ENHANCING STUDENTS' SELF-REGULATED LEARNING IN ONLINE LEARNING ENVIRONMENTS USING LEARNING ANALYTICS

by

DAN YE

(Under the Direction of Lloyd P. Rieber)

ABSTRACT

Online education has become popular in the last two decades and has played an important role in efforts to continue offering education to all levels of students during the coronavirus pandemic. Recent studies indicate that online learners must have the capacity to regulate their own learning in order to succeed in online courses. Measuring a learner's ability to regulate their own learning is difficult. Traditional self-reported self-regulated learning questionnaires have many limitations when used alone. The research study reported in this dissertation collected online learners' digital trace data recorded by a learning management system and compared that data with the results of self-reported data to evaluate which one accurately measures students' self-regulatory ability. In-depth interviews were conducted to better understand the causes accounting for the differences. A manuscript format consisting of three manuscripts has been chosen for the dissertation. The first manuscript (Chapter 2) is a historical review of learning analytics in the learning, design, & technology field, which provides a methodology foundation for the study. The second manuscript (Chapter 3) is a literature review of self-regulated learning theory and measurements, which provides a theoretical foundation for the design of the major study. The third manuscript (Chapter 4) is a mixed-methods study to investigate the differences and similarities between the digital trace data and self-reported SRL data as well as identify key learning behaviors that related to students' self-regulatory ability.

INDEX WORDS: Self-regulated learning, learning analytics, digital trace data, online learning environments

ENHANCING STUDENTS' SELF-REGULATED LEARNING IN ONLINE LEARNING ENVIRONMENTS USING LEARNING ANALYTICS

by

DAN YE

B.S.Ed., Huazhong Normal University, China, 2007

M.S.Ed., Southern Illinois University, 2012

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2021

ENHANCING STUDENTS' SELF-REGULATED LEARNING IN ONLINE LEARNING ENVIRONMENTS USING LEARNING ANALYTICS

by

DAN YE

Major Professor:	Lloyd P. Rieber
Committee:	Janette R. Hill
	Ikseon Choi
	Svoboda V. Pennisi

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2021

DEDICATION

I dedicate this dissertation to my husband Weihua and my kids Jacey and Anna, who have been with me through the journey and be supportive.

In addition, I dedicate this dissertation to my parents, who taught me the importance of education and the value of working hard and perseverance.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the effort and support of many people. I would like to thank both Dr. Jean Bertrand and Dr. Josef M. Broder, who encouraged me to pursue this degree and had been always supportive when I needed to take classes during my work time. To my co-workers in the Office of Online Learning especially James Castle, thank you for inspiring me to investigate the digital trace data from the learning management system. To all the wonderful people in the Learning, Design, and Technology program at the University of Georgia, thank you for helping me become a better scholar through sharing your ideas and experiences, questioning and inspiring me, and more.

To all my dissertation committee members, thank you for being part of my academic journey. I thank Dr. Lloyd Rieber for giving me the freedom to explore and find a research topic that I am interested in for my dissertation. Your support and guidance are always very helpful. Thanks to Dr. Janette Hill for your encouragement along the way. Thanks to both Dr. Janette Hill and Dr. Ikseon Choi for supporting me through the refinements of the research design. Thanks to Dr. Svoboda Pennisi for allowing me to collect data in her courses as well as providing helpful revising suggestions. Without your support, this dissertation would not be the same!

Finally, to my family, thank you for all the love, support, and encouragement. Thanks to my husband Weihua Zhou for all the unconditional support in these past few years. Thanks to my daughters Jacey Zhou and Anna Zhou, you are my motivation to achieve greatness.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The ability to self-regulate is one of the most important human capabilities (Zimmerman, 2005). Self-regulated learning (SRL) has become an essential skill in today's rapidly growing knowledge-driven society (Siadaty, Gasevic, & Hatala, 2016). SRL is defined as a process in which learners set their learning goals, monitor, and control their cognition, motivation, and behaviors to optimize their learning outcomes (Zimmerman,1989; Pintrich, 2000). The ability to self-regulate one's learning is particularly important within online learning contexts. Online education has become popular in the last two decades. The Covid-19 pandemic has pushed this trend further, with a distinctive rise in online learning. Although a variety of pressures have resulted in many students returning to traditional face-to-face learning as the two-year mark of the pandemic approaches, the impact remains and society as a whole seem more accepting of the idea of online learning (Adedoyin, & Soykan, 2020). In order for online learners to succeed in online courses, learners must have the capacity to regulate their learning (Lee, Choi & Kim, 2013; Hew & Cheung, 2014; Kizicec & Schneider, 2015) or receive active self-regulation support from the platform (Kizilcec & Cohen, 2017). Traditionally, online learners are non-traditional students and they generally are more self-disciplined and motivated than traditional students (Jinkens, 2009). Now with the impact of the Covid-19 pandemic, more and more students including traditional students are taking online courses. However, not every student has strong self-regulation abilities. Therefore, it is important to better understand how students self-regulate their learning in an online learning environment in the new situation and find out ways to improve their self-regulatory ability.

Self-regulated Learning Models and Measurements

Self-regulated learning is a topic that has been widely studied. Different scholars have proposed different SRL models (Zimmerman and Campillo, 2003; Pintrich, 2000; Boekaerts, 1996; Winne and Hadwin, 1998; Efklides, 2011; Jarvela and Hadwin, 2013). The evolution of these models shows a trend of shifting the focus from a general SRL process model to a more contextual SRL model. The earlier SRL models are goal-directed top-down self-regulation, such as Pintrich's cyclical model (Pintrich, 2000) of forethought, planning, and activation, monitoring, control, and reaction and reflection. Later, researchers started to emphasize the contextual bottom-up self-regulation, for example, Ffklides' Metacognitive and Affective Model of SRL (MASRL) (Efklides, 2011) proposed two levels of self-regulation: top-down and bottom-up self-regulation.

Meanwhile, several SRL measurement instruments have been developed to capture learners' self-regulation abilities: the Learning and Study Strategies Inventory (Weinstein, Palmer, & Schulte, 1987), the Self-Regulated Learning Interview Scale (Zimmerman & Martinez-Pons, 1986), the Motivational Strategies for Learning Questionnaire (Pintrich, Smith, Garcia, & McKeachie, 1993), the Online Self-regulated Learning Questionnaire (Barnard, Paton, & Lan, 2008), and Self-regulation in Self-paced Open and Distance Learning Environments (Kocdar, Karadeniz, Bozkurt, and Buyuk, 2018). However, most of these instruments are measuring learners' aptitude. Although the Self-Regulated Learning Interview Scale developed by Zimmerman and Martinez-Pons (1986) tries to elicit learners' SRL-related behaviors in a contextualized task, the task was usually fictitious. In addition, all these existing SRL measurements heavily rely on learners' self-reported data, which has been criticized as lacking validity (Cho & Summers, 2012; Zimmerman, 2008; Winne, 2010). There is an urgent need to identify or develop new ways to measure SRL in contextualized reality.

Towards A Solution of Using Learning Analytics to Enhance Self-regulated Learning

With the rapid development of technology and online education, we are able to track learners' digital traces collected from Learning Management Systems (LMS). To overcome the shortcomings of existing SRL measurements, one possible solution is to use learners' digital trace data such as clicking and interaction data collected by learning management systems as a supplement to the self-reported SRL data. Recent studies (Hwu, 2003; Yu & Zhao, 2015) have indicated that online student behavioral data are more accurate because the data collected from modern tracking technologies occur in real spontaneous learning situations. Although learners may be aware of the data collection taking place, it is relatively unobtrusive, so more authentic learning behaviors can be recorded.

However, data do not equate to useful information. Researchers often use different learning analytics methods to analyze data and try to identify useful information from large and messy data sets. Some important questions to consider are: How did the art and science of learning analytics originate? What learning analytics methods should be used to analyze these learners' trace data collected from the LMS? We also need to figure out a way to identify some important variables from the large and messy data so that we do not get "lost in the ocean" of data. The digital trace data from LMS are only some pieces of the big puzzle, which need to be interpreted in the context. How to use self-regulated learning theories and the learning context to interpret the data? These are all the questions we need to address while we are conducting any study using learning analytics to analyze the trace data.

Dissertation Design Rationale and Structure

To address the above questions, a manuscript-style structure is chosen for this dissertation consisting of three manuscripts. In the first manuscript (Chapter 2), I reviewed the related literature to reveal the origins and development of learning analytics from a historical perspective and the types of learning analytics methods were summarized. This manuscript provides a foundation for the learning analytics methods used in the Learning, Design, and Technology field as well as general guidance about conducting learning analytics research in the educational field. The second manuscript (Chapter 3) is a literature review illustrating the key existing SRL models and measurements. After comparing all the key existing SRL models, a new and more contextualized theoretical framework for SRL research in online learning environments is proposed to guide the study. This theoretical framework also embraces the impact of social interactions on learning. All key existing SRL measurements are compared and one instrument is identified to use for the study. The third manuscript (Chapter 4) is a mixed-methods study aimed to understand self-regulated learning using digital trace data from LMS by comparing the digital trace data with students' self-reported SRL data. Semi-constructed interviews were also conducted to better explain the results in context. The final chapter (Chapter 5) of the dissertation provides an overall summary of the results, the implications of this study, and suggestions for future research.

Table 1.1.

Summary of the Manuscripts and Their Purposes

Manuscripts	Purposes

Chapter 2: Manuscript 1 "The History and Development of Learning Analytics in the Learning, Design, and Technology Field"	Review related literature to illustrate the history and development of learning analytics in the Learning, Design, and Technology field.
Chapter 3: Manuscript 2 "A Theoretical Framework for SRL in Online Learning Environments"	Review related literature to explain SRL models and measurements and propose a new theoretical framework for SRL in online learning environments.
Chapter 4: Manuscript 3 "Use of Trace Data to Enhance Students' Self-Regulation: A Learning Analytics Perspective"	Investigate how trace data collected by the learning management system reflect online students' self-regulation by comparing it with students' self-reported SRL data.

Introduction to the Research Study

In order to overcome the limitations of existing SRL measurements, the research study reported in Chapter 4 was designed to use digital trace data from LMS to understand online learners' self-regulated learning in online learning environment by comparing the digital trace data with learners' self-reported SRL data. Specifically, the following research questions guided this study:

(1) How do the digital trace data collected by the learning management system reflect the students' self-reported SRL?

(2) What is the relationship between students' performance and the digital trace data and self-reported SRL data?

(3) What are the patterns of learning behaviors based on the digital trace data and self-reported SRL data?

(4) What are the explanations for any differences between their self-reported SRL data and the digital trace data?

Mixed methods were used to conduct the study. Three types of data were collected to better interpret online learners' self-regulated learning in context: the digital trace data from LMS, self-reported SRL data, and qualitative interview data. First, the digital trace data were compared with self-reported SRL data to see how do the digital trace data reflect the self-reported SRL data. Then, linear regressions were conducted to see which data can predict students' performance more accurately. Next, cluster analysis was conducted to identify the learning behavior patterns of each cluster. Lastly, differences between students' digital trace data and self-reported SRL data were compared and possible explanations were discussed in the context based on the qualitative interview data. Theme analysis was used to identify patterns of high and low self-regulated learners in the hope of providing some practical guidance for online teaching.

Contributions to the Field

This dissertation used a systematic approach to investigate self-regulated learning in online learning environments through a learning analytics lens. It demonstrates that it is important to have a theory to guide the data collection, which provides some guidance for future learning analytics studies in educational field. This dissertation also contributes to the field by emphasizing the importance of studying SRL in a contextual way. The results from the analyses of digital trace data should be interpreted in the context.

This study compares the digital trace data from LMS with self-reported SRL data systematically, which provides supportive evidence that digital trace data can predict performance more accurately than self-reported SRL data. It also shows that student interactions are important learning behaviors that should be included in the digital trace data, which has been neglected in existing studies. Finally, this study provides helpful information for both instructors and instructional designers about how to use the trace data collected from LMS to promote online learners' self-regulation and improve their academic achievements.

References

Adedoyin, O. B., & Soykan, E. (2020). Covid-19 pandemic and online learning: the challenges and opportunities. *Interactive Learning Environments*, DOI: 10.1080/10494820.2020.1813180.

Barnard, L., Paton, V., & Lan, W. (2008). Online self-regulatory learning behaviors as a mediator in the relationship between online course perceptions with achievement. *The International Review of Research in Open and Distance Learning, 9*(2), 1–11.

Boekaerts, M. (1996). Coping with stress in childhood and adolescence. In Zeidner, M. & Endler, N. S. (Ed.), *Handbook of coping: Theories, research, application* (pp. 452 -484). Oxford: John Wiley & Sons.

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

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

Hew K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review, 12*, 45-58.

Hwu, F. (2003). Learners' behaviors in computer-based input activities elicited through tracking technologies. *Computer Assisted Language Learning,16*(1), 5–29.

Jarvela, S., & Hadwin, A. F. (2013). New Frontiers: Regulating Learning in CSCL. *Educational Psychologist*, *48*(1), 25-39.

Jinkens, R. C. (2009). Nontraditional students: Who are they? *College Student Journal, 43*(4), 979-987.

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

Kizilcec, R. K., & Schneider. E. (2015). Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. *Transactions on Computer-human Interactions (TOCHI), 22*(2), 24.

Kocdar, S., Karadeniz, A., Bozkurt, A., & Buyuk, K. (2018). Measuring self-regulation in self-paced open and distance learning environments. *International Review of Research in Open and Distributed Learning*, *19*(1), 25-42.

Lee, Y., Choi, J., & Kim, T. (2013). Discriminating factors between completers of and dropouts from online learning courses. *British Journal of Educational Technology, 44*(2), 328-337. https://doi-org.proxy-remote.galib.uga.edu/10.1111/j.1467-8535.2012.01306.x.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In Boekaerts, M., Pintrich, P. R., & Zeidner, M. (Ed.), *Handbook of self-regulation* (pp. 451-502). San Diego, CA: Academic Press.

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

Siadaty, M., Gasevic, D., and Hatala, M. (2016). Trace-based microanalytic measurement of self-regulated learning processes. *Journal of Learning Analytics, 3*(1), 183-214. https://doi.org/10.18608/jla.2016.31.11

Weinstein, C. E., Palmer, D., & Schulte, A. C. (1987). *Learning and Study Strategies Inventory (LASSI)*. Clearwater, FL: H & H Publishing.

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist*, *45*(4), 267–276. http://dx.doi.org/10.1080/00461520.2010.517150

Yu, Q., & Zhao, Y. (2015). The value and practice of learning analytics in computer assisted language learning. *Studies in Literature and Language, 10*(2), 90–96.

Zimmerman, B. J. (2005). Attaining Self-regulation - A social cognitive perspective. In Boekaerts, M., Pintrich, P. R., & Zeidner, M. (Ed.), *Handbook of Self-Regulation.* (pp. 13–39). Burlington, MA: Academic Press.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166-183.

Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational Psychologist, 48*(3), 135-147.

Zimmerman, B. J., & Campillo, M. (2003). Motivating self-regulation problem solvers. In Davidson, J. E. & Sternberg, R. J. (Ed.), *The psychology of problem solving* (pp. 233–262). New York, NY: Cambridge University Press.

Zimmerman, B. J., & Martinez-Pons, M. (1986). Development of a Structured Interview for Assessing Student Use of Self-Regulated Learning Strategies. *American Educational Research Journal, 23*(4), 614-628.

CHAPTER 2

THE HISOTRY AND DEVELOPMENT OF LEARNING ANALYTICS IN LEARNING, DESIGN, & TECHNOLOGY[1]

[1] Ye, D. Submitted to *TechTrends, 10/15/2021*.

Abstract

This article introduces the evolution of themes and ideas related to the history, theory, and practice of learning analytics within the learning, design, and technology field through four eras. This review provides researchers with a fundamental understanding of the origin of learning analytics from a historical perspective and distinguishes learning analytics from educational data mining. This article also explores the relationship between theory and learning analytics, which provides some guidance for future related learning analytics research. Ethical concerns of learning analytics are also discussed in the hope of providing some guidance for practitioners and future research.

Keywords: learning analytics, educational data mining, course assessment and evaluation, theory

Originating from early work on educational data mining, learning analytics has become a trend in education research over the last few years. The goal of learning analytics is to use large data sets tracking learners as they proceed in completing courses of study in order to understand and optimize their learning and the learning environment. This paper introduces the evolution of themes and ideas related to the history, theory, and practice of learning analytics within the learning, design, and technology field, which provides a foundation for researchers to better understand the history and development of learning analytics.

The research approach consists of several gathering and analysis processes. First, a general timeline of the learning analytics was developed after scanning and analyzing abstracts in the field of Learning Analytics and Educational Data Mining. Then following the order of the timeline, a brief literature review was conducted around each milestone. The topics were selected based on whether the topic has a fundamental connection with or philosophical impact on the development of learning analytics.

Although the strong interest in learning analytics is relatively recent, it builds on a long-standing interest in how to use data to inform instructional design, such as course assessment and evaluation. The next section will introduce the history and evolution of learning analytics over four general eras.

The Pre-digital Era

Impact of the Philosophy of Behaviorism

In the pre-digital era, there was not much use of data in education. However, instructional systems technology evolved with a behaviorist foundation. Recognizing the importance of teaching educators to write objectives appropriately, Robert Mager wrote *Preparing Objectives for Programmed Instruction* in the early 1960s. The book describes how to write objectives for desired learner behaviors by specifying the conditions under which the behaviors are to be performed and the standards (criteria) by which the behaviors are to be judged. In 1963, Robert Glaser introduced the term "criterion-referenced measures" for assessing learners' behavior based on identified learning objectives. Although it may seem that a behaviorist philosophy is not related to learning analytics, the shift to behavioral learning outcomes represents a historical step toward quantitative learning data in the instructional design and technology field. Rodriguez (2013) pointed out that the large-scale online education model – MOOCs, such as Udacity, EdX, and Coursera are based on a behaviorist approach because they take traditional learning methods such as mastery learning and reinforcing concepts through interactive exercises. The use of observable and measurable learning objectives ensures that certain levels of learning can be measured, reinforced, and even graded by a machine such as a computer. If the learning process is a "black box," advocating behavioral learning objectives is like shining a beam of light to a corner of the black box. Although it may not have been the early behaviorists' original intention, the behaviorism philosophy inspires observing and analyzing learners' behavioral data in the education field. This eventually led, in turn, to the collection and analysis of learners' digital behavioral data in learning management systems.

Course Assessment and Evaluation

Course assessment and evaluation was one key area in the pre-digital era in which researchers collected data mainly from learners to evaluate and improve teaching and learning. Cronbach (2001) defined evaluation broadly as "the collection and use of information to make decisions about an educational program" (p.123). The field of educational evaluation has a long history of using standard tests covering the curriculum to assess the efficiency of the teachers or the school systems. The greatest expansion of systematic achievement testing occurred in the

1920s (Cronbach, 2001). After 1930, tests were given almost exclusively for judgments about individuals such as to select students for advanced training and to diagnose individual competencies and deficiencies (Cronbach, 2001). Educational researchers also frequently used tests, exams, or quizzes to assess how individual learning processes were impacted by novel educational interventions.

Two types of evaluation are distinguished by educational research for course assessment and evaluation. Formative evaluation is used to improve an instructional product while it is still in the development stage, while summative evaluation is used to assess the effectiveness of the final version of the product. Although its origins can be dated back much earlier, Scriven (1967) first used the term “formative evaluation” and “summative evaluation.” The idea of collecting data to evaluate an instructional product in the development stage greatly impacts not only the history of education but also the educational data collection and usage history.

The traditional method is to use quantitative data such as student performance to inform instructional adjustments. Later, quantitative course evaluation data extends beyond student performance data by examining students’ frequencies of behaviors from observations, using rubrics or checklists to derive quantitative data, and collecting technology-based course statistics.

Today, course assessment and evaluation still play a critical role in education. Researchers look for ways to use all kinds of related data to improve teaching and learning. For example, Data Wise (Boudett, City, & Murnane, 2005) proposed a step-by-step guide to improve teaching and learning using assessment results: (1) organizing for collaborative work; (2) building assessment literacy; (3) creating a data overview; (4) digging into data; (5) examining instructions; (6) developing an action plan; (7) planning to assess progress; and (8) acting and assessing. Although this step-by-step guide is proposed much later, the general course assessment and evaluation process are similar if not the same in the pre-digital era.

Both behaviorism and course assessment and evaluation still play important role in today's education, but they originated from the pre-digital era and have changed dramatically in the digital era. Nevertheless, both behaviorism and course assessment and evaluation provide a historical foundation for the role of learning analytics in using learners' quantitative data to inform teaching and learning.

The Rise of Computer Assisted Instruction

As early as 1954, B. F. Skinner and J. G. Holland implemented a programmed instruction education model at Harvard University. This model was characterized by self-paced self-administered instruction presented in a logical sequence. This method was applied through teaching machines. In 1954, B. F. Skinner wrote the article *The Science of Learning and the Art of Teaching* in which he pointed out the deficiencies of traditional instructional techniques and indicated that many of those problems could be overcome by using teaching machines. For several reasons, the programmed instruction movement ended in the late 1960s, but it inspired educators to turn attention to other forms of individualized instruction (Reiser, 1986).

The first generalized computer-assisted instruction (CAI) system, Programmed Logic for Automatic Teaching Operations (PLATO), started in the 1960s, but its use was very limited until the late 1970s. PLATO was designed and built at the University of Illinois and offered coursework to UIUC students, local schools, and other universities. According to Lyman (1975), by that time, the system had 950 terminals and users and had access to more than 3500 hours of instructional materials in more than 100 subject areas. The PLATO system had efficient communication systems between users, such as forums for exchanging notes, paging someone for immediate discussion, and sharing screens (Smith & Sherwood,1976). PLATO, as one of the first networked education and communication systems, plays an important historical role in the birth of learning management systems. Its social dimension of education greatly impacts the development of online education.

At the same time, adaptive computer-assisted instruction became popular. Proponents of this approach argued that the effects of instruction could be amplified by using principles of learning and memory, especially when using adaptive instruction based on the needs of individual learners and tasks. Richard Chatham Atkinson was a key scholar in early adaptive computer-assisted instruction. He published several papers to advance methods for optimizing the learning process. His main method was to optimize learning using a mathematical model based on students' performance history. The model selects learning materials and exercises based on the student's existing knowledge and its analysis of the skills required. For example, Atkinson (1972) proposed four strategies to optimize the learning of a large German-English vocabulary: the first strategy was to present items in random order; the second strategy permitted students to determine which item to present; the third and fourth strategies were based on a mathematical model of the learning process. The mathematical model was developed to present items in an order based on students' response history. The results showed that the two computer-controlled strategies based on the mathematical model overperformed the other two. Atkinson (1976) conducted a series of experiments in computer-assisted instruction to discover the optimal learning situation based on a mathematical model. The sequence of instruction varied according to a given student's performance history, and the model would modify itself automatically as more students' records are collected. Atkinson (1976) concluded that "an optimal instructional procedure could be derived if four elements are clearly specified: (1) the set of admissible instruction actions; (2) the instructional objectives; (3) a measurement scale that permits costs to be assigned to each of the instructional actions and payoffs to the achievements of instructional objectives; and (4) a model of the learning process” (p.361).

