Exploring Relations Between Motivation, Metacognition, and Academic Achievement Through Variable-Centered, Person-Centered and Learning Analytic Methodologies

Wonjoon Hong
wonjoon.hong0220@gmail.com

Follow this and additional works at: https://digitalscholarship.unlv.edu/thesesdissertations

Part of the Educational Assessment, Evaluation, and Research Commons, and the Educational Psychology Commons

Repository Citation
https://digitalscholarship.unlv.edu/thesesdissertations/3497

This Dissertation is brought to you for free and open access by Digital Scholarship@UNLV. It has been accepted for inclusion in UNLV Theses, Dissertations, Professional Papers, and Capstones by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.
EXPLORING RELATIONS BETWEEN MOTIVATION, METACOGNITION, AND ACADEMIC ACHIEVEMENT THROUGH VARIABLE-CENTERED, PERSON-CENTERED AND LEARNING ANALYTIC METHODOLOGIES

By

Wonjoon Hong

Bachelor of Science - Computer Education
Kongju National University, South Korea
2008

Master of Art - Educational Technology
Seoul National University, South Korea
2013

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy – Learning and Technology

Department of Educational Psychology and Higher Education
College of Education
The Graduate College

University of Nevada, Las Vegas
December 2018
This dissertation prepared by

Wonjoon Hong

entitled

Exploring Relations Between Motivation, Metacognition, and Academic Achievement Through Variable-Centered, Person-Centered and Learning Analytic Methodologies

is approved in partial fulfillment of the requirements for the degree of

Doctor of Philosophy – Learning and Technology
Department of Educational Psychology and Higher Education

Matthew Bernacki, Ph.D.
Examination Committee Chair

Michael Nussbaum, Ph.D.
Examination Committee Member

Harsha Perera, Ph.D.
Examination Committee Member

Andreas Stefik, Ph.D.
Graduate College Faculty Representative

Kathryn Hausbeck Korgan, Ph.D.
Graduate College Interim Dean
ABSTRACT

The three studies that comprise this dissertation examine relations between student characteristics, motivations, metacognitive learning processes, and academic achievement. Methodologically, the dissertation demonstrated the potential of multiple types of approaches and data resource types. By employing multiple approaches including variable-centered, person-centered, and learning analytics, researchers can understand learning processes from various angles. In addition, through this triangulation by multiple types of methodological approaches, educational theories could be more thoroughly verified and supported by various empirical findings. Multiple types of data resources are related to analytical methods.

The purpose of the first paper was to examine relations between achievement goals and metacognitive learning behaviors using a clustering analysis and visualization. A clustering analysis conducted with achievement goals produced three goal profiles; 1) mastery-approach, 2) performance-approach, and 3) performance-avoidance identified three goal profiles. The profiles include High Approach, High Mastery, and High Goal Endorsement groups. The finding demonstrated that students in the High Mastery group, who had greater use of the self-assessment tool, obtained higher final grades than other groups could be explained from the perspective of SRL. In addition, learners motivated by mastery approach goals engaged in the greater use of self-assessment quizzes. Students in the High Mastery group also used the tools earlier than other two groups for exam 2. As the most frequently used pattern, sequential pattern mining discovered the repeated use of self-assessment quizzes to monitor their learning. More students in the High Mastery group employ this pattern of metacognitive events than students in the High Performance and High-Goal endorsement groups, particularly during sessions in weeks before exams. A subsequent analysis revealed that for all exams, students who conducted a
repeated behavior pattern indicative of metacognitive monitoring and control outperformed those who did not. From the research, it is confirmed that the person-centered analysis provided authentic and generalizable groups and afforded observation of the learning behaviors of learners with typical combinations of goals. In addition, sequential patterns provide instructor more interesting information on learning processes than the frequency of accesses.

The purpose of the second research was to identify motivational profiles based on multiple types of motivations including self-efficacy, achievement goals, and expectancy-value from an integrative perspective. For this research, a LPA was conducted with ten types of motivational constructs and three kinds of metacognitive learning processes. The LPA identified four motivational profiles; 1) High Cost, 2) High Performance Goals, 3) High Goals and Values, and 4) Low Performance Goals, and three metacognitive profiles; 1) Infrequent metacognitive processing, 2) Checking performance and planning, and 3) Self-assessment. Student demographic information significantly influenced the membership of motivational profiles. Older students tend to have higher self-efficacy, mastery-approach, and values, but low cost than younger ones. In addition, compared to Caucasian and Asian students, underrepresented students tend to be more motivated by higher goals and values than high cost or high performance goals. Lastly, female students are more likely to be members of High performance goals and High goals and values than High cost oriented and Low performance goals and cost than males.

