    # pivot wider the dataset to have activity_type as columns
    df <- df %>%
          pivot_wider(names_from = activity_type, values_from = probability)
    df
  }
```

Function `ComputeSRLIndex` for step 6 of the procedure

```
 ComputeSRLIndex <- function(steadyStatesData, vleUseData,
                                                    totalScoreData) {

   ssDf <- steadyStatesData
   scoreDf <- totalScoreData
   vleDf <- vleUseData

   srlScoreDf <- ssDf
   srlScoreDf$srl_index <- 0
   srlScoreDf$cognitive_score <- 0
   srlScoreDf$log_srl_index <- 0
   srlScoreDf$final_result <- "NA"
   for(i in 1:nrow(srlScoreDf)) {
     # step 1. Get the student id
     studentId <-
    as.character(srlScoreDf[srlScoreDf$id_student==srlScoreDf$id_student[i],
1])
     print(paste("Computing the SRL index for Student ", studentId, sep= " "))
     # Step 2. Get the steady state vector of the student
     ss <- ssDf[ssDf$id_student == studentId, 2:ncol(ssDf)]
     # Step 3. Get the vle use vector of the student
     vle <- vleDf[vleDf$id_student == studentId, 2:ncol(vleDf)]
     # Step 4. Get student's total score
     score <- scoreDf[scoreDf$id_student == studentId,]
     # Step 5. Compute the SRL score
     if(length(score$total_score)>0) {
       srlScoreDf$srl_index[i] <- rowSums(vle^ss) * score$total_score
       srlScoreDf$cognitive_score[i] <- score$total_score
     }
     if(length(score$final_result)>0) {
```

```
      srlScoreDf$final_result[i] <- score$final_result
    }
  }
  print("omputing the log SRL index for Students")
  srlScoreDf$cognitive_score[is.na(srlScoreDf$cognitive_score)] <- 0
  srlScoreDf$srl_index[is.na(srlScoreDf$srl_index)] <- 0
  srlScoreDf$log_srl_index <- log2(srlScoreDf$srl_index)
  srlScoreDf[is.infinite(srlScoreDf$log_srl_index), "log_srl_index"] <- 0
  srlScoreDf
}
```