Later, Tennyson and Rothen (1979) designed the computer-based Minnesota Adaptive Instructional System (MAIS). Initially, the system used a program-controlled management system to select the appropriate amount and sequence of instructions for individual students. Realizing the importance of student responsibility in learning, Tennyson and Buttrey (1980) developed a learner-controlled mode of CAI by using the diagnostic and prescriptive information generated from the MAIS. They found that a learner-controlled condition could be an effective CAI management strategy if students received sufficient information about their progress and instructional needs necessary for obtaining the objective (Tennyson & Buttrey, 1980; Tennyson, 1981).

Since the late-1970s, various approaches and methods have been proposed to provide adaptive instruction to individual students. Park and Tennyson (1983) reviewed five adaptive computer-based instruction models: the mathematical model, the regression model, the Bayesian probabilistic model, the testing and branching model, and artificially intelligent instructional systems. The Bayesian probabilistic model and the multiple regression model were designed to select the amount of instruction required to learn a given task using both pre-task and on-task information (Park & Tennyson, 1980; Rothen & Tennyson, 1977). Pre-task information includes pre-task individual student aptitude and prior achievement. On-task information refers to individual student on-task learning data or responses. Participants who received the instruction based on both pre-task and on-task response data needed less learning time and resulted in better posttest performance than participants who were given instruction based on pre-task data alone

(Rothen & Tennyson, 1977). The results of the response-sensitive strategies studied by Park and Tennyson (1980) suggested that, in predicting students' learning needs for a specific learning criterion, the predictive power of on-task data is more efficient than pre-task data or pre-task plus on-task data.

In the early 1980s, with the advances in artificial intelligence and cognitive psychology, intelligent tutoring systems (ITS) diverged from their roots in computer-assisted instruction. The goal of ITSs was to provide adaptive instruction by intelligently diagnosing students' needs and progress in a response-sensitive manner. Generally speaking, an ITS has three main components: a student model for understanding and representing the knowledge state of a student, a domain model for representing the content to be taught, and a teaching model for making decisions about tutoring tactics by adopting an inherent instructional strategy (Shute & Psotka, 1996). Artificial intelligence methods make it possible to generate and present knowledge to students based on their performance on the task rather than selecting the knowledge according to the predetermined branching rules. For example, Anderson, Boyle, and Reiser (1985) introduced an intelligent tutoring system based on a set of pedagogical principles derived from the Adaptive Control of Thought (ACT) theory of cognition. To integrate the intelligent tutoring system with problem-solving, their intelligent tutoring system had a feature to provide new rules by summarizing a problem solution in an episode of learning. The ITS was able to provide personalized tutoring for students. When an incorrect response was given, the tutor provided appropriate remedial instruction. Two ITS tutors were introduced based on the intelligent tutoring systems: one was a Geometry tutor, which was used to help students learn geometry; the other was a LISP (a computer programming language) tutor, which was used to teach introductory programming at

Carnegie-Mellon University. Based on several evaluation studies done at the time, there was no difference between the ITS tutors and the human tutors (Anderson, Boyle, and Reiser, 1985).

In summary, during this era, although the personal computer market had recently been established, researchers were already developing CAI using different models to meet individual students' needs, optimize the learning process, and improve students' performance. The central idea in these efforts was to automatically adjust the learning content in response to individual students' historical performance and their ongoing user activity to personalize the learning process based on learning goals. The development of intelligent tutoring systems was a further step to use artificially intelligent methods to enhance learning and teaching with a focus on personalized learning. The use of individual learner's ongoing activity data to inform teaching and learning provides a solid foundation for the development of educational data mining and learning analytics.

The Rise of Online Learning and Birth of Educational Data Mining

The Rise of Online Learning

Although technologically mediated learning can be dated back to the early 1920s with radio courses, online learning greatly developed in the 1990s through the use of satellite virtual classrooms, video conferencing, and the Internet. In 1997, the "Interactive Learning Network" (ILN1.5) was released and installed at several academic intuitions including Cornell University, Yale Medical School, and the University of Pittsburgh. The ILN was an interactive classroom environment created using PCs and a software application, which was the first e-learning system of its kind. In the same year, Blackboard Inc. was founded, which developed a learning management system for course delivery and management. In 1999, Desire2Learn was founded, which became another widely used learning management system. In 2002, an open-source internal network named Modular Object-Oriented Dynamic Learning Environment (Moodle) was introduced, which was the first open-source learning management system. The emergence of learning management systems provided unique platforms with a great variety of channels and workspaces for participants to communicate and share information in a course. Instructors can share their lectures, prepare assignments and exams, and engage students in online discussions to enable collaborative learning, etc. The expansion of Internet access and advances in social media and mobile communications in the 2000s opened new avenues for online learning. In 2003, WebCT announced over 6 million students and 40,000 instructors teaching 150,000 courses per year at 1,350 institutions in 55 countries. By 2006, 89% of 4-year public colleges in the U.S. offered classes online, along with 60% of private institutions (Gensler, 2014).

The rise of online learning, especially the use of learning management systems, provides a convenient way to collect and record large amounts of data about students' learning including students' log files, their ongoing performance, their interactions with the system and the course content, etc. Record-keeping becomes an automatic process, which provides a necessary condition for the birth of educational data mining as well as learning analytics.

Birth of Educational Data Mining

Data mining, in particular, has become a research interest since the early 1990s. An example of an early approach to data mining is known as Knowledge Discovery in Database (KDD) and is defined as the process of developing methods or techniques to discover novel and potentially useful patterns from large amounts of real-world data (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). The term "Educational Data Mining (EDM)" initially emerged in a workshop series in 2005 (Baker & Inventado, 2014). According to the website of the Educational Data Mining community (www.educationaldatamining.org), educational data mining is a discipline that aims at developing methods for better understanding learners and the learning context by exploring unique and large-scale educational data. The first international conference on educational data mining was held in 2008 and this conference has been held annually since then.

Romero and Ventura published the first book on EDM - *Data Mining in e-Learning* in 2006. Romero and Ventura (2007) proposed a cycle of applying data mining in the educational system (see Figure 2.1). They believed that data mining techniques can discover useful information that can be used to improve the e-learning process from three perspectives: (1) it can be used to recommend learning activities, resources, and learning tasks to students based on the tasks already done by the learner or by other similar learners; (2) educators can get more objective feedback for their instruction and discover useful information to establish a pedagogical basis for decisions when designing or modifying a learning environment or teaching approach; and (3) academic administrators can discover useful information to optimize the network traffic flow and evaluate their educational programs and determine the effectiveness of the new computer-mediated distance learning approach. These three perspectives for applying data mining in education are similar to the purposes of evaluation proposed by Cronbach (2001), in which the author distinguished three purposes of evaluation: (1) course improvement, such as deciding what instructional materials and methods are satisfactory and what needs to be changed; (2) selecting, judging, or grouping pupils; and (3) judging the school system or individual teacher. This provides an important indication that EDM could be considered as a new form of evaluation in the digital era although now we focus more on improving teaching and learning.

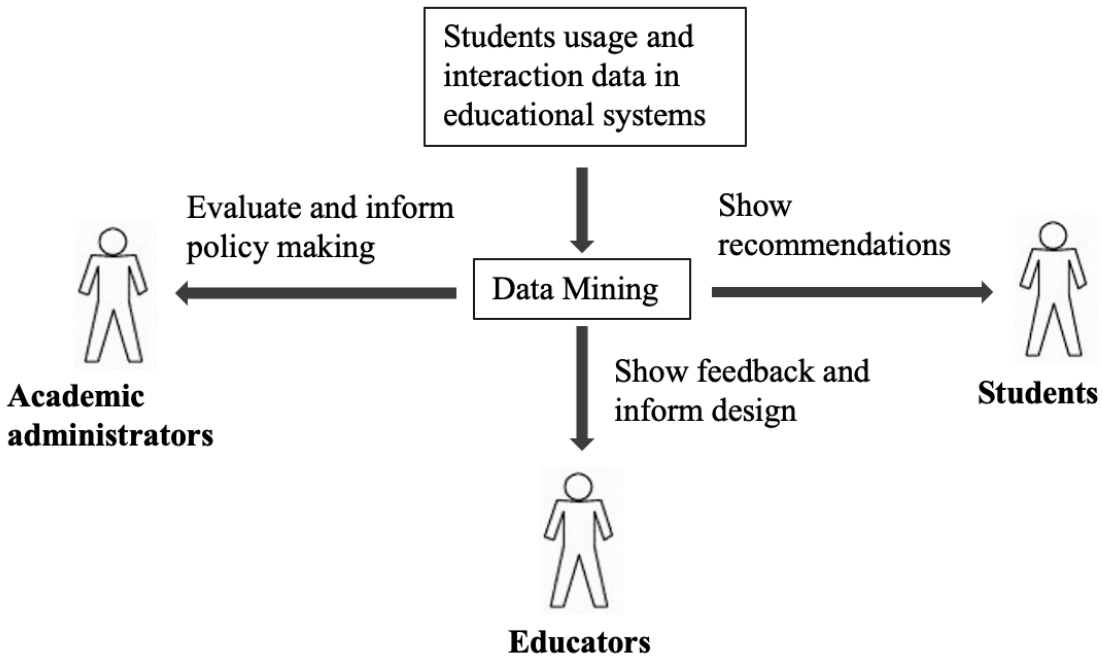


Figure 2.1. The Cycle of Applying Data Mining in Educational Systems. Adapted from Romero and Ventura (2007), p. 136.

Realizing that general data mining tools may not work well for data mining specifically in education, researchers started to explore data mining and machine learning techniques specifically designed for educational data mining. Zaiane (2001) discussed some data mining and machine learning techniques that could be used to enhance web-based learning environments. Merceron and Yacef (2005) presented a tool called the Tool for Advanced Data Analysis in Education (TADA-Ed), which was designed to mine data collected in an educational context. It included classification, clustering, and association rule algorithms to provide sensible information for teachers, such as mistakes made by students as well as concepts involved in these mistakes. Damez, Marsala, Dang, and Bouchon-Meunier (2005) presented a software tool called Tree Analysis for Providing Advice (TAFPA), which used a fuzzy decision tree technique to trace human-computer interactions, discriminate novice from experienced users automatically, and provide advice to the novice computer uses with contextual help.

Published research papers in EDM increased during this era. According to Romero and Ventura (2007), common data mining techniques include statistics and visualization, clustering, classification and outlier detection, association rule mining and pattern mining, and text mining. Baker and Yacef (2009) reviewed a comprehensive list of papers published between 1995 and 2005 and found that relationship mining methods and prediction methods were the most prominent types of EDM research between 1995 and 2005, accounting for 43% and 28% of these papers.

As more EDM-related research was conducted, the first issue of *The Journal of Educational Data Mining* was published in 2009, and the International Educational Data Mining Society was founded in 2011. Both events provide important milestones for the birth of learning analytics.

The Birth of the Society for Learning Analytics Research

The Society for Learning Analytics Research (SoLAR) was established in 2011, which is a leading international network of researchers who devote themselves to improve teaching and learning by analytics as well as promote the collaboration and publication of learning analytics research. SoLAR organized the first international conference on Learning Analytics and Knowledge (LAK) in the same year. In 2014, *The Journal of Learning Analytics* published its first issue, which is another important cornerstone established by SoLAR to support learning analytics research and practice. According to Gasevic, Mirriahi, Long, and Dawson (2014), *The Journal of Learning Analytics* is the first journal dedicated to “research investigating the challenges of collecting, analyzing, and reporting data with the specific intent to understand and improve learning” (p. 1).

Differences between Educational Data Mining and Learning Analytics

Researchers are often confused with the two research communities — educational data mining and learning analytics and sometimes use them interchangeably. So, what are the difference between these two communities? According to Siemens and Baker (2012), both communities share the goals of enhancing education by understanding educational problems and selecting interventions based on analysis of large-scale educational data. Siemens and Baker (2012) also pointed several key distinctions between these two communities. First, EDM has a considerably greater focus on automated discovery while learning analytics has a greater focus on leveraging human judgment. Second, EDM models are often used as the basis of automated adaption while learning analytics models are often designed to inform and empower instructors and learners. Finally, EDM research emphasizes reducing phenomena to components and analyzing individual components and relationships between them while learning analytics research typically emphasizes on understanding systems as wholes. EDM research uses classification, clustering, Bayesian modeling, relationship mining, discovery with models, and visualization while learning analytics research uses social network analysis, sentiment analysis, influence analytics, discourse analysis, learner success prediction, concept analysis, and sensemaking models. Similarly, Hoppe (2017) proposed the trinity of methods (Figure 2.2) for learning analytics, which includes: (1) network analysis including actor-actor (social) network and actor-artifact relations analysis, (2) processes analysis using methods of action patterns and sequence analysis, and (3) content analysis using text mining or other techniques to analyze learner-created artifacts.

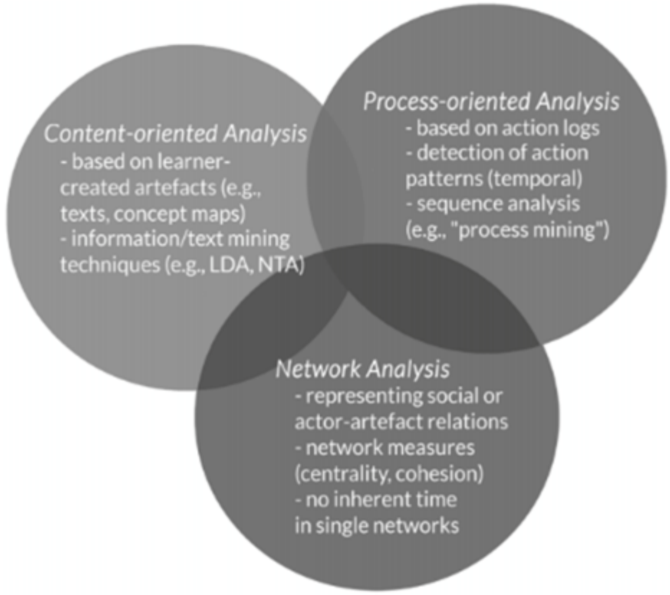


Figure 2.2. The "Trinity" of Methodological Approaches, by Hoppe, H. U. (2017), p. 26, Society for Learning Analytics Research. CC BY 4.0.

Development Trend in Technology-enhanced Learning

So, what do the technology-enhanced learning environments look like in the learning analytics era? Although learning management systems had been widely adopted by universities and colleges for several years along with more comprehensive technical support provided to students, content creation and delivery still heavily relied on experts' prior identification of the kinds of problems learners might encounter. In order to ensure that all learners are supported, a wider range of content needs to be developed to fit the expected needs and goals of these learners. Just as CAI had progressed from a program-controlled to a learner-controlled stage, the data-assisted approach to support instructional interventions in technology-enhanced learning environments moved from relying on the intelligence in the system to a dialogue between the intelligence in the system and intelligence of the instructional expert. The data used by the system include traces of learners' interaction with the online learning platform. These data traces link actors (mainly learners and instructors) and course content (e.g., videos, web pages, quizzes, online discussions) with interaction behaviors (e.g., reviewing, answering, clicking). Instead of using machine learning methods to automatically change the model and adapt to the individual learner, the instructional expert was given a central role in the process. Consequently, the instructional expert needed to form hypotheses of learning activity, validate the pedagogical relevance of patterns, and form instructional interventions (Brooks, Greer, & Gutwin, 2014).

In summary, although educational data mining and learning analytics are different, the development of technology-based learning shows a trend shifting from program-controlled to human-controlled. Therefore, the two research communities may merge into one in the future.

Challenges for Learning Analytics Research

In recent years, learning analytics has been used to provide feedback to students, predict student behavior, improve students' self-regulated learning, support or remedy teaching strategies, make decisions for the educational policymaker, etc. We are in a new era in which increasing numbers of researchers realize the power of learning analytics and begin to adapt learning analytics methods to achieve their educational goals. However, learning analytics research also faces several substantial challenges.

The Relationship with Theory

Greller and Drachsler (2012) pointed out that learners' digital behavioral data from LMS is only a single puzzle piece and needs to be interpreted with other puzzle pieces. In addition, learning analytics algorithms are reductive in nature and the result of which only represent simplified reality. As learning analytics continues to develop, a critical question turns on how theory can or should shape research in learning analytics. Wise and Shafer (2015) suggested that theory is more important in interpreting results and identifying meaningful, actionable results when working with big data. Bearing this question of how theory should guide learning analytics research in mind, Knight and Shum (2017) argued that the use of learning analytics tools is always aligned with assessment regimes and grounded in epistemological assumptions and pedagogical practices. They also provided six questions to be considered in the development of learning analytics: (1) Epistemology — what are we measuring? (2) Epistemology — how are we measuring? (3) Pedagogy — why is this knowledge important to us? (4) Pedagogy — who is the assessment/analytic for? (5) Assessment —where does the assessment happen? (6) Assessment — when do the assessment, and feedback, occur? These questions could serve as a theoretical foundation for learning analytics research.

While it is important to use theory to guide and illustrate the learning analytics results, we also need to be aware that different theoretical constructs may interpret the raw data differently, leading to different consequences in terms of decision making (Greller & Drachsler, 2012).

Ethical Concerns

With the wide use of learning analytics, more and more researchers have recognized the importance of the privacy of student information and started to pay attention to the ethical issues of using digital trace data in education. Hakimi, Eynon, and Murphy (2021) conducted a systematic qualitative analysis of research in the ethics of digital trace data use in education and found that privacy, informed consent, and data ownership are the most commonly and extensively discussed ethical issues. The authors also pointed out that digital trace data tend to only include information that is easily measurable and quantifiable, which may lead to biased interpretations and discrimination. Greller and Drachsler (2012) pointed out one of the principal shortfalls of learning analytics is that it tends to focus on predict average behaviors not outliers, which may reinforce peer pressure, segregation, and conformism.

When conducting research using digital trace data, it is important to follow ethical rules by getting consent from the participants and protecting their privacy by anonymizing the data. One common approach is to analyze possible consequences of the study on participants and society based on consequentialist ethics theory and weigh if the study is beneficial for most people (Willis, 2014). In addition, it is important to examine the power relationships in individual contexts and provide safeguards to the vulnerable individuals affected by the consequences of the study. Hakimi, Eynon, and Murphy (2021) also suggested the need for greater transparency in data privacy, access, and ownership as well as analyzing the power relationships to eliminate potential bias or discrimination in the interpretation of digital trace data. Similarly, Greller and Drachsler (2012) suggested that learning analytics should take a bottom-up approach by focusing on the interest of the learners because they are vulnerable individuals among the power relationships.

Conclusion

In summary, from a philosophical point of view, learning analytics was heavily influenced by behaviorism. From a historical perspective, learning analytics emerged from the development of technology and the birth of learning management systems and online education. Traditionally, teachers' observations of learners' learning behaviors were valued. Through everyday observation in the classroom, teachers tried to determine what students struggle with the most and subsequently how much practice students would need to master the material. With the advance of technology, learning management systems are able to record every learning behavior exhibited by learners as they interact with the system. Of course, an important difference is that a learning management system cannot judge which behavior is important, so additional data analysis is needed to identify meaningful information and apply it to enhance the teaching and learning practice. Learning analytics emerged during the process and can be considered as a new form of course assessment or evaluation in the digital era, which will play an increasingly important role in educational research and practices. But we also should be cautious of using any digital trace data for any sort of student evaluation or assessment simply because we are not ready yet to understand the full ethical implications of its use.

References

Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent Tutoring Systems. *Science, 228*(4698), 456-462.

Atkinson, R. C. (1972). Optimizing the learning of a second language vocabulary. *Journal of Experimental Psychology, 96*(1), 124 – 129.

Atkinson, R.C. (1976). Adaptive instructional systems: Some attempts to optimize the learning process. In Klahr, D. (Ed.), *Cognition and Instruction* (pp. 81-108). New York: Wiley.

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In Larusson, J. A. & White, B. (Eds.), *Learning Analytics from Research to Practice* (pp.61-75). New York: Springer.

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*(1), 3-17.

Brooks, C., Greer, J., & Gutwin, C. (2014). The data-assisted approach to building intelligent technology-enhanced learning environments. In Larusson, J. A. & White, B. (Ed.), *Learning Analytics from Research to Practice* (pp.123-156). New York: Springer.

Cronbach, L. J. (2001). Course improvement through evaluation. In Ely, D. P., & Plomp, T. (Ed.), *Classic Writings on Instructional Technology* (pp.123-134). Englewood, Colo: Libraries Unlimited, Inc.

Damez, M., Marsala, C., Dang, T., & Bouchon-Meunier, B. (2005). Fuzzy decision tree for user modeling from human–computer interactions. *In International conference on human system learning: Who is in control?* 287–302.

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery: An overview. In Advances in *Knowledge Discovery and Data Mining* (pp. 1-30). Menlo Park, Calif.: AAAI/MIT Press.

Gasevic, D., Mirriahi, N., Long, P., & Dawson, S. (2014). Editorial: Inaugural issue of the Journal of Learning Analytics. *Journal of Learning Analytics, 1*(1), 1-2.

Gensler, L. (2014). From correspondence courses to MOOCs: The Highlights of Distance Learning over the ages. *Forbes Magazine*, 12 Feb.

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

Hakimi, L., Eynon, R., and Murphy, V. A. (2021). The ethics of using digital trace data in education: A thematic review of the research landscape. *Review of Educational Research, 91*(5), 671-717.

Hoppe, H. U. (2017). Computational methods for the analysis of learning and knowledge building communities. In Lang, C., Siemens, G., & Wise, A., & Gasevic, D. (Ed.), *Handbook of Learning Analytics* (pp.23-33). SOLAR.

Knight, S. & Shum, S. B. (2017). Theory and Learning Analytics. In Lang, C., Siemens, G., & Wise, A., & Gasevic, D. (Ed.), *Handbook of Learning Analytics* (pp. 17-22). SOLAR.

Lyman, E. R. (1975). PLATO Curricular Materials (Computer-Based Education Research Laboratory, Urbana, Ill.), report X-41, No. 3.

Merceron, A. & Yacef, K. (2005). Tada-ed for educational data mining. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning, 7*(1), 267-287.

Park, O. and Tennyson, D. R. (1980). Adaptive design strategies for selecting number and presentation order of examples in coordinate concept acquisition. *Journal of Educational Psychology, 72*(3), 362-270.

Park, O., & Tennyson, R. D (1983). Computer-based instructional systems for adaptive education: A review. *Contemporary Education Review, 2*(2), 121-135.

Reiser. A. R. (1986). Instructional technology: a history. In Gagne R. M. (Ed.), *Instructional Technology: Foundations*. Hillsdale, NJ: L. Erlbaum Associates.

Rodriguez, O. (2013). The concept of openness behind c and x-MOOCs (Massive Open Online Courses). *Open Praxis, 5*(1), 67-73.

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

Rothen, W. and Tennyson, D. R. (1977). Pre-task and on-task adaptive design strategies for selecting number of instances in concept acquisition. *Journal of Educational Psychology, 69*(5), 586-592.

Scriven, M. (1967). The methodology of evaluation. In Tyler, R., Gagne, R. & Scriven, M. (Ed.), *Perspectives on Curriculum Evaluation* (AERA Monograph Series – Curriculum Evaluation). Chicago: Rand McNally and CO.

Shute, V. J. & Psotka, J. (1996). Intelligent tutoring systems: past, present, and future. In Jonassen, D. (Ed.), *Handbook of Research on Educational Communications and Technology* (pp. 570-600). New York: Macmillan.

Smith, S. G. & Sherwood, B. A. (1976). Educational uses of the PLATO computer system, *Science, 192*(4237), 344-352.

Siemens, G., & Baker, R. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 252–254.

Tennyson, R. D. (1981). Use of adaptive information for advisement in learning concepts and rules using computer-assisted instruction. *American Educational Research Journal*, *18*(4), 425-438.

Tennyson, R. D., & Buttrey, T (1980). Advisement and management strategies as design variables in computer-based instruction. *Educational Communications and Technology Journal*, *73*(3), 326-334.

Tennyson, R. D., & Rothen, W (1979). Management of computer-based instruction: Design of an adaptive control strategy. *Journal of Computer-Based Instruction*, *5*, 126-134.

Willis, J. E. (2014, August 25). Learning analytics and ethics: A framework beyond utilitarianism. EDUCAUSE Review. https://er.educause.edu/articles/2014/8/learning-analytics-and-ethics-a-framework-beyond-utilitarianism

Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics, 2*(2), 5-13.

Zaïane, O. (2001). Web usage mining for a better web-based learning environment. *In Proceedings of conference on advanced technology for education*, 60–64.

CHAPTER 3

A THEORETICAL FRAMEWORK FOR SELF-REGULATED LEARNING IN ONLINE LEARNING ENVIRONMENTS IN THE DIGITAL ERA[2]

[2] Ye, D. To be submitted to *The International Review of Research in Open and Distributed Learning.*

Abstract

With the development of technology and online education, learners' digital trace data can be collected and used to improve teaching and learning. Being able to self-regulate is very critical for students to succeed in online learning due to the isolated nature of online learning. This paper proposes a comprehensive theoretical framework for self-regulated learning research in online learning environments by incorporating learners' digital trace data collected from a learning management system. This theoretical framework is developed based on a comprehensive literature review on existing self-regulated learning models and measurements, which could provide a fundamental understanding for interpreting and the use of digital trace data for educational research.

Keywords: self-regulated learning, online learning environments, digital trace data, self-regulated learning models, self-regulated learning measurement instruments

Online education has become popular in the 2000s and it has been playing an important role during the coronavirus pandemic. However, not every student succeeds in online education. Recent studies indicate that self-regulated learning appears to have a positive impact on students' academic performance, which is especially important for online learning due to the isolated nature of online learning (Barnard-Brak, Paton, & Lan, 2010; Zhu, Zhang, Au, & Yates, 2020). In general, online students are usually adults (professionals with a full job), mature and motivated (for example, career advancement) to complete the online program or certificate they choose to pursue. However, due to the coronavirus, millions of traditional college students had no choice but to take courses online, which increased the concerns about the quality of their learning. Even when as the pandemic situation improves and the health crisis ends, it is very unlikely the impact of online learning will be reduced. More students will likely choose to take online courses for many reasons, such as due to convenience. While it is important to improve the quality of online teaching from the teachers' perspective, it is also worthy to investigate more about learners' Self-Regulated Learning (SRL) in online learning environments during this period. This leads to the question of how to measure and foster students' self-regulated learning in online courses. With the development of technology and the wide use of learning management systems, learners' digital trace data can be easily collected from online courses thereby providing a new approach to investigating online teaching and learning. So how do we measure and foster online students' self-regulated learning in the digital era? What learning behavioral variables should be collected through the digital trace data and why? A good place to start is to review the history and theory of SRL to identify or develop a theoretical framework for conducting related research.

Origins of Self-regulated Learning

Self-regulated learning (SRL) is a concept developed in the field of educational psychology and cognitive psychology in the 1980s (Siadaty, Gasevic, & Hatala, 2016). Zimmerman initially considered "three possible definitions of SRL: as an ability, as a behavior, or as a self-belief" (Zimmerman, 2013, p.137). However, the definition of SRL was broadened by adding the learner's dynamic interplay within the social and physical environment. By further including key metacognitive processes and motivational variables, Zimmerman (1989) defined SRL as the degree to which students are "metacognitively, motivationally, and behaviorally active participants in their own learning process" (p. 329).

After examining different SRL models, Pintrich (2000) argued that these models share four basic assumptions about learning and regulation. Based upon these assumptions, Pintrich (2000) defined SRL as an active constructive learning process in which learners set learning goals and then attempt to monitor, regulate, and control their cognition, motivation, and behavior, guided and constrained by their goals and the contextual features in the environment.

Self-directed Learning and self-regulated Learning

It is worth noting that some scholars use SRL and self-directed learning (SDL) interchangeably (Gandomkar & Sandars, 2018), so it is important to distinguish between these two terms and to explain why SRL is the chosen term in this article. SDL is a concept that originated from adult education in the 1970s-1980s. Knowles (1975) described SDL as a process in which individuals take the initiative in diagnosing their learning needs, formulating learning goals, identifying human and material resources for learning, choosing and implementing appropriate learning strategies, and evaluating learning outcomes. Due to its adult education roots, SDL is mostly used for describing the learning activities outside traditional school environments and involves aspects of designing learning environments (Saks & Leijen, 2014).

Research on SRL, however, mostly involves studies in a school environment (Loyens, Magda, & Rikers, 2008). Nevertheless, SRL should not exclude the possibility of designing personal learning environments. SDL has been treated as a broader concept in which learners have more freedom to manage and control their learning activities and learning process. In SDL, it is the learner who defines the learning task, while in SRL it may be a teacher who defines the learning task (Loyens, Magda, & Rikers, 2008). Based on the differences between these two terms and the research context of the school environment, SRL is the more appropriate term for this paper.

The Evolution of Self-regulated Learning Models

Several researchers have proposed SRL models. This section will describe the historical evolution of these models.

Zimmerman's Cyclical Phases Model

Influenced by Bandura's triadic analysis of human functioning (i.e. personal, behavioral, and environmental) (Bandura, 1986), Zimmerman initially proposed a triadic social cognitive model for SRL (Zimmerman, 1989), which includes three forms of SRL: the behavioral forms of self-regulation refer to "self-observing one's performance and adapting it strategically" (Zimmerman, 2013, p. 137); the environmental forms of self-regulation involve "monitoring the effects of varying environmental conditions and controlling those conditions strategically" (Zimmerman, 2013, p. 137); the convert forms of self-regulation refer to "observing and adapting specific feelings and thoughts" (Zimmerman, 2013, p. 138). During this process, feedback plays a critical role to guide strategic adaptations in each form of self-regulation.

Zimmerman (2013) also pointed out that these three forms of self-regulation are interdependent, and an optimal self-regulatory intervention would target all three forms.

By incorporating causal relationships, motivational beliefs, and metacognitive strategies, Zimmerman (2000) later proposed a cyclical model of SRL based on social cognitive theory. According to this model, a student's learning processes fall into three phases shown in Figure 3.1: forethought, performance, and self-reflection. In the forethought phase, learners analyze the task, set goals, and execute strategic planning. A number of self-motivational beliefs, such as self-efficacy, intrinsic interest or value, outcome expectations, and goal orientation, energize the process and influence the activation of learning strategies. In the performance phase, the students actually execute the task while they monitor how they are progressing using a number of self-control and self-monitoring strategies to keep themselves cognitively engaged and motivated to finish the task. In the self-reflection phase, students evaluate how they have performed the task, making attributions about their success or failure. These attributions generate self-reactions that can positively or negatively influence the forethought process in later performance, thus, leading to a self-regulatory cycle.

Zimmerman (2013) pointed out that this cyclical property not only explains the results of a repeated effort to learn from a quantitative perspective but also explains a major qualitative difference in students' self-regulation by comparing proactive learners and reactive learners. Proactive learners can plan strategically by setting up specific challenging goals and self-evaluate based on mastery of the goals, while reactive learners rely heavily on post-performance by comparing with others to self-evaluate (Zimmerman, 2013). Because of their superior task analysis and self-control process, proactive self-regulators are expected to display a superior cyclical pattern of processes than reactive self-regulators (Zimmerman, 2013).

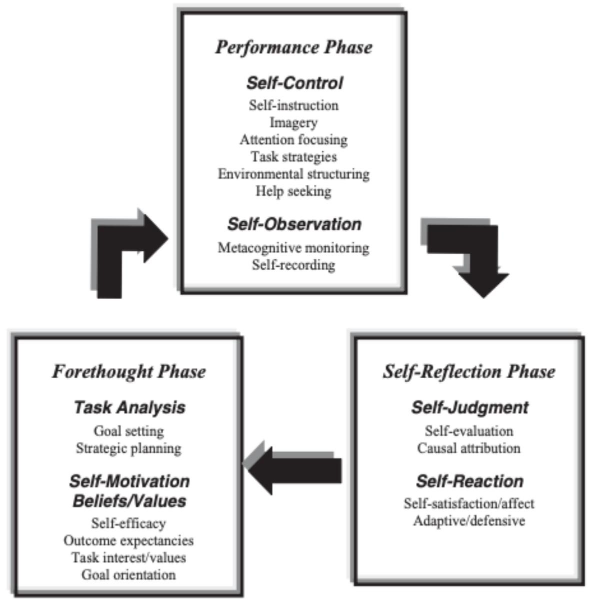


Figure 3.1. Phase and Subprocesses of Self-regulation. From *Zimmerman and Campillo (2003),* p. 239. Used with the written permission of the publisher.

Pintrich's Self-regulated Learning Model

Pintrich (2000) extended Zimmerman's cyclic SRL model into four phases: 1) forethought, planning, and activation, 2) monitoring, 3) control, and 4) reaction and reflection.

In the phase of forethought, planning, and activation, learners perceive upcoming academic tasks, set target goals, activate prior knowledge, and form plans on how to achieve the tasks and goals. Goal setting is an important part of this phase (Zimmerman 2000, 2008). Goals can be divided into two categories: proximal (short term) goals and distal (long term) goals. Zimmerman (2008) showed that appropriate proximal goals are more effective than distal goals to improve immediate task performance. Learners' two major goal orientations have also been studied in self-regulation literature: mastery approach goals and performance approach goals (Pintrich, 2000). Mastery approach goals concern learning, improvement, and increasing one's skills and knowledge, while performance approach goals involve demonstrating competence and outperforming others.

In the monitoring phase, learners monitor their cognition, motivation, effort, behavior, and context conditions. According to Pintrich, Wolters, and Baxter (2000), cognitive monitoring involves judgments of learning and comprehension monitoring. For example, learners may become aware that they do not understand something they have just read or become aware that they are reading too quickly or slowly given the text and their goals. Learners also cognitively monitor their motivation, effort, behavior, and context conditions. For example, learners may become aware that they need to read the required readings more carefully in order to discuss reading-related topics in the online discussion activity and therefore may adjust time and effort accordingly.

In the controlling phase, learners select and adapt cognitive strategies for learning and thinking, select and adapt motivational strategies, decide to increase or decrease effort, make decisions about whether to seek help or persist or give up, and determine whether to change or renegotiate the task. The control of cognition involves the actual selection and use of various cognitive strategies for memory, learning, reasoning, problem solving, and thinking (Pintrich, 2000). Pintrich (2000) also summarized some strategies of motivational control based on existing research, which include learners' control of self-efficacy through the use of positive self-talk, increasing extrinsic motivation by promising themselves extrinsic rewards, increasing the task value by making it more relevant or useful to them, and the use of defensive pessimism to deal with negative motivations like anxiety and fear.

In the reaction and reflection phase, learners take actions based on their cognitive judgments and motivational strategies and then evaluate the task to see whether their reactions are effective or not. Reactions are mainly reflected in time and effort patterns that learners spend studying the learning task. Seeking help is another reaction to the self-regulated learning process. Learners also make attributions for their performance and make reflections in this phase. Studies have shown that good self-regulated learners are more likely to make adaptive attribution for their performance, which means that they attribute lower performance to low effort or poor strategy use, not a lack of general ability (Zimmerman, 1998).

Boekaerts' Dual Processing Self-Regulation Model

By arguing that it is important to create a powerful learning environment to enable students' self-regulation, Boekaerts (1997) proposed a six-component SRL model. The model consists of six cubes: Three cubes on the left represent cognitive self-regulation, and the other three on the right represent motivational self-regulation. Three levels are distinguished from bottom to top: the domain-specific level, the strategic level, and the goal level. Boekaerts also distinguished three aspects of motivational self-regulation: motivational beliefs, motivation strategies, and motivational regulatory strategies (Boekaerts, 1997).

Later, assuming that all learning processes are behavioral changing processes, Boekaerts proposed a dual processing self-regulation model shown in Figure 3.2 (it is also called the model of adaptable learning), which focuses on explaining the role of appraisals and affective states (Boekaerts, 2011). One of the underlying assumptions of this model is that individuals inherently self-regulate in terms of two parallel goals: (1) they want to extend their knowledge and skills so that they can expand their personal resources; (2) they wish to maintain their available resources, thus preventing loss and distortions of well-being (Boekaerts, 1995). It was theorized that each learning situation triggers a dynamic internal working model, which is constantly fed information from three main sources: (1) task in context, which is the perception of the learning situation, including the task, the instruction given by the instructor, and the physical and social context; (2) knowledge and skills, which covers activated domain-specific knowledge and skills, including declarative and procedural knowledge, cognitive strategies that have been successful in that domain, and metacognitive knowledge relevant to the learning situation; and (3) the learners' self-system, which includes their goal hierarchy, values, and motivational beliefs of the domain that is activated by the situation. The model introduces two pathways: the growth pathway and the well-being pathway. Students' appraisals are crucial in determining which pathway students will take. If students perceive that they might not be successful in a task, negative cognition and emotions are triggered, and students move onto a well-being pathway to protect their ego from damage. On the other hand, if the task meets with students' needs and interests, it may trigger positive cognitions and emotions, and thereby they move onto the mastery/growth pathway.

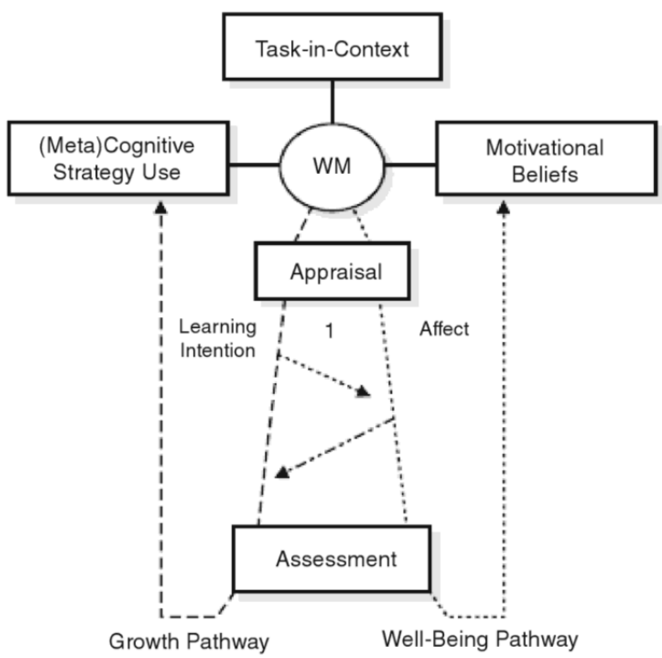


Figure 3.2. Dual Processing Self-Regulation Model. From *Boekaerts (2011),* p. 410. Used with the written permission of the publisher.

Winne and Hadwin's COPES Model

Winne and Hadwin (1998) used a five-facet typology as a schema to characterize instances of self-regulated learning as shown in Figure 3.3: conditions, operations, products, evaluations, and standards (COPES). Viewing SRL as an event, Winne and Hadwin (1998) theorized a complete model of SRL with four basic phases: task definition, goal setting and planning, enactment, and adaptation. The COPES model assumes that learning is goal-directed. In phase 1, the learner generates perceptions of the task at hand, what constraints and resources are in place. Two kinds of perceptions are included in this model: (1) task conditions, providing information about the task in the environment; (2) cognitive conditions, which are memorial representations of similar past tasks. In phase 2, learners frame goals and assemble a plan for addressing the study task. In phase 3, learners apply tactics and strategies identified in phase 2 to achieve the settled goals. Here, tactics are bundles of memories comprised of conditional knowledge and cognitive operations. In phase 4, learners make major adaptations based on the situations by restructuring cognitive conditions, tactics, and strategies to create different approaches to address the task. It is worthy to note that SRL is a recursive process that information generated in each or any subsequent phase may jump phases or recur to the same phase.

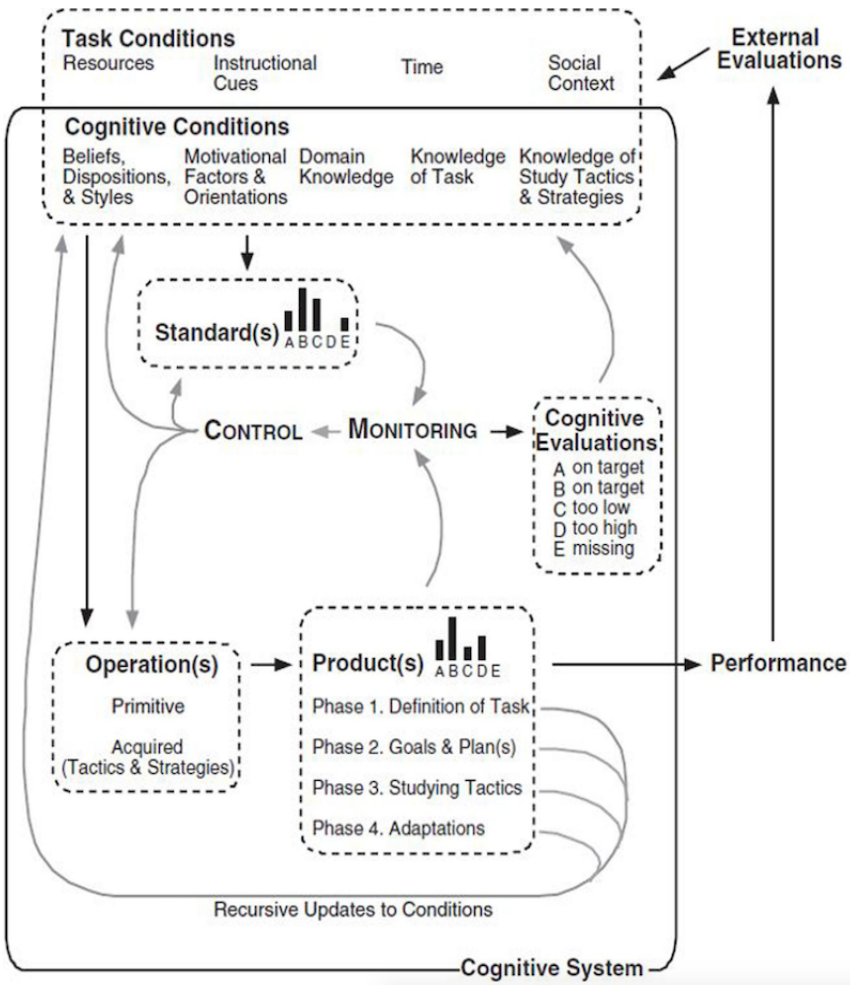


Figure 3.3. The COPES Model. From Winne and Hadwin (1998), p.127. Used with the written permission of the publisher.

Efklides' MASRL

By highlighting the interactions among metacognition, motivation, and affect in SRL, Efklides (2011) proposed the Metacognitive and Affective Model of SRL (MASRL). In this model, the interactions serve two modes of self-regulation – top-down and bottom-up self-regulation. More specifically, these two levels in this model are the person level/macro level and the Task × Person level/micro level. Efklides described that the person level represents "a generalized level of SRL functioning" (Efklides, 2011, p. 10), which sets goal-directed, top-down self-regulation. The Task × Person level is the level where the task is processed, functioning in the data-driven, bottom-up self-regulation. Efklides criticized that Zimmerman's description of the self-regulated model was limited to a macro level, which ignores spontaneous mastery experience such as metacognitive feelings and affects that serve a bottom-up self-regulation mode (Efklides, 2011). He argued that the resources of the personal level are relatively stable, while learners can more easily modify or change their decisions in the bottom-up self-regulation depending on the monitoring of task processing and subjective experiences. Efklides (2011) believed that four basic functions operate at the bottom-up self-regulation level – cognition, metacognition, affect, and regulation of affect and effort. Efklides (2011) also pointed out that "the Person and the Task × Person levels interact and inform each other" (p.10). This model emphasizes the role of subjective experiences, particularly mastery experience, which is believed as "the manifestation of online monitoring and trigger control processes" (Efklides, 2011, p. 21).

Jarvela and Hadwin's SRL Model

Jarvela and Hadwin proposed three modes of regulation in the context of collaborative learning: self-regulation (SRL), co-regulation (CoRL), and shared regulation (SSRL) (Jarvela & Hadwin, 2013). SRL refers to the individual learner's regulatory actions that emerge through a series of exchanges among group members. CoRL refers to regulatory actions guided or directed by a particular group member(s). SSRL refers to regulatory actions that occur within a group. Because co-regulation and shared regulation are not the primary focus of this paper, this model will not be covered at length hereafter.

Comparing these SRL Models

A summary comparison of these SRL models is shown in table 3.1. Although Pintrich's cyclical phases model has been widely used, it is more of a general process model. Zimmerman's SRL model and the COPES model are similar to Pintrich's cyclical phases model in that they all focus on the self-regulatory process and are goal-driven. Boekaers, Pintrich, and Zeidner (2000) also agreed that Zimmerman's cyclical phase model is a general self-regulation model. Based on the history of the SRL model development, Pintrich's cyclical phase model is an extension of Zimmerman's model, so both of these two models could be considered as general self-regulation models.

Table 3.1.

Self-regulated Learning Models Comparison

Model name	Perspective	Key Phases/Structure
Zimmerman's cyclical phases model	Social cognition	A cyclical model of forethought, performance, and self-reflection.
Pintrich's SRL model	Social cognition	A cyclical model of forethought, planning, and activation, monitoring, control, and reaction and reflection.
Boekaerts' dual processing model	Goals & motivation	The growth pathway and well-being pathway.
The COPES model	Goal-directed	Task definition, goal setting and planning, enactment, and adaptation.
Efklide's MASRL	The interactions among metacognition, motivation, and affect	Two levels: the person level (top-down self-regulation) and the Task×Person level (bottom-up self-regulation).
Jarvela and Hadwin's SRL Model	Collaborative learning	Three modes: self-regulation (SRL), co-regulation (CoRL), and shared regulation (SSRL).

Boekaers' dual processing model proposes a different perspective to investigate self-regulated learning behaviors, which might better fit online learning environments than the models reviewed above. Online learners' digital traces in online courses represent their behaviors and the patterns of these digital traces may reflect their self-regulated learning process. Boekaerts' dual processing model provides an excellent model for analyzing learners' behavior changes based on two types of goals: learning goals and ego-protecting goals. However, if we take a deep look at these models, it is clear that Zimmerman's cyclical phases model, Pintrich's SRL model, Boekaerts' dual processing model, and Winne and Hadwin COPES model are all goal-driven and focus on top-down self-regulation. Three of these four models categories learners into two groups: Zimmerman (2013) pointed out two types of learners – proactive and reactive learners; Pintrich (2000) distinguished two goal approaches – mastery and performance approach goals; Boekaerts (1995) defined the growth and well-being pathway. It appears that all learners can always be divided into two groups based on any criteria. Similar to asking a yes/no question, there will be always a group of "yes" people and a group of "no" people. It is worth questioning if this reflects reality. These four SRL models also have a common limitation: they all focus on the learners' top-down level of self-regulation but neglect the importance of learning context and learners' subjective experiences.

It is worth to note that Efklides' MASRL model distinguishes two levels of self-regulated learning: the person level/ macro level and the Task × Person level/micro level. The Task × Person level incorporates the contextual influence on learners' self-regulated learning. In an online learning environment, the automatically recorded learners' digital trace data reflect the Task × Person level/micro level. The knowledge domain, the interactions with the instructor, the interactions with peers, the cognition and motivation of the learner all impact the learner's self-regulated learning process. Although it is impossible to observe or record the data of learners' metacognition and motivation, learners' digital traces could offer at least a modest reflection of their self-regulated learning process from an authentic learning context. The learning patterns might vary from what the existing SRL models proposed because the context makes the SRL process more dynamic. Therefore, when looking at the digital trace data collected from LMS, it may be more helpful to jump out of the general top-down self-regulation to focus on the contextual bottom-up self-regulation.

The next section reviews existing SRL measurements and the degree to which they are appropriate to use in online learning environments in the digital era.

Self-regulated Learning Measurements

Winne and Perry (2000) believed that SRL measurements could be categorized into two broad types: as an aptitude and as an event. According to Winne and Perry (2000), an aptitude is described as “a relatively enduring attribute of a person that predicts future behavior” (p. 534). Aptitude data can be collected by a self-reported questionnaire, a structured interview, or teacher judgments. An event is “like a snapshot that freezes activity in motion, a transient state embedded in a larger, longer series of states unfolding over time” (p. 534). Events data can be collected by doing think-aloud activities, error detection tasks, trace methodologies, or observations. When judging from these two formats of SRL measurements, the aptitude data can be collected through questionnaires while event data can be compiled from learners’ digital traces collected by the learning management system.

It is helpful at this point to find out what kind of SRL measurements exist, so in the next section, key existing SRL measurements will be illustrated and compared.

The Learning and Study Strategies Inventory

In order to diagnostically assess students’ academic strengths, Weinstein, Palmer, and Schulte (1987) developed an instrument named The Learning and Study Strategies Inventory

(LASSI), which initially had 77 items in 10 subscales. The second version (Weinstein & Palmer, 2002) was an update from the first edition with 80 items by broadening the scope of the scales and incorporating current educational research findings. The third version of LASSI has been reduced to 60 items, which focuses on learners' awareness about their use of learning and study strategies related to skill, will, and self-regulation components of strategic learning (Weinstein, Palmer, & Acee, 2016). The self-regulation component of strategic learning includes four scales: concentration, self-testing, time management, and using academic resources. The skill components of strategic learning include three scales: information processing, selecting main ideas, and test strategies. The will components of strategic learning include three scales: anxiety, attitude, and motivation.

Zimmerman's SRLIS

Zimmerman and Martinez-Pons developed a Self-Regulated Learning Interview Scale (SRLIS) to solicit students' verbal responses to typical academic problems or contexts (Zimmerman & Martinez-Pons, 1986). It is a theory-guided, structured interview protocol. Data about SRL-related behaviors are elicited by having students consider a contextualized but fictitious task. The students' responses are categorized into 15 self-regulated learning strategies, including self-evaluation, organizing and transforming, goal-setting and planning, seeking information, keeping records and monitoring, environmental structuring, self-consequences, rehearsing and memorizing, seeking social assistance, reviewing records, and other. Zimmerman and Martinez-Pons (1986) asked the teachers to rate each of their students on a scale entitled Rating Student Self-Regulated Learning and found that the teachers' ratings were highly correlated with students' reports (R=.70).

The Motivational Strategies for Learning Questionnaire

In order to assess college students' motivational orientations and their use of different learning strategies for college courses, Pintrich, Smith, Garcia and McKeachie (1993) proposed a final version of the Motivational Strategies for Learning Questionnaire (MSLQ) based on 10 years of research involving many rounds of data collection. The MSLQ includes two sections: a motivation section and a learning strategies section. It has 81 items that are scored on a 7-point Likert scale, from 1 (not at all true of me) to 7 (very true of me). The motivation scale consists of 31 items to assess students' goals and value beliefs for a course, self-efficacy and control of learning beliefs, and their anxiety about tests in a course. The learning strategy section includes 50 questions: 31 items regarding students' use of different cognitive and metacognitive strategies and 19 items concerning student management of different learning resources.

The Online Self-regulated Learning Questionnaire

The Online Self-regulated Learning Questionnaire (OSLQ: Barnard, Paton, & Lan, 2008) is a 24-item scale with a 5-point Likert response format having values ranging from strongly agree (5) to strongly disagree (1). According to the authors, the OSLQ was developed based on the 1998 work of Zimmerman to reflect a multi-dimensional conception of self-regulated learning and has been examined for its internal consistency. Exploratory factor analyses were used to examine its structure validity. The OSLQ consists of six subscales including environment structuring, goal setting, time management, help-seeking, task strategies, and self-evaluation.

Self-Regulation in Self-Paced Open and Distance Learning Environments

Kocdar, Karadeniz, Bozkurt, and Buyuk (2018) proposed a 30-item scale to measure Self-Regulation in Self-Paced open and distance Learning Environments (SRSPLE) based on the literature review, expert opinions, and learner questionnaires. The initial scale had 62 items, which were generated by the analysis of previous studies and the responses obtained from a questionnaire that was sent to distance learners. The 62-item scale was presented to seven experts for evaluation and it was reduced to 55 items. Both the exploratory and confirmatory factor analyses were conducted to validate the scale. Five factors composed of goal setting, help-seeking, self-study strategies, managing the physical environment, and effort regulation emerged in the 30-item scale. This instrument includes items that imply whether learners seek help from the Internet and social media, friends, as well as subject experts, which reflect the characteristics of online learning.

Comparing these SRL Measurement Instruments

Table 3.2 shows a summary of these SRL measurement instruments. The LASSI measurement has been widely used, however, it has been commercialized and the questionnaire is not available to the public. Zimmerman's SRLIS is more of an interview protocol instrument. MSLQ is the most widely used self-report questionnaire for self-regulated learning and a manual of how to use this measurement is available. However, it is very long and includes some unnecessary or redundant items. Of the two instruments developed for online learning environments, the OSLQ seems to have the most empirical support.

All these SRL measurement instruments are focused on two key components: (1) learners' motivation orientation, which includes learners' self-beliefs of learning and learners' intrinsic and extrinsic goals (goals are the driver of the motivation); and (2) learners' cognitive strategies, which includes time management/effort regulation, study environment management, and help-seeking.

Table 3.2.

Self-regulated Learning Measurement Instruments Comparison

Instrument	Key components	Item #
LASSI	Self-regulation: concentration, self-testing, time management, and using academic resources; skill: information processing, selecting main ideas, and test strategies; will: anxiety, attitude, and motivation.	60
SRLIS	Self-evaluation, organizing and transforming, goal-setting and planning, seeking information, keeping records and monitoring, environmental structuring, self-consequences, rehearsing and memorizing, seeking social assistance, reviewing records, and other.	NA
MSLQ	**Motivation orientation**: intrinsic goal orientation, extrinsic goal orientation, task value, control beliefs about learning, self-efficacy for learning and performance, and test anxiety; **cognitive strategy**: rehearsal, elaboration, organization, critical thinking, metacognitive self-regulation, time and study environment management, effort regulation, peer learning, and help-seeking.	81
OSLQ	Environment structuring; goal setting; time management; help-seeking; task strategies; and self-evaluation.	24
SRSPLE	Goal setting, help-seeking, self-study strategies, managing physical environment, and effort regulation.	30

All these SRL measurement instruments heavily rely on students' self-reported data, which has been criticized as lacking validity (Cho & Summers, 2012; Zimmerman, 2008; Winnie, 2010). For example, besides the possible biased information arising from incomplete and reconstructed memories, Winne (2010) pointed out two shortcomings of self-reported SRL measurements: (1) they are weak in representing variance in learners' intentions and experiences; and (2) they only can describe an adaptive behavior as a homogeneous, static state with no accounting for context. Given the latest developments of eLearning, current learning management systems can automatically record students' digital traces. Some researchers refer to these digital traces as learners' behavioral data in their online courses. Students' digital trace data are passive (i.e. not intentionally measuring any specific domain), which could reflect their true self-regulation ability. However, data do not equate to useful information. With the development of the new extensive educational media and advances in computation, learning analytics is at the cusp of becoming a new educational research method using large data sets to understand learning and the context (learning environment) in which learning occurs. Learning analytics could be used as a method to supplement traditional self-reported SRL data and help us advance our understanding of SRL. In the next section, a theoretical framework is proposed to illustrate how the learners' digital trace data collected from LMSs could be interpreted from a self-regulated learning theory perspective.

A New Theoretical Framework for SRL in Online Learning Environments

Figure 3.4 shows a new theoretical framework for SRL in an online learning environment in the digital era. It is an interactive system between learners and their online learning environments. It begins when a learner logs into the learning management system (LMS) and starts to explore the system and access the learning materials. The LMS provides responses to the learner although these responses are mainly just the learner's performance scores as the LMS is not an artificially intelligent assisted system. The instructor then communicates with the learner and provides feedback to him or her. The learner will also communicate and interact with his or her classmates. The interaction between the learner and the instructor, the learner and the content, the learner and his or her classmates are mutual. Efklide's MASRL is reflected here: the top-down is at the learner's personal level, which includes the learner's intrinsic and extrinsic goals, self-belief of learning, the knowledge of the task, and cognitive strategies. The bottom-up self-regulation happens when the learner interacts with the content, the instructor, and other learners. Top-down self-regulation and bottom-up self-regulation are "two sides of the coin." The learner's behavioral data are collected by the LMS mainly during the bottom-up self-regulation process, which happens when the learner is interacting with the learning materials. The bottom-up self-regulation should also include the interactions between the learner and the instructor as well as the interactions between the learner and other learners. Boekaerts' dual processing model – whether the learner chooses the growth pathway or well-being pathway – intends to focus on the end results, so could be considered as the result of the SRL process. Therefore, the digital traces collected by the learning management system could reflect the learner's self-regulation process, but it reflects the process dynamically because it involves the learner's interactions with the instructor and other learners while completing the learning task. If the self-reported SRL measurements are assessing the learner's aptitude or perception and are stable and static, then learners' trace data are more of a kind of checking learners' self-regulation ability in a contextualized format that fluctuates and are dynamic.

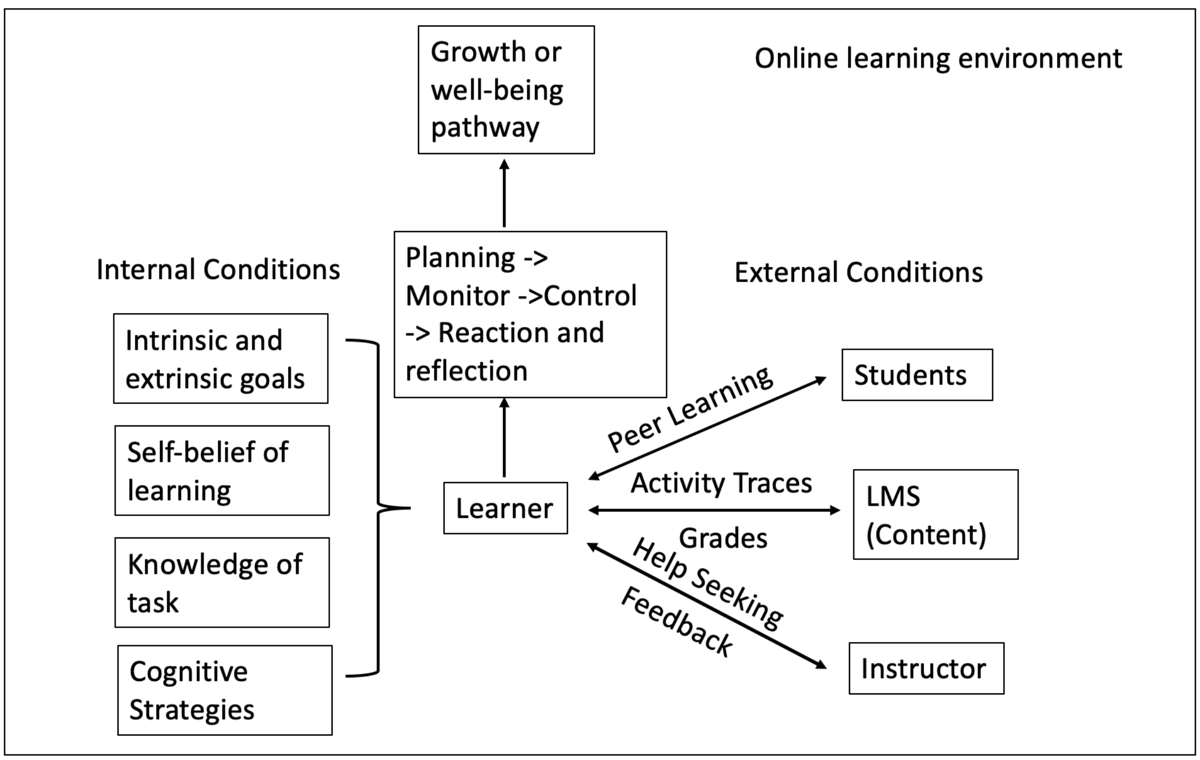


Figure 3.4. The Theoretical Framework for Self-regulated Learning in Online Learning Environments

Conclusion

In summary, based on a comprehensive literature review on existing self-regulated learning models and measurements, key components of SRL have been identified. Finally, a theoretical framework explaining the relationship between the learner's digital traces collected by LSMs and the key components of SRL has been proposed. Learners' top-down self-regulation is driven by the learner's goals, internal beliefs, cognitive strategies, and knowledge of the task, while bottom-up self-regulation is reflected by the learner's interactions with the content, the instructor, and his or her classmates. The digital trace data collected by LSMs can reflect the learner's self-regulation dynamically and contextually, which may fluctuate somewhat based on the learning environment and the learner's interactions with the instructor as well as other learners.

References

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Barnard, L., Paton, V., & Lan, W. (2008). Online self-regulatory learning behaviors as a mediator in the relationship between online course perceptions with achievement. *The International Review of Research in Open and Distance Learning, 9*(2), 1–11.

Barnard-Brak, L., Paton, V. O., & Lan, W. Y. (2010). Profiles in Self-Regulated Learning in the Online Learning Environment. *International Review of Research in Open and Distance Learning*, *11*(1), 61–80. https://doi.org/10.19173/irrodl.v11i1.769

Boekaerts, M. (1995). Self-regulated learning: Bridging the gap between metacognitive and metamotivation theories. *Educational Psychologist, 30*(4), 195-200.

Boekaerts, M. (1996). Coping with stress in childhood and adolescence. In Zeidner, M. & Endler, N. S. (Ed.), *Handbook of coping: Theories, research, application* (pp. 452 -484). Oxford: John Wiley & Sons.

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction, 7*(2), 161–186.

Boekaerts, M. (2011). Emotions, emotion regulation, and self-regulation of learning. In Zimmerman, B. J. & Schunk, D. H. (Ed.), *Handbook of self-regulation of learning and performance* (pp. 408-425). New York, NY: Routledge.

Boekaerts, M., Pintrich, P. R., & Zeidner, M. (2000). *Handbook of self-regulation*. San Diego: Academic Press.

Boekaerts, M. (2002). Bringing about change in the classroom: strengths and weaknesses of the self-regulated learning approach — EARLI Presidential Address, *Learning and Instruction, 12*, 589–604.

Cho, M., Kim, Y., & Choi, D. (2017). The effect of self-regulated learning on college students' perceptions of community of inquiry and affective outcomes in online learning. *The Internet and Higher Education, 34*, 10-17.

Cho, M. & Yoo, J. S. (2017). Exploring online students' self-regulated learning with self-reported surveys and log files: a data mining approach. *Interactive Learning Environments, 25*(8), 970-982.

Colthorpe, K., Zimbardi, K., Ainscough, L., & Anderson, S. (2015). Know thy student! Combining learning analytics and critical reflections to increase understanding of students' self-regulated learning in an authentic setting. *Journal of Learning Analytics, 2*(1), 134-155.

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

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *The International Journal of Technology Enhanced Learning, 4*(5/6), 304-317.

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

Gandomkar, R., & Sandars, J. (2018). Clearing the confusion about self-directed learning and self-regulated learning. *Medical Teacher, 40*(8), 862–863.

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

Gelan, A., Fastre, G., Verjans, M., Martin, N., Janssenswillen, G., …Thomas, M. (2018). Affordances and limitations of learning analytics for computer-assisted language learning: a case study of the VITAL project. *Computer Assisted Language Learning, 31*(3), 294-319.

Graham, S., Harris, K. R. & Troia, G. A. (1998). Writing and self-regulation: Cases from the self-regulated strategy development model, In Schunk, D. H. & Zimmerman, B. J. (Ed.), S*elf-regulated Learning from Teaching to Self reflective Practice* (pp. 20-41). New York: The Guildford Press.

Jarvela, S., & Hadwin, A. F. (2013). New Frontiers: Regulating Learning in CSCL. *Educational Psychologist*, *48*(1), 25-39.

Hoppe, H. U. (2017). Computational methods for the analysis of learning and knowledge building communities. In Lang, C., Siemens, G., & Wise, A., & Gasevic, D. (Ed.), *Handbook of Learning Analytics* (pp.23-33). SOLAR.

Knowles, M. S. (1975). *Self-directed learning: A guide for learners and teachers*. New York: Association press.

Kocdar, S., Karadeniz, A., Bozkurt, A., & Buyuk, K. (2018). Measuring self-regulation in self-paced open and distance learning environments. *International Review of Research in Open and Distributed Learning*, *19*(1), 25-42.

Lawanto, O., Santoso, H. B., Lawanto, K. N., & Goodridge, W. (2014). Self-regulated learning skills and online activities between higher and lower performers on a web-intensive undergraduate engineering course. *Journal of Educators Online, 11*(3).

Li, H., Flanagan, B., Konomi, S., & Ogata, H. (2018). Measuring behaviors and identifying indicators of self-regulation in computer-assisted language learning courses. *Research and Practice in Technology Enhanced Learning, 1*, 13-19.

Loyens, S. M. M., Magda, J., & Rikers, R. M. J. P. (2008). Self-directed learning and problem-based learning and its relationships with self-regulated learning. *Educational Psychology Reviews*, *20*(4), 411-427.

Pardo, A., Han, F., & Ellis, R. A. (2017). Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. *IEEE Transactions on Learning Technologies, 10*(1).

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In Boekaerts, M., Pintrich, P. R., & Zeidner, M. (Ed.), *Handbook of self-regulation* (pp. 452-502). San Diego, CA: Academic Press.

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

Pintrich, P. R., Wolters, C., & Baxter, G. (2000). Assessing metacognition and self-regulated learning. In G. Schraw (Ed.), *Metacognitive assessment*. Lincoln: University of Nebraska Press.

Saks, K. & Leijen A. (2014). Distinguishing self-directed and self-regulated learning and measuring them in the e-learning context. *Procedia – Social and Behavioral Sciences*, *112*, 190-198.

Weinstein, C. E., & Palmer, D. R. (2002). *Learning and Study Strategies Inventory (LASSI): User's manual (2nd ed.)*. Clearwater, FL: H & H Publishing.

Weinstein, C. E., Palmer, D. R., & Acee, T. W. (2016). *Learning and Study Strategies Inventory (LASSI): User's manual (3rd ed.)*. Clearwater, FL: H & H Publishing.

Weinstein, C. E., Palmer, D., & Schulte, A. C. (1987). *Learning and Study Strategies Inventory (LASSI)*. Clearwater, FL: H & H Publishing.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.), *Metacognition in educational theory and practice* (pp. 279 – 306). Routledge.

Winne, P. H. & Perry N. E. (2000). Measuring self-regulated learning. In Boekaerts, M., Pintrich, P. R., & Zeidner, M. (Ed.), *Handbook of self-regulation* (pp. 531-566). San Diego: Academic Press.

You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *The Internet and Higher Education, 29*, 23–30.

Zhu, Y., Zhang, J. H., Au, W., & Yates, G. (2020). University students' online learning attitudes and continuous intention to undertake online courses: a self-regulated learning perspective. *Educational Technology Research and Development*, *68*(3), 1485–1519. https://doi.org/10.1007/s11423-020-09753-w

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology, 81*(3), 329-339.

Zimmerman, B. J. (1998). Developing self-fulfilling cycles of academic regulation: An analysis of exemplary instructional models. In D. H. Schunk & B. J. Zimmerman (Ed.), *Self-regulated learning: From teaching to self-reflective practice* (pp. 1-19). New York: Guilford.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In Boekaerts, M., Pintrich, P. R., & Zeidner, M. (Ed.), *Handbook of self-regulation* (pp. 531-566). San Diego: Academic Press.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166-183.

Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational Psychologist, 48*(3), 135-147.

Zimmerman, B. J., & Campillo, M. (2003). Motivating self-regulation problem solvers. In Davidson, J. E. & Sternberg, R. J. (Ed.), *The psychology of problem solving* (pp. 233–262). New York, NY: Cambridge University Press.

Zimmerman, B. J., & Martinez-Pons, M. (1986). Development of a Structured Interview for Assessing Student Use of Self-Regulated Learning Strategies. *American Educational Research Journal, 23*(4), 614-628.

CHAPTER 4

USING TRACE DATA TO ENHANCE STUDENTS' SELF-REGULATION: A LEARNING ANALYTICS PERSPECTIVE[3]

[3] Ye, D. To be submitted to *Educational Technology Research and Development*.

Abstract

The purpose of this study was to explore possible ways to enhance online students' self-regulated learning ability through learning analytics methods. This study took place in an upper-level college agriculture course delivered in an asynchronous online format. The majority of the participants were undergraduate students. By comparing online students' digital trace data with their self-reported data, this study found that digital trace data from LMS could predict students' performance more accurately than self-reported SRL data. Through cluster analysis, students were classified into three levels based on their self-regulatory ability: low, moderate, and high. The characteristics of each group were analyzed. By incorporating qualitative interview data, we found that lower self-regulated learners tend to not be fully aware of their weaknesses and are reluctant to make any changes. The implications of this study are discussed.

Keywords: self-regulated learning, online learning environments, digital trace data, self-reported SRL data, cluster analysis

Introduction

Online education has been growing tremendously in the past decade (Hart, 2012; Van Rooij & Zirkle, 2016), and it has been playing a dominant role in education during the coronavirus pandemic. Despite the popularity of online education, not all students are equally successful in asynchronous online courses. The situation has been even worse during the coronavirus pandemic because most students have had no choice but to take their courses online. Dray, Lowenthal, Miszkiewicz, Ruiz-Primo, and Marczynski (2011) indicated that students' engagement with technology and the personal traits of self-direction and initiative are significant predictors of online learners' success. Recent studies also demonstrate that in order for online learners to succeed in online courses, online learners must have the capacity to regulate their learning (Hew & Cheung, 2014; Kizicec & Schneider, 2015) or receive active self-regulation support from the learning platform (Kizilcec & Cohen, 2017). With the continuous growth of online courses and online programs offered by higher education, it is important to understand online students' self-regulated learning (SRL) processes so that we can implement strategies to enhance students' self-regulation abilities and thus improve their academic performance in online learning environments.

Although numerous studies about SRL have been conducted in online learning environments, existing SRL research has heavily relied on self-reported surveys (Winne & Perry, 2000). Self-reported data from students have been criticized as lacking validity (Cho & Summers, 2012; Zimmerman, 2008; Winne, 2010). One possible solution to address this issue is to use learners' trace data collected by learning management systems as a supplement to the self-reported SRL data. Traces are defined as "observable indicators about cognition that students create as they engage with a task" (Winne & Perry, 2000, p. 551). Recent studies (Hwu, 2003; Yu & Zhao, 2015) have indicated that online students' behavioral data are more accurate because the data collected from modern tracking technologies occur in actual learning situations in real-time. Learners may be aware of the data collection taking place, but it is relatively unobtrusive and difficult for learners to alter, so one can assert that more authentic learning behaviors can be recorded on a large scale using this approach. Winne and Perry (2000) proposed two different conceptualizations of SRL: as an aptitude and as an event. Winne (2010) believed that self-reported SRL should be considered as an aptitude and the trace data could be treated as an event. Trace data becomes the raw material for researchers to track aptitudes "in action" and how aptitudes may evolve as students make progress in their study.

The purpose of this study is to investigate whether students' self-reported SRL aligns with their behavior as indicated by the digital trace data collected by the learning management system. This study will also explore possible ways of using trace data to enhance the course design and improve learners' self-regulatory ability in online learning environments. The research questions of this study are as follows:

(1) How do the digital trace data collected by the learning management system reflect the students' self-reported SRL?

(2) What is the relationship between students' performance and the digital trace data and self-reported SRL data?

(3) What are the patterns of learning behaviors based on the digital trace data and self-reported SRL data?

(4) What are the explanations for any differences between their self-reported SRL data and the digital trace data?

Theoretical Framework

Several existing SRL models have been examined and compared, including Zimmerman's cyclical phases model (Zimmerman, 2000), Pintrich's SRL model (Pintrich, 2000), Boekaerts' dual processing model (Boekaerts, 2011), the Conditions, Operations, Products, Evaluations, and Standards (COPES) model (Winne & Hadwin, 1998), Efklide's Metacognitive and Affective Model of SRL (MASRL) (Efklides, 2011), and Jarvela and Hadwin's SRL model (Jarvela & Hadwin, 2013). The development of SRL models shows a trend of shifting from general SRL process models to more context-based models. Figure 4.1 shows a theoretical framework for SRL in online learning environments, which is proposed based on existing SRL models while being more context focused. It is an interactive system between the learner and the online learning environment. Each learner has his or her own learning goals and beliefs. He or she also has some knowledge of the task and his or her own cognitive strategies. When a learner logs into the learning management system (LMS), he or she starts the interactions with the content, the instructor, and other students. The interactions between the other learners, the content, and the instructor are mutual although here only important information related to the study is illustrated in the framework. Efklide's MASRL is reflected here: the top-down self-regulation is at the learner's personal level, which includes the learner's goals, internal beliefs, knowledge of the task, and cognitive strategies; the bottom-up self-regulation happens when the learner interacts with the content, the instructor, and other learners. Top-down self-regulation and bottom-up self-regulation are two aspects of the same behavior. The general SRL process is the results of both the top-down and bottom-up self-regulation, which leads to the results of a growth or well-being pathway. This theoretical framework indicates that the trace data collected by the LMS can reflect students' SRL dynamically and contextually. Given the volume of data that learning management systems collect, it can be difficult to decide which learning behavior variables to study. A review of the existing literature provides some guidance.

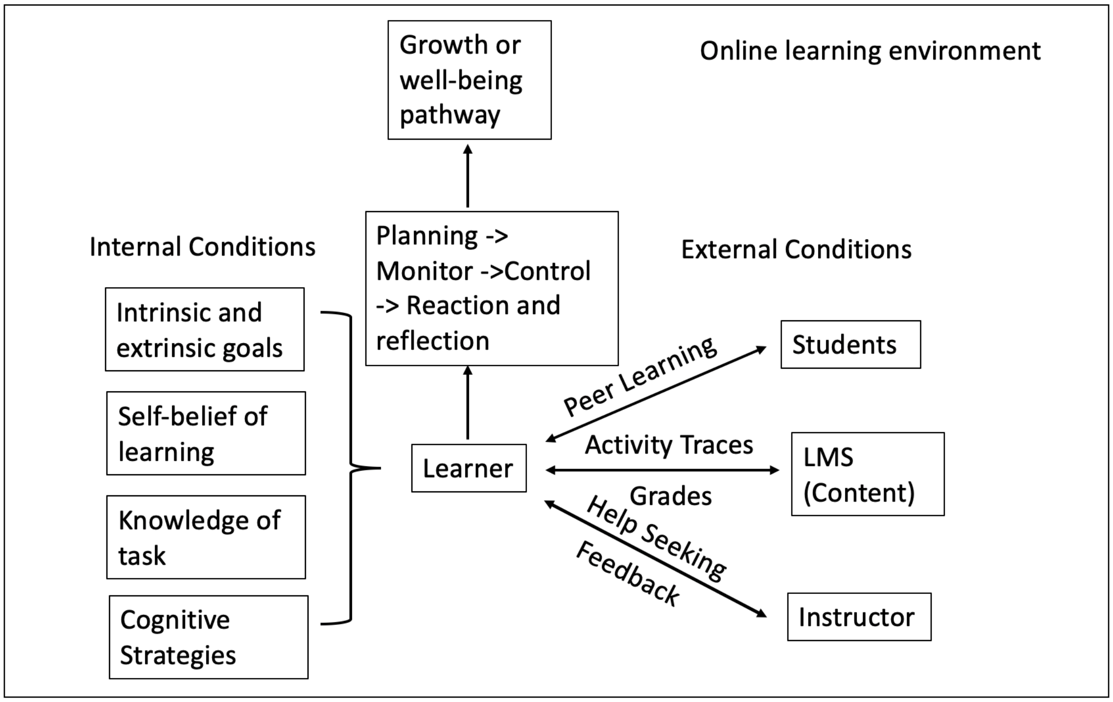


Figure 4.1. The Theoretical Framework for Self-regulated Learning in Online Learning Environments

Research on SRL-based Learning Analytics

This section reviews the literature of research on SRL-based learning analytics. This review focuses on key learning variables and analysis methods used in SRL research.

Literature Searching Methods

The database searching method was used to locate existing research of SRL using LMS data. After searching the words "self-regulated learning," "online," "learning analytics," and "data" in two databases (Education Research Complete and ERIC) with published dates from 2011 to 2021, 21 scholarly (peer-reviewed) articles were found and reviewed. The criteria used to select these articles were: (1) the focus is SRL, and (2) data from LMS are used and analyzed in an online learning environment.

Learning Variables/Measurements

It is neither efficient nor necessary to analyze all the digital trace data gathered within LMS. Therefore, it is important to distinguish the important learning behavior variables from all the other variables by only focusing on the important learning behavior variables. To achieve this goal, a literature review of related studies was conducted. A summary of learning behavioral variables used in these studies is shown in Table 4.1. The most commonly used learning behavior variables are study regularity, procrastination, time investment, completion, and help-seeking. The definition of these terms in different studies varies slightly. Study regularity generally refers to the frequency of the student accessing various learning materials or LMS login frequency. Procrastination usually measures whether students submit the assignments on time or whether they study the learning materials on time. Time investment often refers to the total time students spend studying the course or completing a task. Completion usually measures the portion of the course or the task students complete. Help-seeking generally measures how many times students reach out to other learners or the instructor for help.

Table 4.1.

Learning Behavior Variables Used in Existing Studies

Variables	Data Collection	Studies
Study regularity	Computed using the standard deviation of login (or access) intervals.	Li, Flanagan, Konomi, and Ogata (2018); Kim, Yoon, Jo, and Branch (2018).
	Measured by the virtual attendance score.	You (2016).

	Spacing – the standard deviation of the studying in advance of the unit and the unit quiz deadline.	Li, Baker, and Warschauer (2020)
	The frequency of the student accessing various learning materials.	Lawanto, Santoso, Lawanto, and Goodridge (2014).
	LMS login in interval regularity and online lecture access interval regularity.	Kim, Yoon, Jo, and Branch (2018).
Procrastination	Late submission indicates the student's failure to submit assignments on time.	You (2016); Colthorpe, Zimbardi, Ainscough, and Anderson (2015).
	The number of on-time, late, and early submitted course assignments.	Lawanto, Santoso, Lawanto, and Goodridge (2014).
	Whether the students completed the quiz units in advance and how early the students completed the quiz units.	Li, Flanagan, Konomi, and Ogata (2018).
	Studying on time: the proportion of units accessed before the deadline.	Li, Baker, and Warschauer (2020)
	Studying in advance – students' first time to visit a unit page in advance of the unit quiz deadline.	Li, Baker, and Warschauer (2020)
Time investment	Time spent on online lectures and online lecture access frequency.	Kim, Yoon, Jo, and Branch (2018).
	The total time spent on the course.	Gelan et al. (2018).
	Time on task	Li, Baker, and Warschauer (2020)
	Number of times students access the course	You (2016); Lawanto, Santoso, Lawanto, and Goodridge (2014); Gelan et al. (2018).
Completion	Whether a student downloads and reads the course information packets (important course information).	You (2016).
	Whether the students completed a quiz unit within the scheduled stage.	Li, Flanagan, Konomi, and Ogata (2018).
	The number of completed quizzes.	Li, Flanagan, Konomi, and Ogata (2018).

Help-seeking	Time spent on the Q & A Board, Q & A board visit frequency, and number of messages posted on the Q & A board.	Kim, Yoon, Jo, and Branch (2018).

Key Findings

Key findings from existing SRL-based learning analytics research in higher education are discussed below.

Digital trace data from LMS are more powerful than self-reported SRL in predicting academic performance (Cho & Yoo, 2017; Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007). Cho and Yoo (2017) used a classification model to predict students' final grades based on 17 log attributes, five selected log attributes, and five selected log attributes as well as four self-reported SRL attributes respectively and found that the accuracy of correctly classified instances (58.33%) of the prediction model from log attributes was higher than that of the prediction model from self-reported SRL (41.67%). Using regression analysis, Pardo, Han, and Ellis (2017) found that the variation of the students' final scores for their course is better explained when combining self-reported SRL data with data of seven student engagement events based on the digital footprints left in the learning management system.

By using the regression or decision-tree method, the following learning behavioral variables explain the most variance of students' academic performance: the frequency of course login (You, 2016; Lawanto, Santoso, Lawanto, & Goodridge, 2014; Gelan et al., 2018), study regularity (You, 2016; Kim, Yoon, Jo, & Branch, 2018; Li, Flanagan, Konomi, & Ogata, 2018), procrastination (You, 2016; Li, Flanagan, Konomi, & Ogata, 2018; Lawanto, Santoso, Lawanto, & Goodridge, 2014; Colthorpe, Zimbardi, Ainscough, & Anderson, 2015), time investment (Gelan et al., 2018; Cho &Yoo, 2017), and completion (You, 2016; Cho &Yoo , 2017; Flanagan, Konomi, & Ogata, 2018).

However, most of these studies have not provided a solid pedagogy or theory support for why these learning behavioral variables are chosen. The researchers generally chose these learning behavioral variables based on what existing studies used.

Learning Behavior Variables Proposed in This Study

Although learning analytics appears very promising, existing research has demonstrated several limitations. Several researchers have argued that data should be interpreted from the learners' perspective (Ferguson, 2012) and that the results should provide actionable recommendations (Gasevic, Dawson, & Siemens, 2015). One key reason why existing learning analytic research gets these criticisms is that we do not have a solid pedagogy or theory support for data collection. Based on a review of the literature, the majority of related studies have not provided a solid pedagogy or theoretical support for why these learning behavioral variables are chosen. Most of the existing research studies use the same or similar learning behavior variables, such as frequency and procrastination. Several studies pointed out that it is important to align the data collection with SRL model or theory (Yu & Zhao, 2015; Siadaty, Gasevic, & Hatala, 2016). Therefore, a theory-based data collection is proposed in Table 4.2. Using Pintrich's (2000) cyclical SRL model of (1) forethought, planning, and activation, (2) monitoring, (3) control, and (4) reaction and reflection, I developed a matching map between the SRL phases and the corresponding learning behavioral data in LMS. In a recent pilot study, I found that some students chose to download some learning materials (e.g., course syllabus, lectures, PDF file) and read them locally. This implies that the trace data may not reflect the actual time they spent on studying. For this reason, the study monitored whether the learner downloads any files, and if so, these data are collected and included in the analysis.

Table 4.2.

A Matching Map between the SRL Phases, Learning Behavior Data in LMS, and OSLQ

Pintrich's SRL Model Phase	Learning Behavioral Data in LMS	OSLQ
Forethought, planning, and activation	Time spent on reviewing the syllabus (if it has been downloaded) Time spent on reviewing the rubrics Syllabus visit frequency Rubrics visit frequency	Goal setting
Monitoring	Weekly course logins Weekly page visits Number of access to optional resources	Task strategies
Control	Time spent on the course weekly Total time spent on the course Number of threads created in online discussions Number of replies created in online discussions Weekly lecture viewing time Number of late submissions Studying in advance Unit completion portion before the deadline	Time management
Reaction and Reflection	Number of threads created in Q & A forum Number of replies created in Q & A forum Number of emails to the instructor asking for help	Help-seeking
	Time on revisiting a unit after the unit grades published Completion rate of optional self-assessment	Self-evaluation

Methods

Existing research mainly focuses on quantitative data, such as the clickstream data or the interaction between the learner and the content. This study includes the interaction data between the learner and the content, the learner and other learners, and the learner and the instructor. In addition, both quantitative and qualitative research methods were used in this study because both types of data can support and supplement each other while also serving different purposes. Maxwell (2013) pointed out that qualitative and quantitative methods are not simply different ways of doing the same thing, but are best used to address different kinds of questions and goals. In this study, quantitative data were used to address the "what" questions, such as "what learning behavior patterns can we identify?" In contrast, qualitative data were used to address the "why" question. In this study, the question is, "why there are differences between students' self-reported SRL data and the digital trace data?" Although data gathered within the learning management system are authentic and provide critical information about learners' learning behaviors, the behavioral data are just part of a more complex learning process. As Greller and Drachsler (2012) pointed out, "To judge a learner's performance merely on, e.g., LMS quantitative data, is like looking at a single puzzle piece" (p. 52). Pardo, Ellis, and Calvo (2015) is a good example of using a mixed research method. In their study, they explored the benefits of designing activities for a learning environment using quantitative data in the form of the digital footprint data of students as well as qualitative data examining the student approaches to learning framework. These authors argued that the meaning of learning analytics is improved when combining quantitative data with qualitative data. Therefore, this study combined self-reported SRL measurement results with LMS behavioral data and qualitative interview data in the hope that they can supplement each other and provide a more complete picture.

Research Context

The data were collected from an upper-level, undergraduate/graduate split level science course, cross-listed with horticulture and crop and soil sciences. The course was taught in an asynchronous, online format. The content of the course is core to both disciplines as well as required for several graduate programs. However, approximately half of the students who registered for the course major in other disciplines and take the course as an upper-level science elective. The course is offered in modules. In each module, there are readings, lectures, additional resources, some optional self-assessments, a quiz, and an online discussion or assignment. The final grades are calculated based on 10 quizzes (35.3% of the final grade), four practical lab written assignments (33% of the final grade), six online discussions (28.2% of the final grade), and one applied concept activity (3.5% of the final grade). The quizzes cover basic knowledge students need to master, while the practical lab written assignments, online discussions, and applied concept activity require students to observe, do some hands-on experiments, record data, and explain their understanding by applying the concepts they learned in the course. Based on the structure and components of the grading policy, final grades measure both students' knowledge understanding and application and cover both low and high levels of cognitive learning, which can well represent students' learning in this course.

Participants. Participants were students who registered for this course. They came from various majors and included both undergraduate and graduate students although the majority were undergraduate students. A total of 91 students were enrolled in the course. Of these, 67 students agreed to participate in the study, but one student withdrew from this class in the middle of the semester and another student did not complete the SRL post-course survey. Among these 65 participants, there were 20 male students and 45 female students. A total of 18 participants were graduate students and the remaining 47 were undergraduate students. Among these 65 participants, 29 students indicated they were willing to participate in the interview at the beginning of the course. By the end of the semester, 20 students were randomly selected among these 29 and invited to participate in the interview. A total of 11 eventually participated in the interview.

Data Collection

Instruments. The Online Self-Regulated Learning Questionnaire (OSLQ: Barnard, Paton, & Lan, 2008) was used to collect students' self-reported self-regulation data. It consists of 24 items using a 5-point Likert response format ranging from strongly agree (5) to strongly disagree (1). Although the use of Motivated Strategies for Learning Questionnaire has a long history and is the most widely used self-report questionnaire for self-regulated learning, it has a large number of items, making its use to collect data impractical in an online course. In contrast, the OSLQ is short and concise, which led to the assumption that it would yield higher completion rates and therefore more data. In addition, it is an SRL questionnaire specifically designed for online learning environments, which fits well with the context of the study. Most importantly, the available psychometric data shows it to be a reliable and valid way to assess the self-regulatory learning skills of students in both blended and online courses (Barnard, Lan, To, Paton, & Lai, 2009).

A semi-structured interview was conducted to focus on students' typical self-regulated learning behaviors while taking this course. Specifically, participants were asked to recall "what they typically did while taking this course" based on the six scales of self-regulated learning: goal setting, environmental structuring, task strategies, time management, help-seeking, and self-evaluation. An interview protocol (see appendix B) was used during the interview. The purpose of the interview was to explore possible contextual explanations for the differences between the trace data and the self-reported data by using qualitative data.

Data Collection Process. Three types of data were collected in this study in the hope that the data would complement each other and help overcome the inherent limitations of each type of data. Students were invited to participate in the study voluntarily in the first week of the course through a course announcement. Once the participants agreed, they were asked to complete the online self-regulated learning questionnaire (OSLQ) in the course. During the semester, the learning management system recorded learners' behavior data automatically. Researchers collected data recorded by the learning management system based on modules. Most modules were offered weekly, but a few modules were offered every two weeks. Li, Baker, and Warschauer (2020) argue that compared to the self-report questionnaires after the course, it is more problematic to use self-report questionnaires before a course starts because it requires students to predict their SRL behaviors later on. Similarly, Winne (2010) also believed that researchers should not presume aptitudes self-reported before an intervention to be constant throughout the intervention and instead should gather data during the interventions to document whether aptitudes change. Therefore, in this study, students were asked to complete the OSLQ again in the last week of the course. By the end of the course, learners' completion data of each individual item, interaction data in online discussions, interactions with the instructor, login history, time spent on reviewing content, final grades, etc. were recorded in an Excel file and used for data analysis. The participants who agreed to be interviewed were contacted and interviewed via Zoom meetings. Eleven students participated in a semi-structured interview. The length of the interviews ranged from 30 to 50 minutes. The interview focused on recalling the most typical situations students experienced while taking this course from the following six perspectives: goal setting, environment structuring, task strategies, time management, help-seeking, and self-evaluation.

Data Analysis

Table 4.3 shows an overview of the data analysis methods used in this study based on the research questions. Both quantitative and qualitative data analysis methods were used.

Table 4.3.

An Overview of the Data Collection and Analysis Methods

Research Questions	Data collection methods	Analysis Methods
	1.1 Record data from LMS	Correlation

RQ1: How do the trace data reflect the self-reported SRL data?	1.2 OSLQ Pre- and post-course	
RQ2: What's the relationship between performance and the trace data and self-reported SRL data?	1.1 Record data from LMS 1.2 OSLQ Pre- and post-course	Regression analysis Feature selection
RQ3: What learning patterns are present based on the trace data?	1.1 Recorded data from LMS 1.2 OSLQ Pre- and post-course	Cluster analysis
RQ4: What are the explanations for any differences between their self-reported SRL data and the digital trace data?	1.1 Recorded data from LMS 1.2 OSLQ Pre- and post-course 1.3 Interview	Correlation Thematic Analysis

Quantitative Data Analysis. Trace data were organized in an Excel file, and the average values were calculated for each learning variable. Based on Table 4.2, students' learning behavior data recorded by LMS were collected according to the module schedule. This course had 11 modules scheduled, but the last module was specifically designed for graduate students. Consequently, the undergraduate students were not required to complete module eleven. Module eleven was also available to students at the beginning of the course and had the same due dates as module ten for all its assignments. When analyzing the data, the majority of learning behavior data were based on an average of ten modules, but there were a few variables such as average topics visited per module, average time accessing the eLC course per module, and average items visited per module which included the module eleven data, and the average values were calculated based on 11 modules if the student was a graduate student. For the self-reported SRL data, the average rating of each scale was calculated. So for both the trace data and self-reported SRL data, the average values were used in the analysis.

Pearson correlations were calculated to see how the trace data reflected students' self-reported SRL data. Pearson correlations were also used for comparing the pre-course self-reported SRL data with the post-course self-reported SRL data. In addition, Linear regression analysis and feature selection methods were used to construct prediction models and identify identify key attributes from self-reported SRL data and the digital trace data. Three methods were used to identify key features: stepwise, forward, and backward.

Cluster analysis was used to analyze both the trace data and the self-reported SRL data. Cluster analysis is a class of techniques that are used to classify a set of data objects based on similarity and dissimilarity. In clustering analysis, there is no prior information about the group or cluster membership for any of the objects. In this study, clustering analysis, specifically the K-means clustering method, was used to group students based on relatively homogeneous learning behavioral patterns. The K-means clustering method is a classification method to group points by computing the distance between points and group centers. All the values of these variables were standardized by converting them into z-scores, and then the k-mean clustering analysis was conducted using the SPSS software. After the completion of the cluster analysis, the characteristics of each group of learners were summarized based on the data. Finally, the results were compared with that of the self-reported SRL data. Similarity and differences were identified and analyzed.

Qualitative Data Analysis. The interview data were transcribed and analyzed using the thematic analysis method. Braun and Clarke (2006) definesd thematic analysis as "a method for identifying, analyzing and reporting patterns (themes) within data" (p.79). The thematic analysis method proposed by Braun and Clarke (2006) was used to analyze the data, which includes six phases: (1) familiarizing yourself with your data; (2) generating initial codes; (3) searching for themes; (4) reviewing themes; (5) defining and naming themes; (6) producing the report. The thematic analysis approach was used to code the topics of the interview data and categorize the codes into themes. The interview recordings were first transcribed. After that, the transcribed data were read repeatedly while coding each topic. After that, these codes were categorized into themes. The audio recording was also listened to as needed to pick up nuances in meaning based on the participants' voice inflections and tones. Matrices strategy was used to list all these topics in a summary format for each participant in an Excel file. Then, after a reading across all these participants, common themes were found and marked. Finally, these themes were reported. While analyzing the common themes, an effort was made to align them with these six scales of self-regulated learning: goal setting, environment constructing, task strategies, time management, help-seeking, and self-evaluation.

Validity and Reliability

Different approaches were used to ensure the validity and reliability of both quantitative and qualitative research methods used in this study. The Online Self-Regulated Learning Questionnaire (OSLQ) was developed from an 86-item pool and reduced to 24 items after examining their internal consistency and exploratory factor analysis results based on data collected. Barnard, Lan, To, Paton, and Lai (2009) have conducted two studies to examine the instrument's reliability and validity across students' experience of a blended course as well as an online course. Both studies demonstrated adequate internal consistency and sufficient score reliability. Confirmatory factor analyses were performed with supportive evidence towards the construct validity of the instrument with respect to students enrolled in a blended course as well as in an online course. Therefore, the instrument has been proved to be reliable and valid to assess the self-regulatory learning skills of students in both blended and online courses.

The behavioral data automatically recorded from the LMS were reviewed carefully to make sure they were accurately recorded. For the clustering analysis, both cluster cohesion and separation values were checked to validate clusters. For the qualitative interview, the same questions were consistently asked to all interviewees, except some additional questions were added after the first interview in order to make sure they closely aligned with trace data. The process was recorded in detail, and to enhance the validity of the qualitative research process, all memos, drafts, notes, and analytic notes have been maintained in an organized file and are available upon request for review.

Results

RQ1: How Do the Trace Data Reflect the Self-reported SRL Data?

The self-reported SRL data were collected twice: once at the beginning of the course (pre-course) and again at the end of the course (post-course). Based on the Online Self-regulated Learning Questionnaire (OSLQ: Barnard, Paton, & Lan, 2008), six perspectives of self-regulated learning data were self-reported by the participants, including goal setting, environment structuring, task strategies, time management, help-seeking, and self-evaluation. Each scale has three to five items. The question is whether the students' self-reported SRL data collected before they began the course significantly differ from their self-reported SRL data collected after the course ended. In order to compare these two sets of scores, the mean score of each scale was calculated. Table 4.4 shows the comparison of the pre- and post-course self-reported SRL data. An informal inspection indicates that the means and standard deviations of each scale appear similar. A review of the correlations of the pre- and post-course data show a range of between 0.428 and 0.545 (Table 4.4). The critical r (two-tailed) value is 0.244 ($p < .05$) and 0.399 ($p < .001$) with a sample size $n = 65$. Therefore, students' self-reported SRL data are positively correlated. According to the guidelines for social sciences proposed by Cohen (1988), the effect size is medium. Based on the correlation results, it is apparent that there are some differences between the pre- and post-course data.

Table 4.4.

Comparison of the Pre- and Post-course Self-reported SRL Data

Scale	Mean	Std. Deviation	Correlations
GoalPre	4.258	.555	.473
GoalPost	4.305	.511	
EnvirPre	4.220	.585	.510
EnvirPost	4.29	.646	
StratPre	3.139	.706	.450
StratPost	3.169	.679	
TimeMPre	3.569	.676	.522
TimeMPost	3.590	.790	
HelpSPre	3.196	.705	.428
HelpSPost	2.919	.777	
EvaluPre	3.268	0.829	.545
EvaluPost	3.223	0.771	

In order to check if there are any significant differences between the means of pre- and post-course self-reported SRL data, normality tests were conducted, and the results showed that three pairs of data were normally distributed, while the other three pairs of data violated the normal distribution assumption. Therefore, both paired samples test and Wilcoxon signed rank tests were conducted. The results (Table 4.5) show that only one scale, help-seeking, has significantly different means between the pre- and post-course self-reported data. Therefore, based on the results, there are no significant mean differences between the pre- and post-course self-reported SRL data except the help-seeking scale. The results indicate that students thought they would be more likely to ask for help in the pre-course survey than in the post-course survey.

Table 4.5.

The Mean Differences of Pre- and Post-course Self-reported SRL Data

Scale	Goal setting	Environment structuring	Time Management	Task Strategies	Self-Evaluation	Help-seeking
Test		Wilcoxon signed rank tests			Paired samples tests	
Sig.	.497	.358	.952	.734	.671	.003*

$*p < .05$

To better answer the question of how the trace data reflect the self-reported SRL data, correlations of the trace data and self-reported SRL data were reported in Table 4.6 using $r_{critical} = .244, p < .05$ and $r_{critical} = .399, p < .001$. Because there are no corresponding trace data related to environment structuring, this scale was excluded from the data analysis. Both pre- and post-course self-reported SRL data were analyzed. More significant correlations were found between the post-course self-reported SRL data and the learning behavior data than between the pre-course data and the learning behavior data. This may indicate that students' post-course self-reported SRL data are more accurate than that of the pre-course. Based on the results shown in Table 4.6, it is apparent that the following three learning behavior variables aptly reflect students' self-reported SRL data: (1) bonus quiz completion rate; (2) average time accessing the eLC course per module; (3) lecture completion rate; and (4) quiz retake rate. The reason is that they are significantly correlated to at least two self-reported SRL scales. There are another eleven learning behavior variables that are significantly correlated to at least one self-reported SRL scale: (1) Time spent on viewing syllabus, (2) time spent on viewing rubrics, (3) average time spent on additional resources, (4) the average number of days visiting the course per module, (5) average topics visited per module, (6) total items visited per module, (7) number of discussion threads created, (8) number of discussion replies created, (9) average late submission number per module, (10) average advance module completion, (11) unit completion rate. The majority are positively correlated, but there are some negative correlations. It is not surprising that the average late submission rate is negatively correlated with the goal setting of the post-course survey ($r(64)=-.253$, $p<.05$), but other negative correlations are interesting and worth some discussion. First, students' self-reported help-seeking scale (post-course survey) is negatively correlated with the average lecture completion rate ($r(64)=-.326$, $p<.05$) as well as average time spent on additional resources ($r(64)=-.269$, $p<.05$). Secondly, students' self-reported self-evaluation scale (post-course survey) is negatively correlated with the average topics visited per module ($r(64)=-.250$, $p<.05$), total items visited per module ($r(64)=-.348$, $p<.05$), and lecture completion rate ($r(64)=-.387$, $p<.05$). Lastly, students' self-reported task strategies scale (post-course survey) is negatively correlated with the average time spent on viewing syllabus ($r(64)=-.250$, $p<.05$) and average login per module. It could be that the students who needed viewing the course syllabus more often had downloaded the course syllabus so that they could view it offline. These supervising negative correlations may imply that the SRL questionnaire may need some improvement to well reflect students' self-regulatory ability.

Overall, the trace data in LMS correlate with the self-reported SRL data to some degree. The learning behavior data collected in LMS correlated well with the goal-setting scale in the SRL questionnaire but did not correlate well with the task strategies and time management scales. The worst were the help-seeking and the self-evaluation scales because at least two negative correlations existed between these two scales and the trace data. It also shows that most of the time management learning behavior variables positively correlated with the goal-setting of the self-reported SRL data.

Table 4.6.

The Correlations of the Learning Behavior Variables and Self-reported SRL Data

	GoaPr	GoaPo	StrPr	StrPo	TimPr	TimPo	EvaPr	EvaPo	HelPr	HelPo
SylFre	.136	.117	.078	.002	.090	.133	-.001	.059	-.031	-.005
SylTim	-.127	-.004	-.160	**-.250***	.080	-.073	-.167	-.139	-.014	-.085
SylDow	.145	-.039	.196	-.015	.123	.131	.126	.068	.151	.029
RubFre	.160	.185	.101	-.106	.038	-.041	.147	-.115	.081	.111
RubTim	.214	.047	.094	-.071	.065	-.044	**.254***	-.060	.232	-.032
LoginM	.042	**.244***	-.125	**-.255***	.018	.168	.082	.094	.119	.109
AddRes	.177	.084	.047	-.018	.044	-.104	.174	-.119	.123	**-.269***
DayVisM	.193	**.315***	.113	-.010	.219	.247	.080	-.140	.074	.030
TopVisM	.202	.202	.072	.180	.143	.049	-.070	**-.250***	.023	-.145
TotVisM	.175	.233	.058	-.008	.072	.005	-.036	**-.348***	-.034	-.159
TimeM	**.261***	.214	.106	.077	**.250***	.180	-.121	-.138	-.113	-.100
#PostRea	.115	.166	.060	.114	.017	.045	.056	-.044	.078	.178
#DisThr	**.364****	.195	.155	.145	.215	.084	.075	-.062	.096	-.002
#DisRep	.214	**.287***	.170	.154	.111	.038	.014	-.120	.040	.016
LecComR	.085	.086	-.057	-.098	.105	-.040	-.160	**-.387***	-.079	**-.326***
LatSubAv	-.106	**-.253***	.058	-.005	-.018	-.053	-.060	-.041	-.048	-.023
Advance	.112	**.322***	-.078	.163	.131	.123	-.118	-.204	-.016	-.020
UniComAv	.206	**.248***	.092	.215	.156	.089	-.010	-.169	.039	-.102
RetakeQ	**.246***	.173	.108	.118	.141	.207	.034	**.293***	.057	.168
Revisit	.002	-.130	-.040	-.155	-.070	-.041	-.067	-.201	.026	-.059
SelAssComR	.185	.133	-.091	.073	.119	.077	-.148	-.143	.028	-.164
BonQComR	.238	**.247***	**.250***	**.263***	.236	**.295***	.125	.004	.219	.109

#QAskIns	.005	-.091	-.178	-.012	.060	.068	-.073	-.099	-.133	-.098
#AskExt	-.053	-.129	.070	.038	.126	.189	.074	.063	-.032	-.087

Note. See Appendix C for the full description of the codes used in this table.

$*p < .05$ $** p < .001$

RQ2: What Is the Relationship between Performance and the Trace Data and Self-reported SRL Data?

The Relationship between Performance and the Trace Data. Multiple linear regression was conducted to predict final grades based on all the learning behavioral data collected from the LMS. A significant regression equation was found ($F(24,40) = 4.454, p < .001$), with an R^2 of $.728$.

Because the learning behavior data include 25 variables, three feature selection methods were used to eliminate unnecessary predictors: stepwise, backward, and forward. The stepwise and forward methods ended with the same model ($F(5,59) = 15.484, p < .001$), with an R^2 of $.568$, while the backward method ended with a different model ($F(9,55) = 13.451, p < .001$), with an R^2 of $.688$. These two models are similar, and both models include five predictors: (1) the average late submission number per module, (2) the average number of discussion replies per online discussion activity, (3) the average logins per module, (4) the average number of posts read per online discussion activity, and (5) the average number of discussion threads created per online discussion activity. Besides the five key predictors used in the first model, four additional variables were included in the second model as predictors: (1) self-assessment completion rate, (2) average time accessing the course per module, (3) average unit completion rate before deadline, and (4) average topics visited per module.

Therefore, we conclude that the trace data collected from LMS can be used to predict students' course learning outcome. The results show that the trace data could explain around 73% of the variance of students' final grades. The key learning behavior predictors are: (1) late submission; (2) online discussion participation including creating threads and reading and commenting on others' posts; and (3) course login frequency. Additional predictable learning behavior variables may include self-assessment completion rate, average time accessing the course, average unit completion rate before the deadline, and average topics visited. These factors cover three of these six SRL scales: task strategies, time management, and self-evaluation.

Table 4.7.

The Multiple Linear Regression Results of Three Feature Selection Methods

Method	Variables/Predictors	R^2	F	Sig.
Stepwise	LatSubAv, #DisRep, LoginM, #PostRea, #DisThr	.568	15.484	$p < .001$
Backward	#PostRea, LoginM, #DisThr, SelAssComR, TimeM, LatSubAv, #DisRep, UniComAv, TopVisM	.688	13.451	$p < .001$
Forward	LatSubAv, #DisRep, LoginM, #PostRea, #DisThr	.568	15.484	$p < .001$

Note. See Appendix C for the full description of the codes used in this table.

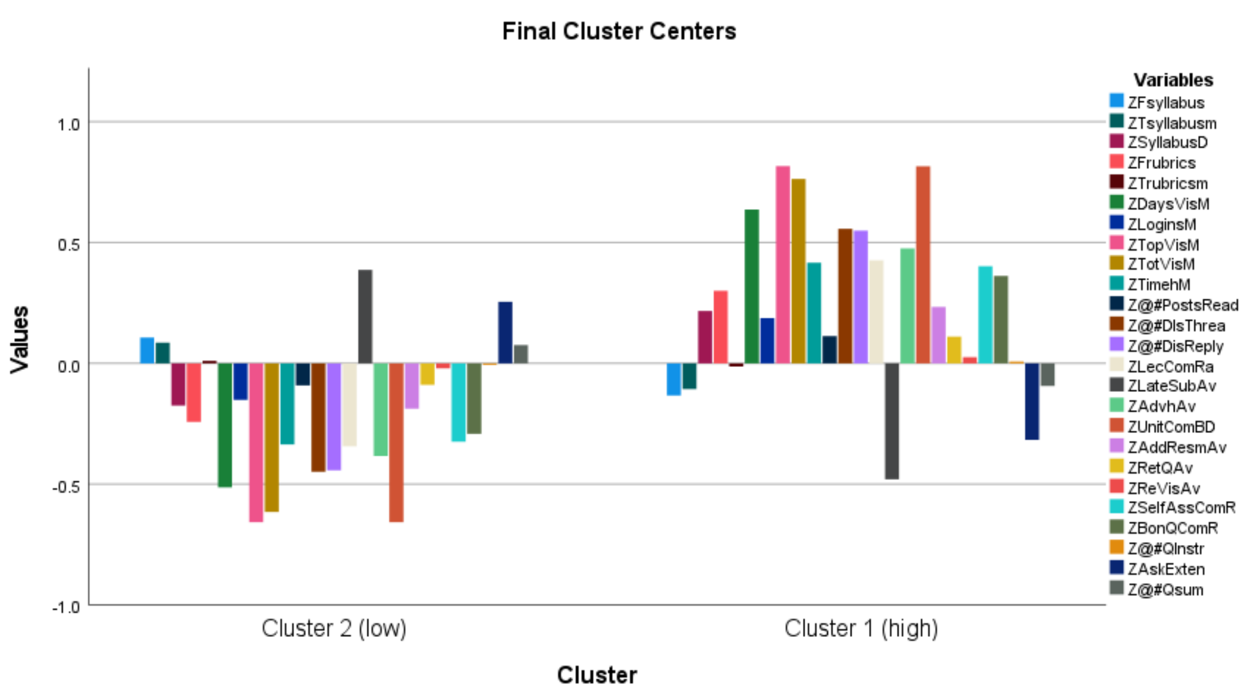
The Relationship between Performance and the Self-reported SRL Data. Multiple linear regression was used to predict final grades using all the self-reported SRL data including both the pre- and post-course survey data. No significant regression equation was found ($F(10,54) = 1.068, p > .05$), with an R^2 of $.165$. Because the data include ten variables, three feature selection methods were used to eliminate unnecessary predictors: stepwise, backward, and forward. All three methods ended with the same model ($F(1,63) = 7.625, p < .05$), with an R^2 of $.108$. The model includes only one predictors: the mean rate of the goal setting in the post-course survey. The results show that the mean rate of the goal setting in the post-course survey could explain around 11% variance of students' final grades. This result supports that the self-reported SRL data collected post-course is more accurate than that collected pre-course. Therefore, it can be concluded that students' self-reported SRL data collected at the end of the course can be used to predict students' final grades, but with some limitations.

In summary, it seems that students' trace data collected from LMS are much more powerful than students' self-reported SRL data in predicting students' final grades.

RQ3: What Learning Patterns Are Present based on the Trace Data?

All the trace data were standardized, and then K-means cluster analysis was conducted based on the standardized z-scores. The K-means cluster analysis was conducted based on two assumptions: (1) students could be categorized into two groups: one group with high self-regulatory ability and the other with low self-regulatory ability; (2) students could be categorized into three groups: one group with high self-regulatory ability, one group with moderate self-regulatory ability, and one group with low self-regulatory ability.

Figure 4.2 shows the final cluster centers of two or three clusters based on students' trace data collected from LMS. The two-cluster results show that these students can be categorized into two groups: one group has 76% (19 out of 25) of the behavioral learning variables with positive cluster centers, while the other group. Apparently, cluster one includes students with high self-regulatory ability, while group two includes students with low self-regulatory ability. Among these 65 students, 29 students were classified into the high self-regulated learners, and 36 students were classified as low self-regulated learners. Based on the ANOVA results, among these 25 behavioral learning variables, 15 of them have significantly different cluster centers in the two clusters. The significant results show that students with low self-regulatory ability tend to submit assignments late and ask for an extension more often, while students with high self-regulatory ability tend to access rubrics, the course, topics, and items more frequently, spend more time accessing the course per module, post more discussion threads and replies, have higher lecture and unit completion rate before deadline, start to work on assignment earlier, and have higher self-assessment and bonus quiz completion rate. The results also show that although there are no significant differences, students with low self-regulatory ability tend to access the syllabus more frequently and spend more time viewing the syllabus, but students with high self-regulatory ability tend to download the syllabus. This counterintuitive result could be interpreted such that students with high self-regulatory ability actually visit the syllabus more frequently or spend more time on the syllabus, so they prefer to download the syllabus and read it offline, but the data of students' study time outside the LMS was not collected. Therefore, this could be attributed to a limitation of the data collection method.



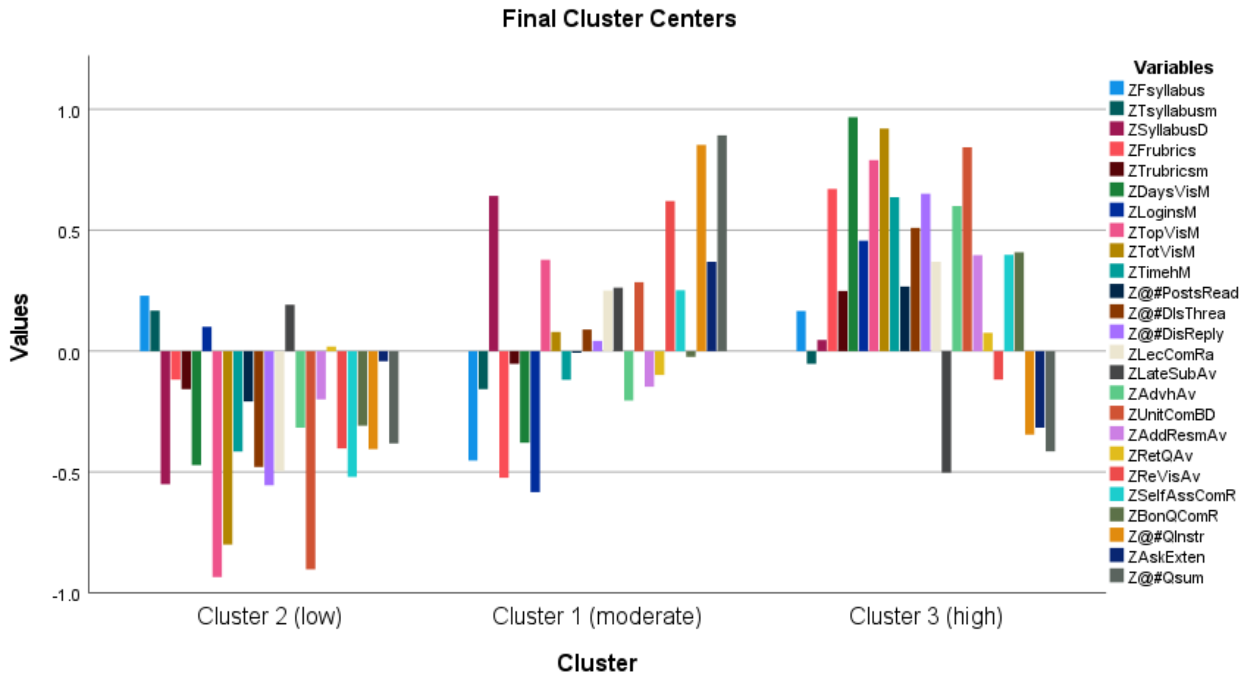


Figure 4.2. The Final Cluster Centers of Two and Three Clusters Using Students' Trace Data

The three-cluster results show that the students can be categorized into three groups: one group has 76% (19 out of 25) of the behavioral learning variables with positive cluster centers, another has 52% (13 out of 25) of the behavioral learning variables with positive cluster centers, and the third group has 20% (5 out of 25) of the behavioral learning variables with positive cluster centers. The trace data cluster analysis results indicate students could be categorized into three levels based on their self-regulatory abilities: high self-regulatory ability (cluster 3), moderate self-regulatory ability (cluster 1), and low self-regulatory ability (cluster 2). Among these 65 participants, 20 students were classified as high self-regulated learners, another 20 students were moderate self-regulated learners, and the remaining 25 students were low self-regulated learners. Based on the ANOVA results, among these 25 behavioral learning variables, 18 have significantly different cluster centers among three clusters. The significant results show that high self-regulated learners in general accessed the course more often, visited more topics and items, actively participated in online discussions more often, and spent more time on the course. They also tended to work on assignments or modules ahead of time to avoid late submissions, and the average unit completion before the due date is also higher than the other two clusters. It is surprising that they tended to ask fewer questions and seldom revisited previous modules.

The moderate self-regulated learners had medium cluster center values in most of these behavioral learning variables, but they also had some extreme cluster center values in some of the behavioral learning variables. They tended to be the group that visited the syllabus and rubrics the least frequently, but this may be explained by the highest cluster center value of syllabus downloading. They may prefer to download the syllabus and read it offline. The case may be the same with the rubric, but no rubric downloading data was collected. They also logged into the course the least frequently and had the largest late submission value, but they have the highest cluster center values of revisiting previous modules and asking the instructor questions. Please note that the variable #QAskIns excluded the questions that asked for extension. In general, the moderate self-regulated learners spent less time studying this course and were not good at time management, but they tended to improve their performance through strategies like seeking help from the instructor, revisiting previous modules, reading the syllabus to better understand course requirements and policies, etc.

The low self-regulated learners had the lowest cluster center values in most of these behavioral learning variables. They spent the least time studying, visiting topics and items, and watching lectures. They were the least active group of people in the online discussions. They were the latest group of people to begin to study a module. They also tended to submit assignments late but had a lower cluster center value than the moderate self-regulated learners.

They also logged into the course more frequently than the moderate group. However, they had the lowest cluster center values in self-evaluation and help-seeking. Overall, they are a group of people that spend the least time studying and do not use many learning strategies. They spend the least time on studying but do not submit assignments late as often as the moderate self-regulated learners. It appears that they tend to get the work done in the least time and do not care that much about the quality of their submitted work.

Table 4.8.

Behavioral Learning Variables with Significant Different Cluster Centers in Three Clusters

Variables	Cluster center value in low self-regulated learner	Cluster center value in moderate self-regulated learner	Cluster center value in high self-regulated learner	Matching SRL scale
SylFre	.229	**-.453**	.166	Goal setting
SylDow	-.550	**.642**	.046	
RubFre	-.118	**-.523**	.671	
DayVisM	-.471	-.379	.968	Task strategy
LoginM	.101	**-.583**	.457	
AddRes	-.120	**-.147**	.397	
TimeM	-.415	-.118	.637	Time management
#DisThr	-.479	.089	.510	
#DisRep	-.555	.042	.651	
TopVisM	-.935	.378	.790	
TotVisM	-.800	.080	.920	
LecComR	-.496	.250	.369	
LatSubAv	.192	**.263**	**-.502**	
Advance	-.317	-.204	.600	
UniComAv	-.903	.285	.843	
ReVisit	-.402	**.621**	-.118	Self-evaluation

SelAssComR	-.520	.252	.399	
#QAskIns	-.406	**.853**	-.346	Help-seeking
#QSum	-.382	**.892**	**-.415**	

Note. See Appendix C for the full description of the codes used in this table.

In summary, based on the cluster analysis results of trace data collected from LMS, high self-regulated learners tend to spend the most time and effort in studying and do not seek help much; moderate self-regulated learners tend to spend moderate time and effort in studying, and they also seek help the most often; low self-regulated learners tend to spend the least time and effort in studying and seek help the least often.

As a comparison, cluster analysis was also conducted based on students' self-reported SRL data. Here, both the pre-course and post-course survey data were used in the K-means cluster analysis. The two-cluster results show that these students can be categorized into two groups: one group has 100% (10 out of 10) of the SRL scales with positive cluster centers, while the other group has 100% (10 out of 10) of the SRL scales with negative cluster centers. Among these 65 students, 25 students were classified into the low self-regulated learners, and the remaining 40 students were high self-regulated learners. Based on the ANOVA results, nine of these ten SRL scales have significantly different cluster centers in the two clusters.

The three-cluster results show that these students can be categorized into three groups based on their self-reported SRL data: one group has 100% (10 out of 10) of the SRL scales with positive cluster centers, one group has 60% (6 out of 10) of the SRL scales with positive cluster centers, and the third group has 100% (10 out of 10) of the SRL scales with negative cluster centers. Among these 65 participants, 29 students were classified as high self-regulated learners, another 23 students were moderate self-regulated learners, and the remaining 13 students were low self-regulated learners. The ANOVA results indicate that all ten of the SRL scales have significantly different cluster centers in these three clusters. The results show: (1) high self-regulated learners have positive cluster centers in all five SRL scales; (2) the moderate self-regulated learners have negative cluster centers in goal setting, task strategies, and time management but positive cluster centers in self-evaluation and help-seeking; (3) low self-regulated learners have negative cluster centers in all five SRL scales.

Based on the cluster analysis of both the trace data and self-reported SRL data, similar learning patterns were found among low, moderate, and high self-regulated learners although the trace data results appeared messier than the self-reported SRL data. However, after checking the classification results of each case based on these two types of data, the results showed that only 53.846% (35 out of 65) cases were classified into the same cluster in the two-cluster analysis, and only 30.769% (20 out of 65) cases were classified into the same cluster in the three-cluster analysis. This indicates that although the trace data and self-reported SRL data have similarities, they are also different. This leads to the next research question: what are the explanations for any differences between students' self-reported SRL data and the digital trace data?

RQ4: What Are the Explanations for Any Differences between Their Self-reported SRL Data and the Digital Trace Data?

Among these 65 participants, 11 students voluntarily participated in a semi-structured interview. The semi-structured interview focused on their typical behaviors while taking this course based on the six SRL scales: goal setting, environment structuring, task strategies, time management, self-evaluation, and help-seeking. The interview also included some questions regarding their typical learning behaviors for completing a module (see Appendix B).

According to the cluster analysis results of trace data and self-reported SRL data, these 11 participants were classified into different clusters. Table 4.9 shows the cluster results, with these 11 students represented by letters $S_a - S_k$. It is apparent that students S_c, S_d, and S_i were classified as high self-regulated learners by all these four cluster analysis results. Student S_a was classified as a high self-regulated learner when using the two-cluster analysis method and as a moderate self-regulated learner when using the three-cluster analysis method. It is fairly straightforward that the difference is mainly caused by the number of clusters, not the data type. For students S_e and S_f, three of the four cluster analysis results classified them as high self-regulated learners, while the trace data 3-cluster analysis classified them as moderate self-regulated learners. For student S_g, three of the four cluster analysis results classified him/her as a low self-regulated learner, while the self-reported SRL data 3-cluster analysis classified him/her as a moderate self-regulated learner. Although students S_e, S_f, and S_g were classified into different groups, the results are close. The big differences exist between the results for students S_b, S_h, S_j, and S_k. For both students S_h and S_k, the trace data classified them as low self-regulated learners, while the self-reported SRL survey classified them as high self-regulated learners. For student S_b, the trace data classified him/her as a high self-regulated learner, while the self-reported SRL survey classified him/her as a low or moderate self-regulated learner. For student S_j, the trace data classified him/her as a high or moderate self-regulated learner, while the self-reported SRL data classified him/her as a low or high self-regulated learner. A large difference exists between two- and three-cluster analyses of self-reported SRL survey data, which may indicate that trace data can reflect a student's self-regulatory ability more accurately than self-reported SRL data. Overall, it can be concluded that both trace data and self-reported SRL data are generally able to classify high self-regulated learners more accurately than moderate or low self-regulated learners. The logical next step is to compare high self-regulated learners with moderate or low self-regulated learners using these two data types to see what causes the differences by incorporating the interview data collected.

Table 4.9.

The Cluster Analysis Results of These Eleven Students

Students	Trace Data 2 clu	Trace Data 3 clu	SRL survey 2 clu	SRL survey 3 clu
S_a	High	Moderate	High	Moderate
S_b	High	High	Low	Moderate
S_c	High	High	High	High
S_d	High	High	High	High
S_e	High	Moderate	High	High
S_f	High	Moderate	High	High
S_g	Low	Low	Low	Moderate
S_h	Low	Low	High	High
S_i	High	High	High	High
S_j	High	Moderate	Low	High
S_k	Low	Low	High	High

Based on the cluster analysis results, students S_c, S_d, and S_i can easily be classified as high self-regulated learners. Three out of the four cluster analysis methods identify students S_e and S_f as high self-regulated learners, and the other method identifies them as moderate self-regulated learners, so they can roughly be categorized as high self-regulated learners in order that these eleven students can be divided into two groups with a similar number of students in each group and compare them. Therefore, students S_c, S_d, S_e, S_f, and S_i are grouped into higher self-regulated learners and the rest six students are categorized as lower self-regulated learners. Here, "higher" and "lower" are used because these students are classified comparably and there is no intention to put a final mark on them. Let's first compare these two groups by looking at the correlations between the three types of data: trace data, self-reported SRL survey data, and interview data. Due to the fact that only limited quantitative data can be collected through interviews, the comparison mainly focused on students' time management ability. Table 4.10 shows the correlation values comparison between the higher and lower self-regulated learners achieved by cross-checking these three types of data. Based on previous analysis, trace data reflect students' self-regulated learning more accurately than the self-reported SRL survey, so trace data is compared with both the interview data and the self-reported SRL data. As shown in Table 4.10, overall, higher self-regulated learners tend to self-report their SRL more accurately, and lower self-regulated learners' self-reported data tend to have more negative correlations with trace data. This explains why trace data and self-reported SRL data tend to classify high self-regulated learners accurately. The negative correlations found among lower self-regulated learners also indicate that these students tend to self-rate SRL scales higher.

Table 4.10.

Correlation Comparison between Higher and Lower Self-regulated Learners

Correlations	Higher self-regulated learners	Lower self-regulated learners
Ave. time spent on each module Trace Vs. Interview	.135	.681
Ave. time spent on each module Trace Vs. Survey time management scale pr	.015	**-.575**
Ave. time spent on each module Trace Vs. Survey time management scale po	**-.239**	**-.623**
Ave. days visit per module Trace Vs. Interview	**-.295**	**-.865**
Total # of questions asked instructors Trace Vs. Survey help seek pr	.199	**-.629**

Total # of questions asked instructors Trace Vs. Survey help seek po	.410	**-.625**
Ave. assignment advance submission Trace Vs. Interview	.824	**-.240**
Ave. assignment advance submission Trace Vs. Survey time management scale pr	.380	**-.215**
Ave. assignment advance submission Trace Vs. Survey time management scale po	.622	.060

Patterns of Higher Self-regulated Learners. The theme analysis results show that higher self-regulated learners have the following two patterns: (1) getting the assignments done early and (2) being able to identify his/her weakness through self-reflection.

Pattern One: Getting the Assignments Done Early. Students S_c, S_d, and S_i all tended to get the assignments done early. Getting assignments done early is an important sign of high self-regulatory ability.

Student S_c said:

"I am the type of person who doesn't like to procrastinate too much. So especially for this course, I would try and get all my work done in the first half of the week so that I wouldn't have to worry about it for the second half of the week… So all the course assignments were due on Sunday nights, and I will usually work on them on Mondays. And I would have everything done, or I would try to have everything done … well, I would try to have the lectures and the quizzes done by Wednesday. And I would try to have the practical lab assignments and the discussion posts done by Friday, so that I could have the weekend to myself, but sometimes with the practical lab assignments, because they were a little bit more writing, I would try to have them done by like Saturday."

Student S_d expressed a similar thought:

"So if it was a typical week, the assignment opens on Monday and they are closed on Sunday, then I would usually have it done by Friday so I did not have any work to do over the weekend."

Student S_i also indicated that he/she would get all the assignments done by Friday:

"I would start reviewing the PowerPoint and just like broadening my understanding of what we're doing for this course. And then we had our learning activities that come along with the course, such as discussion posts and like assignments. So I would try to do that in the middle of the week. And if I didn't understand those reading assignments or needed more information for my discussion post, I would revisit the lectures that I had initially seen on Monday, but later in the week. And I made my work due by Friday, just in case like I email the professor when I was confused; I can edit it before the due date on Sunday morning."

Based on the above excerpts, although for different reasons, these students all aimed at completing the work at least two days before it was due so that there would be some flexibility when they needed it. The trace data show that for students S_c, S_d, and S_i, their average hours of submitting assignments before the deadline were 58, 91, and 18 hours, respectively. It is noteworthy that although student S_i was taking five courses in the semester, he/she was still able to complete all the assignments 18 hours before the deadline on average, which is not easy. Based on the excerpts from students S_c and S_d, they also seem to be able to balance study and personal life well. They tend to get all their school work done by Friday and enjoy the weekend without the need to worry about studying.

Pattern Two: Being Able to Identify His/Her Weakness through Self-reflection. Being able to identify one's weakness through self-reflection of learning and being honest with oneself is another important sign of high self-regulatory ability.

Student S_c admitted that he/she should work on the graduate school assignment earlier, saying:

> "I think achieving goals is a little bit harder just because I would get lazy sometimes… I think if I were to do anything differently, it would be the graduate school stuff, because I did that all in the last, like month of classes. And so if I had maybe looked at that before, it would be eased my stress a little bit."

Student S_d expressed a similar thought:

> "There was a module for the graduate students that had open all year, all semester and I didn't really start working on it until the end. No, it was like the beginning of April. I gave myself like a month and I should have started it earlier or started thinking about the topics earlier because that module was much more challenging for me from a biology standpoint."

Student S_i also planned to change, saying:

> "Um, I would definitely try to get in contact with my classmates more often just so because I know there's other people out there who probably don't understand some things or just wanted to keep like … keep each other accountable for our work like, 'oh, you are doing this. You're doing this this day.' So I definitely love to do that and probably be more in contact with the professor to just like sending frequent emails and just like trying to like, like a teacher student relationship like this with that again."

Student S_c admitted that he/she sometimes gets lazy. He/She also reflected that she did not start to work on the graduate students' assignments early on in the semester, which was a mistake. Student S_d also reflected that he/she should have started to work on the graduate students' assignments early. Student S_i reflected that he/she should interact with classmates or the professor more frequently. In addition, the reflections of students S_c, S_d and S_i were very specific and obtainable. They admitted their mistakes frankly and did not use any excuse.

Patterns of Lower Self-regulated Learners. The theme analysis results show that lower self-regulated learners tend to procrastinate or study just enough to get their desired grade. They also do not plan to change their study approaches or strategies much. Because the regression analysis shows trace data can reflect students' learning more accurately, students S_g, S_h, and S_k are used here to represent lower self-regulated learners.

Pattern One: Procrastinate or Not Studying Hard. Procrastination or not studying hard is an important sign of low self-regulatory ability.

Student S_g admitted that he/she is dilatory, saying:

> "I mean, there's a couple of weeks where, you know, I just get busy. And I had to take the quizzes on Sunday or to do the discussions on Sunday, but there was always preparation prior to doing them, even if I did have to wait till the last minute to do those, which didn't happen very often."

Student S_g also admitted that he/she sometimes do not want to study although he/she knows it may harm his/her grade and bring stress:

> "I mean, there are times like I may not want to study, but it's also, you know, of a thing where it's like for the grade or good. And I have an understanding of that, like, ok, if I could just go ahead and knock it out now, you know, tonight or later

this week, it won't be as big of a stressor and I'll have a grasp on it and be able to perform well."

Student S_h also admitted that he/she sometimes tend to study in the last minute:

"The only thing is really just procrastination. Sometimes discussion, they would just be like, like…I would feel like it's just like so much work, so I would just kind of like sometimes in the weeks I would just be like, ok, no, just do it the next day because it was just too much and I just didn't want to do it. But I mean, I would get to it at the end of the day. Sometimes I would wait… like for some of them, I waited till last minute. But I mean, I finished everything else."

Student S_k expressed frankly that he/she just study enough to get the wanted grades:

"I did not, I don't I'm not very good at studying. I kind of just know how to do work. Like I calculate what exactly I would need to get the grade that I want and study and study just enough to get there. I wish I was an overachiever, but I'm not."

Both students S_g and S_h tend to procrastinate sometimes. Student S_k did not submit any assignments late, but he/she wanted to study just enough to get the grade he/she wanted.

Pattern Two: Do Not Plan to Make Much Change. Although they tend to procrastinate or not study hard, they do not plan to make much change when asked what they would do differently if they took another online course.

Student S_g did not plan to change much except make some adjustments based on the learning materials:

"Um, oh, I don't know. I think I would probably keep a lot of stuff the same. I mean, I might… depending on what the class like, you know, there might be a little bit like different things. But I think, like, I have a pretty good indication of, like how I learn, how I study, what drives me, how to manage time. And so I think just like sticking to that and ensuring that I'm doing the things right is kind of the same approach I would take. I don't know that would be like a tone of things I'd do different, but it is kind of adjusting to the material of the course or what's at hand."

Student S_h expressed a similar opinion:

"Um, I don't think I would do anything different. I would really just do what I do right now, which is just, you know, planning everything out. Planning, I think, to me is key, (that) works for me for sure, to finish things off.

Student S_k expressed that he/she would watch the lecture videos, saying:

"I would make a commitment to watching the lecture videos more because maybe the experience might be different, maybe that they'll actually have pertinent information that's different from what's in the textbooks in the lecture videos. What else could I do? I actually like online courses … I don't think I'd do anything else differently over that."

Both students S_g and S_h do not plan to change anything although student S_g mentioned that he/she would adjust his/her study approach based on the materials of the course. Student S_k said he/she would commit to watching the lecture videos if they had relevant information that was different from what was in the textbook because he/she didn't watch lecture videos in this class. According to the trace data, student S_k's average lecture completion rate was 2.36%. The question is if he/she does not watch all or most of the lecture videos, how does he/she know that the lecture videos do not have additional useful information other than what's covered in the textbook? It sounds like more of an excuse for not watching lecture videos. Student S_k was taking six courses during the semester, so the main reason maybe is that he/she didn't have much time to watch lecture videos and he/she was fine with the grade he/she got without watching lecture videos. Overall, lower self-regulated learners tend to use excuses more often to make themselves look good or may be even feel good.

Discussion

Overall, the results of this study suggest that digital trace data is more powerful in predicting learners' performance than self-reported SRL data. Learners can be classified into three groups: high self-regulated learners who spend a lot of time and effort in studying, moderate self-regulated learners who spend moderate time and effort in studying but use a lot of study strategies to improve their performance, and low self-regulated learners who spend the least time and effort in studying.

Self-reported SRL Data Pre-course vs. Post-course

Through comparing students' self-reported SRL data between pre-course and post-course, this study found that although they are close in general, post-course self-reported SRL data could reflect students' self-regulation more accurately. This result is consistent with existing research results (Li, Baker, and Warschauer, 2020). Although students' self-regulatory ability is relatively stable, students' reactions in different courses (context) may be different. So, when it is necessary to collect students' self-reported SRL data, collecting them at the end of the course may be a better option than collecting them at the beginning of the course.

Digital Trace Data vs. Self-reported SRL Data

This study also shows that the trace data collected from LMS can reflect learners' self-reported SRL data to some degree. The learning behavior data from LMS correlate well with the goal-setting scale in the self-reported SRL data, but do not correlate much with the task strategies and time management scales, and correlate the worst with the self-evaluation and help-seeking scales. It is worthwhile to focus additional attention on the self-evaluation and help-seeking scales because some negative correlations were found between these two scales and the learning behavior variables. The help-seeking scale could be interpreted differently in different learning contexts. For example, it could be interpreted that students who ask more questions are more responsible for their learning, but it also could be interpreted that students who ask more questions do not master the learning materials or fully follow the instructions so that they need to ask more questions to get clarifications or guidance from the instructor. In the context of this study, the latter seems to be true because this course provided a clear structure and detailed instructions. For all the online discussions and lab report assignments, rubrics were given with grading criteria and performance levels. During the interview, students also agreed that this course was structured well and had clear instructions. But the results of this study indicate that students' self-reported help-seeking scale was negatively correlated with the average lecture completion rate as well as average time spent on additional resources. It could be that students who spent more time on additional resources and lectures tended to have fewer questions, so they rated their help-seeking scale lower than the rest of the students. It also could be that students who spent less time on lectures and additional resources tended to ask more questions, so they rated their help-seeking scale higher than the rest of the students. Therefore, the help-seeking scale in self-reported SRL data cannot well reflect students' self-regulatory ability.

For the self-evaluation scale, it is also arguable if it truly reflects learners' self-regulation ability because different interpretations can occur. It is generally believed that students who do self-evaluation more often tend to be higher in self-regulation. However, this study shows that students' self-reported self-evaluation scale (post-course survey) is negatively correlated with the average topics visited per module, average items visited per module, and lecture completion rate, but positively correlated with average rate of retaking quizzes. This indicates that students tended to rate their self-evaluation higher when they retook quizzes more often, but their lecture completion rate was lower and they visited fewer topics and items per module on average. These students seemed to take advantage of retaking quizzes to get higher grades instead of mastering the learning materials by watching lectures or studying materials.

It is surprising that among these five self-reported SRL scales, only the goal-setting scale predicts students' learning performance. Both self-evaluation and help-seeking scales are problematic in predicting students' learning performance. These two scales can be interpreted differently in different contexts because they can be used as strategies to improve learners' grades. Therefore, in future self-regulated learning studies, it is important to be cautious with the use of these two scales when using self-reported SRL questionnaires to collect data. It is also questionable if the self-reported SRL instrument is valid in its ability to reflect students' true self-regulatory ability.

By comparing students' self-reported SRL data with their trace data collected through the course, we found that students' digital trace data from the LMS reflect students' learning more accurately than self-reported SRL, which is consistent with existing research results (Cho & Yoo, 2017; Hadwin, Nesbit, Jamieson-Noel, Code, and Winne, 2007; Li, Baker, & Warschauer, 2020). Digital trace data can predict students' performance more accurately than self-reported data, but they are heavily impacted by the context of the course. For example, a student who spends little time and effort on this course may spend a lot of time and effort on a different course due to different perceived demands and interests. Therefore, it is questionable to interpret students' self-regulatory ability mainly based on the digital trace data of a particular course. A combination of digital trace data from multiple courses may be more reliable than that of a single course. We also should be cautious about potential ethical issues when using digital trace data to interpret students' self-regulated learning abilities. Besides the potential bias that exists in using digital trace data from a limited number of courses, some students may prefer to download learning materials and study offline and digital trace data will not account for this. Using digital trace data alone may lead to inaccurate judgments about the participation and performance of these students. Just as Perrotta (2013) pointed out: "the decontextualized analysis of student data and the powerful performativity arguments that underpin them may subvert concerns for social equity and justice" (p. 119). More research should be conducted to explore and establish possible ethical practices and policies in relation to the use of digital trace data. For example, increasing the transparency in data ownership, analysis, and use may be one ethical practice that could be used (Pardo, A. & Siemens, G., 2021).

Important Learning Behavior Variables

By using linear regression analysis, we found the following learning behavioral variables explain the most variance of students' academic performance: login frequency, online discussion participation including creating and reading posts as well as commenting, the timing of assignment submission (termed "procrastination" in the literature), time investment, unit and/or optional assessment completion, and the number of topics visited. This result is also consistent with existing research results (Cho &Yoo, 2017; Colthorpe, Zimbardi, Ainscough, & Anderson, 2015; Gelan et al., 2018; Kim, Yoon, Jo, & Branch, 2018; Lawanto, Santoso, Lawanto, & Goodridge, 2014; Li, Flanagan, Konomi, & Ogata, 2018; You, 2016). The results indicate that the most important learning behavioral variables that impact students' learning are the following:

study regulation, online discussion interactions, the timing of assignment submission, time investment/effort regulation, and completion. According to Pintrich's SRL Model Phase, these learning behavioral variables mainly belong to the monitoring and control phases. Compared to the results of previous research results, the results of this study show that students' online discussion interactions are the most important learning behavioral variables, which have been neglected by some existing studies. However, online discussions are worth about 28% of the final grade, which may be the reason why online discussion interactions have been found to be the most important behavioral learning variable. More research should be conducted to determine whether or not this is true.

Learning Behavior Patterns in Different Self-regulation Levels

Through cluster analysis, similar patterns were found among both the digital trace data and the self-reported SRL data. Because trace data do not have environment structuring data, the remaining five scales of SRL are the main focus here. The results show: 1) high self-regulated learners can do well in five perspectives of SRL; 2) the moderate self-regulated learners can do moderately well in goal setting, task strategies, and time management but are especially good at self-evaluation and help-seeking; and 3) low self-regulated learners cannot do well in all five perspectives of SRL. Existing self-regulated learning theories tend to categorize students into two groups, such as proactive and reactive learners (Zimmerman, 2013) and learners with mastery approach and performance approach goals (Pintrich, 2000). Existing research also tend to cluster students into two or three groups, such as Kim, Yoon, Jo, and Branch (2018) discovered three clusters based on student log data, while Pardo, Han, and Ellis (2017) identified two clusters based on students' self-reported SRL variables and academic performance. Although this study shows evidence to support both two and three clusters, it is more practical to categorize students into three clusters. Bain (2004) suggested that most learning approaches fall into three categories: 1) the surface approach, in which students are interested primarily in surviving the course; 2) the strategic approach, in which students are driven by a desire to receive good grades; and 3) the deep approach, in which students are learning for mastery, conceptual understanding, and critical thinking. The results of the three cluster analysis in this study are consistent with what Bain (2004) suggested: high self-regulated learners use the deep approach, so they spend a lot of time and effort in studying; the moderate self-regulated learners use the strategic approach, so they tend to use learning strategies such as self-evaluation and help-seeking more often to improve their grades; and low self-regulated learners use the surface approach, studying just enough to get the grades they wanted.

By comparing the trace data with self-reported SRL data and the interview data, it was found that high self-regulated learners tend to be more consistent when reporting their self-regulatory ability. Through thematic analysis, it was found that high self-regulated learners tend to submit assignments early and are able to self-reflect and see their own shortcomings clearly. In contrast, low self-regulated learners tend to procrastinate and, surprisingly, do not plan to change their behaviors much. One possible explanation is that low self-regulated students may know their own weakness in self-regulation, but they choose not to admit it. Based on Boekaerts' dual processing model (Boekaerts, 2011), these low self-regulated learners might choose a well-being pathway to protect their ego from damage. According to Zimmerman (2013), it also could be that high self-regulated learners are proactive learners who can plan their learning strategically in order to see their limitations, while low self-regulated learners are reactive learners who cannot identify their own weaknesses without comparing themselves with others. Based on Pintrich (2000), this also shows that high self-regulated learners are concerned more about learning and improvement with mastery approach goals, while low self-regulated learners focus on demonstrating competence with performance approach goals. It is not clear if low self-regulated learners truly are unable to identify their own weaknesses or, instead, they are choosing not to admit that these weaknesses exist. How to help low self-regulated learners to identify their own limitations and motivate them to change might be worth further exploration in future research. Specifically, the following research questions are worth further research: What factors lead students to be proactive or reactive learners? How to train them to be proactive learners? How to change students from performance oriented to mastery oriented?

Limitations of the Study

Although students in this course came from several different disciplines, one of the limitations of this study is that data were collected from a single online course. Further studies in different disciplines and multiple online courses should be conducted to further validate the results of this study. In addition, only part of the students in the class agreed to participate in this study. For the interview, only 11 students were involved, which may bring bias. Further studies with more complete data should be conducted to yield more convincing results.

Another limitation of this study is that the course final grade was used as the dependent or outcome variable in the linear regression. It may be better to use students' GPAs as the outcome variable because students' self-regulatory ability is a relatively stable ability and GPAs are a kind of longitudinal data, which may match better than the final grades of a particular course. In this study, final grades were used because they reflect students' learning very well, as explained in the methods section. In addition, it is difficult to acquire students' GPAs. However, this study involved three types of data resources, and data triangulation was used to ensure the validity of the findings.

Conclusion

By comparing digital trace data with self-reported SRL data, this study found that digital trace data from LMS could predict students' performance more accurately that self-reported SRL data. This indicates that using digital trace data from LMS to measure students' self-regulatory ability might be a better approach than self-reported SRL data through questionnaires. Through cluster analysis, this study found that students can be categorized into three groups: high, moderate, and low self-regulated learners, which aligns well with three different learning approaches: the deep approach, the strategic approach, and the surface approach. This study also found that low self-regulated learners cannot identify their own limitations in their self-regulatory ability, or they may even try to hide them to protect their ego from damage. They also do not have the intention to change. Further research should be conducted to identify practical approaches to help them.

References

Bain, K. (2004). *What the Best College Teachers Do*. Harvard University Press.

Barnard, L., Lan, W. Y., To, Y. M., Paton, V. O., & Lai, S. L. (2009). Measuring self-regulation in online and blended learning environments. *The Internet and Higher Education*, *12*(1), 1-6.

Barnard, L., Paton, V., & Lan, W. (2008). Online self-regulatory learning behaviors as a mediator in the relationship between online course perceptions with achievement. *The International Review of Research in Open and Distance Learning, 9*(2), 1–11.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*, 77-101.

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

Cho, M., Kim, Y., & Choi, D. (2017). The effect of self-regulated learning on college students' perceptions of community of inquiry and affective outcomes in online learning. *The Internet and Higher Education, 34*, 10-17.

Cho, M. & Yoo, J. S. (2017). Exploring online students' self-regulated learning with self-reported surveys and log files: a data mining approach. *Interactive Learning Environments, 25*(8), 970-982.

Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.

Colthorpe, K., Zimbardi, K., Ainscough, L., & Anderson, S. (2015). Know thy student! Combining learning analytics and critical reflections to increase understanding of students' self-regulated learning in an authentic setting. *Journal of Learning Analytics, 2*(1), 134-155.

Crandall, B., Klein, G., & Hoffmann, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis.* Cambridge, MA, US: MIT Press.

Crocker, L. M., Algina, J (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart, and Winston.

Dray, B. J., Lowenthal, P. R., Miszkiewicz, M. J., Ruiz-Primo, M. A. & Marczynski, K. (2011). Developing an instrument to assess student readiness for online learning: A validation study. *Distance Education, 32*(1), 29-47.

Frank, E., Hall, M. A., & Witten, I. H. (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufman, Fourth Edition.

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

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., and Winne, P.H. (2007). Examining trace data to explore self-regulated learning. *Metacognition Learning, 2*, 107-124.

Hew K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review, 12*, 45-58.

Hoppe, H. U. (2017). Computational methods for the analysis of learning and knowledge building communities. In Lang, C., Siemens, G., & Wise, A., & Gasevic, D. (Ed.), *Handbook of Learning Analytics* (pp.23-33). SOLAR.

Hwu, F. (2003). Learners' behaviors in computer-based input activities elicited through tracking technologies. *Computer Assisted Language Learning,16*(1), 5–29.

Kim, D., Yoon, M., Jo, I.-H., & Branch, R. M. (2018). Learning analytics to support self-regulated learning in asynchronous online courses: A case study at a women's university in South Korea. *Computers & Education*, , 233 – 251.

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

Kizilcec, R. K., & Schneider. E. (2015). Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. *Transactions on Computer-human Interactions (TOCHI), 22*(2), 24.

Lawanto, O., Santoso, H. B., Lawanto, K. N., & Goodridge, W. (2014). Self-regulated learning skills and online activities between higher and lower performers on a web-intensive undergraduate engineering course. *Journal of Educators Online, 11*(3).

Li, H., Flanagan, B., Konomi, S., & Ogata, H. (2018). Measuring behaviors and identifying indicators of self-regulation in computer-assisted language learning courses. *Research*

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Pardo, A., Ellis, R., & Calvo, R. A. (2015). Combining observational and experiential data to inform the redesign of learning activities. *LAK'15 Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, 305-309.

Pardo, A., Han, F., & Ellis, R. A. (2017). Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. *IEEE Transactions on Learning Technologies, 10*(1).

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology, 45*(3), 438-450.

Perrotta, C. (2013). Assessment, technology and democratic education in the age of data. *Learning, Media and Technology, 38*(1), 116-122. DOI: 10.1080/17439884.2013.752384

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In Boekaerts, M., Pintrich, P. R., & Zeidner, M. (Ed.), *Handbook of self-regulation* (pp. 452-502). San Diego, CA: Academic Press.

Siadaty, M., Gasevic, D., and Hatala, M. (2016). Trace-based microanalytic measurement of self-regulated learning processes. *Journal of Learning Analytics, 3*(1), 183-214. https://doi.org/10.18608/jla.2016.31.11

Van Rooij, S. W., & Zirkle, K. (2016). Balancing pedagogy, student readiness and accessibility: A case study in collaborative online course development. *The Internet and Higher Education*, *28*, 1-7.

Winne, P. H. & Perry N. E. (2000). Measuring self-regulated learning. In Boekaerts, M., Pintrich, P. R., & Zeidner, M. (Ed.), *Handbook of self-regulation* (pp. 531-566). San Diego: Academic Press.

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist*, *45*(4), 267–276. http://dx.doi.org/10.1080/00461520.2010.517150

You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *The Internet and Higher Education, 29*, 23–30.

Yu, Q., & Zhao, Y. (2015). The value and practice of learning analytics in computer assisted language learning. *Studies in Literature and Language, 10*(2), 90–96.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166-183.

Appendix A: Online Self-regulated Learning Questionnaire

Instructions: Please answer this questionnaire in the context of you completing this course. Please note that here the term "online courses" means "asynchronous online courses." (In contrast, synchronous online courses use zoom or other conferencing meeting tools.)

Item	Subscale
1. I set standards for my assignments in online courses	Goal setting
2. I set short-term (daily or weekly) goals as well as long-term goals (monthly or for the semester).	
3. I keep a high standard for my learning in my online courses.	
4. I set goals to help me manage studying time for my online courses.	
5. I don't compromise the quality of my work because it is online.	
6. I choose the location where I study to avoid too much distraction.	Environment structuring
7. I find a comfortable place to study.	
8. I know where I can study most efficiently for online courses.	
9. I choose a time with few distractions for studying for my online courses.	
10. I try to take more thorough notes for my online courses because notes are even more important for learning online than in a regular classroom.	Task strategies
11. I read aloud instructional materials posted online to fight against distractions.	
12. I prepare my questions before joining in the chat room and discussion.	
13. I work extra problems in my online courses in addition to the assigned ones to master the course content.	
14. I allocate extra studying time for my online courses because I know it is time-demanding.	Time management
15. I try to schedule the same time every day or every week to study for my online courses, and I observe the schedule.	
16. Although we don't have to attend daily classes, I still try to distribute my studying time evenly across days.	

17. I find someone who is knowledgeable in course content so that I can consult with him or her when I need help.	Help-seeking
18. I share my problems with my classmates online so we know what we are struggling with and how to solve our problems.	
19. If needed, I try to meet my classmates face-to-face.	
20. I am persistent in getting help from the instructor(s) through e-mail.	
21. I summarize my learning in online courses to examine my understanding of what I have learned.	Self-evaluation
22. I ask myself a lot of questions about the course material when studying for an online course.	
23. I communicate with my classmates to find out how I am doing in my online classes.	
24. I communicate with my classmates to find out what I am learning that is different from what they are learning.	

Appendix B: Interview Protocol

- Have you ever tried to set up any goals (for this course or for any learning events) during this course? (Goal setting)

Possible probes:

  - o If so, could you identify the most memorable (negative or positive) and the most typical situations where you set up the goals? Could you describe each moment in detail (e.g., when, where, what, why, and the context)?
    - ▪ If the student did not describe it, ask the following questions:
      - What goals did you set up?
      - How often did you set up the goals?
      - What resources did you use in setting up your goals?
      - Did it help your learning? If so, how? If not, why?
      - What challenges did you experience while setting up/achieving the goals?
  - o If not, then could you explain why you did not set up any goals?

- Have you ever tried to set up/choose a study space for this course? (Environment structuring)

Possible probes:

  - o If so, could you identify the most memorable and the most typical situations where you set up/chose a study space? Could you describe each moment in detail (e.g., when, where, what, why, and the context)?
    - ▪ If the student did not describe it, ask the following questions:
      - What kind of study space were you trying to set up/choose?
      - Did it help your learning? If so, how? If not, why?
      - What challenges did you experience while setting up a study space?
  - o If not, then could you explain why you did not set up a study space?

- Have you ever tried to use any study strategies in this course? (Task strategies)

Possible probes:

- If so, could you identify the most memorable and the most typical situations where you used study strategies? Could you describe each moment in detail (e.g., when, where, what, why, and the context)?
    - If the student did not describe it, ask the following questions:
        - What study strategies did you use?
        - What information did you use in choosing the study strategies?
        - What other study strategies were considered or were available to you?
        - How was/were the study strategy/strategies chosen or others rejected?
        - How often did you use study strategies?
        - Did they help your learning? If so, how? If not, why?
        - What challenges did you experience while choosing study strategies?
- If not, then could you explain why you did not use any study strategies?

- Have you ever tried to use any time management strategies in this course? (Time management)

Possible probes:

- If so, could you identify the most memorable and the most typical situations where you used time management strategies? Could you describe each moment in detail (e.g., when, where, what, why, and the context)?
    - If the student did not describe it, ask the following questions:
        - What time management strategies did you use?
        - What information did you use in choosing the time management strategies?
        - What other time management strategies were considered or were available to you?
        - How was/were the time management strategy/strategies chosen or others rejected?
        - How often did you use time management strategies?
        - Did they help your learning? If so, how? If not, why?

            - What challenges did you experience while choosing time management strategies?
    - If not, then could you explain why you did not use any time management strategies?
- Have you ever sought any help from anyone when studying this course? (Help-seeking)

Possible probes:
    - If so, could you identify the most memorable and the most typical situations where you sought help from others? Could you describe each moment in detail (e.g., when, where, what, why, and the context)?
        - If the student did not describe it, ask the following questions:
            - What information did you use in choosing [whom] to ask for help? Why?
            - How did you know that you need to seek help from others?
            - How often did you seek help from others?
            - Did it help your learning? If so, how? If not, why?
            - What challenges did you experience while seeking help from others?
    - If not, then could you explain why you did not seek help from others?
- Have you ever done any kind of self-evaluation when studying this course? (Self-evaluation)

Possible probes:
    - If so, could you identify the most memorable and the most typical situations where you did self-evaluation? Could you describe each moment in detail (e.g., when, where, what, why, and the context)?
        - If the student did not describe it, ask the following questions:
            - How did you know you are on the right track while studying this course?
            - How often did you do self-evaluation?
            - Did it help your learning? If so, how? If not, why?
            - What challenges did you experience while conducting self-evaluation?

- If not, then could you explain why you did not self-evaluate?

- What did you typically do to complete a module?

Possible probes:

- How many hours did you spend on this course per week online/offline?
- How many days did you visit the course per week?
- When would you start to work on a module? When would you finish the module?
- Have you skipped watching any lectures?
- Have you checked any additional resources in each module?
- How many replies do you usually contribute to an online discussion?

- Reflecting back to this online course experience, what did you learn?

Possible probes:

- What would you do the same?
- What would you do differently?

Appendix C: Learning Variable Codes

This table includes the learning variable codes used in the study and their meanings.

Learning Variables	Meaning
AddRes	Average time spend on reviewing additional resources per module.
Advance	Average hours of completing module assessments before the deadline
#AskExt	Number of times asking for submitting assignments after due.
BonQComR	Bonus quiz completion rate.
DayVisM	Average days accessing the course per module.
#DisThr	Number of discussion threads created.
#DisRep	Number of discussion replies created.
LatSubAv	Average number of late submissions per module.
LecComR	Lecture completion rate.
LoginM	Number of logins per module.
#QAskIns	Number of times asking the instructor(s) questions excluding asking for extension.
#PostRea	Number of posts read in discussions.
#QSum	Number of total questions asking the instructor(s).
RetakeQ	Average rate of retaking quizzes.
Revisit	Average number of revisiting the module after a module ends.
RubFre	Number of accessing the rubric(s) during the course.
RubTim	Time spend on accessing the rubric(s).
SelAssComR	Self-assessment completion rate.
SylDow	If the syllabus has been downloaded by the student or not.
SylFre	Number of times visiting the course syllabus.
SylTim	Time spend on accessing the course syllabus.
TimeM	Average time spends on each module.
TopVisM	Number of topics visited per module.
TotVisM	Total items visited per module.

UniComAv	Average unit (module) completion rate.

CHAPTER 5

CONCLUSION

This dissertation investigated how to use the digital trace data collected in a LMS to measure and improve online students' self-regulated learning by comparing them with self-reported SRL data. After identifying the limitations of existing SRL measurements, I decided to use the naturally occurring digital trace data in LMS as a supplement to investigate how online learners self-regulate their learning by comparing it with students' self-reported SRL data. This idea was developed over several years and a pilot study was conducted about three years ago.

The chapters in this dissertation highlight different elements of this study. The first manuscript (Chapter 2) provided a historical review of learning analytics development in the learning, design, and technology field. It concluded that learning analytics could be considered as a new form of course assessment and evaluation in the digital era and provides historical support for using digital trace data to investigate students' self-regulated learning. It also summarized commonly used learning analytic methods and general concerns regarding learning analytics research, which guided the design and data analysis of the study.

The second manuscript (Chapter 3) is a literature review of self-regulated learning theory and measurements, which provided a theoretical foundation for the design of the research study reported in Chapter 4. After examining and comparing all the key existing SRL models, I concluded that the development of these models follows a trend of shifting from general process models to more contextual-based models. By comparing all key existing SRL measurements, it is found that the majority of these measurements focus on learners' self-reported SRL data and are measuring learners' aptitude. The limitation of existing SRL measurements urges us to look for new measurements. Therefore, a new and contextualized theoretical framework for SRL research in online learning environments was proposed to guide the research design.

The third manuscript (Chapter 4) investigated self-regulated learning using digital trace data from LMS by comparing the digital trace data with students' self-reported SRL data. The results showed that the digital trace data is the powerful of the two in predicting students' performance. Key learning behaviors related to students' self-regulatory ability were identified, which include study regulation, online discussion interactions, timing of assignment submission, time investment/effort regulation, and completion. Through comparison of the digital trace data, self-reported SRL data, and qualitative interview data, possible explanations of the differences between the digital trace data and self-reported SRL data were explored and discussed.

Implications

Theoretical Implications

Although several SRL models have been proposed by researchers, the majority of them focus on top-down self-regulation and neglect the learning context and learners' subjective experiences. The theoretical framework proposed in chapter 3 emphasizes the importance of examining SRL in context by embracing the mutual interactions among the key players in the system, which is a new theoretical contribution to the field. It also considers digital trace data as learning behaviors of SRL events, which dynamically and contextually reflect learners' self-regulatory ability. It provides a new perspective for future research to investigate SRL in online learning environments.

It also emphasized the importance of using theory to guide learning analytics study, which provides foundational guidance for future research.

Implications for Practice

Implications of this research could be illustrated from two aspects: one is from the course design and administration perspective, and the other is from the students' perspective.

For the instructors and administrators, it is important to monitor students' following learning behaviors, such as the time students invest in learning, their login frequency, the number of late submissions, discussion participation rate, and unit competition rate. These could serve as alarm signals to instructors indicating that additional support needs to be provided to specific online students. This might improve online students' retention rate as well as their academic performance.

For online students, right now, they only have access to their own digital trace data, which can provide some insights into their own learning habits. But it is not clear if online students know how to use these data for their own benefit or if they are even aware of the availability of these data. It may be more helpful to overly make students aware of some data collected by the LMS such as the average time students spent in each unit, the average student logins per week, etc. so that they can reflect on whether this matches their own perceptions in order to make adjustments to their course and study habits accordingly. By doing this, students can benefit more from the data generated by themselves. The ultimate goal of using the digital trace data from LMS is to help students learn and succeed in all of their college courses. However, as the digital trace data involves ethical issues, more research should be conducted to decide what kind of data can be available to students and what should not. Potential of benefits and damages should be fully assessed before releasing the data to students.

Recommendation for Future Research

Much of existing self-regulated learning research still heavily replies on using students' self-reported SRL data. This study highlights the need for researchers to explore new ways to measure students' self-regulatory ability. The research results of this study show that digital trace data could reflect students' learning accurately. Thus, in the future, digital trace data from LMSs could be used more often in both SRL research and educational research. With the continued development and improvement of technology, learning analytics will likely be used widely to improve teaching and learning. However, we also need to follow ethical rules and protect participants' privacy while using digital trace data to conduct research. More related research should be conducted to both identify related ethics guidance and distribute this guidance to researchers.

As mentioned above, a lot of digital trace data were not available to students. Further research should explore what kind of digital trace data should be available to students and how they impact students' learning. Both ethical issues and potential benefits should be considered when deciding what kind of digital trace data to let students have access to.

The research results also show that low self-regulated learners have much difficulty identifying their own limitations and are reluctant to change. Further research should be done to find possible approaches to help these students overcome this challenge. Is it that low self-regulated learners do not want to admit their limitations in order to protect their ego or is it the case that they really cannot identify their own limitations? What are ways to help them overcome this limitation? What are ways to help students who study with the surface or strategic approach switch to the deep approach? What factors impact the learning approach students choose? These are the questions that are worth further research.

Reflection

It took me a long time to finally identify this research topic and related research methods. Reflecting on the process, I think my philosophical and epistemic beliefs ultimately determine what I choose to research. I am fortunate to have had the freedom to do the research that I am interested to pursue. Through the study development process, I realized that the development process is iterative and progressive in nature. The earlier you try out the idea, the better the outcome. I learned many lessons by completing a pilot study and improved the final study design based on the results of the pilot study. My advice to others is to try out your ideas as early in the research process as possible and complete iterations in order to improve the final research design.