In terms of the relations profiles with academic achievement, Low Performance Goals group showed the best performance. Among metacognitive profile groups, students in Checking performance and planning, and Self-assessment demonstrated similar academic performance. The investigation of relations between two profile groups demonstrated that students in the High cost group are more likely to be a member of self-assessment group than checking performance
and planning as well as of a member of an infrequent metacognitive process than checking performance and planning. In addition, students in high performance and goals and high goals and values groups relative to the low performance goals group more likely to be a member of the infrequent metacognitive process than checking performance and planning. The findings of this research provide authentic motivation status and metacognition learning process as well as their relations. Addition, this research figured out specific motivational profiles through the multiple types of motivations from the integrative perspective. Therefore, instructors can provide more effective and specific interventions to students who have difficulty utilizing metacognitive learning processes, considering motivational status based on multiple motivations. In addition, instructors can understand motivational profiles by demographics so at the beginning of the semester in which the information on students is not enough to identify students learning processes, they intervene students based on demographic information.

The purpose of the third paper was to consider the relative importance of capturing demographic, motivational and metacognitive processes as potential predictors of learning outcomes, and appraises them alongside both traditional prediction modeling approaches in higher education, and emergent methods, sequence pattern mining, arising from the field of educational data mining. The sequence pattern mining discovered the repeated use of self-assessment quizzes in Biology and repeated use of planning contents in Math. A regression model with combined resource types demonstrated the improved predictive power than models with individual resource types. Also, theory-aligned behaviors designed based on metacognitive learning processes better improved the accuracy of the model than non-theory-aligned behaviors automatically provided by the system. Lastly, when applying the same prediction model, the model better explained the variance of academic achievement in Biology in which metacognitive
supporting tools designed based on an educational theory than that in Math that has few theory-aligned behavior variables.

Therefore, this study emphasizes the importance of existing ambient data from university systems. Also, log data generated by systems such as LMS allows researchers to examine the same data in different ways with no need for additional data collection. Lastly, educational theory and contexts should be taken into consideration in designing courses and developing the prediction models. Therefore, instructors and researchers, in designing courses, the consideration of educational theories and contexts is the essential process.

This dissertation provides insight regarding authentic relations between motivation, metacognition, and academic achievement. Specifically, instructors can understand how multiple types of motivations work together and the motivational profiles influence metacognitive learning strategies. In courses, by examining motivational profiles, instructors can provide more effective intervention with which students change their resolve their weak learning easier.

Practically, by investigating each type of predictor from data resources including demographic, motivation, and behavioral variables, findings from this dissertation can enable researchers to prioritize development of prediction models to identify students who are more likely to experience failure in courses. Additionally, instructors can figure out the importance of interpreting variables through educational theories and in context through the comparison of courses with differing instructional designs. Further, by appraising these results in light of theory, instructors can take action to improve student’s learning outcomes by adjusting the design of their courses.
ACKNOWLEDGMENTS

I am grateful to the Department of Educational Psychology and Higher Education for giving me the opportunity to develop my academic career. I would like to express the deepest appreciation to my committee chair, Dr. Matthew Bernacki, who has encouraged me to pursue my doctoral program without giving up and has provided advice, comments, and support. Additionally, he has been sincere and helpful when I am in trouble with academic and even non-academic challenges.

Dr. Nussbaum, your advice helps me with the improvement of my dissertation. Your suggestions and comments have contributed to my achievement.

Dr. Perera, your valuable methodological comments have led to my intellectual growth. Your suggestions and advice have given me the opportunity to improve my dissertation.

Dr. Stefik, your comments, encouragement, kindness were greatly appreciated.

To my lovely family, I wanted to express the biggest thanks. To my son, Woora, who has expressed great encouragement and his love to me. To my daughter, Yeeun, who always gives me lovely smiles and hugs. Last, but not least, I would like to express special thanks to my wife, Hyeyoung, who has supported and believe in me for last four years. If she did not sacrifice her career for my kids and me, I could not make this achievement.

I promise that I will support my family for their happiness from now.
TABLE OF CONTENTS

ABSTRACT ........................................................................................................................................... iii

ACKNOWLEDGMENTS ......................................................................................................................... vii

TABLE OF CONTENTS ....................................................................................................................... viii

LIST OF TABLES .................................................................................................................................... xiii

LIST OF FIGURES ............................................................................................................................... xv

Chapter 1: Exploring Relations between Motivation, Metacognition, and Academic Achievement through Variable-centered, Person-centered and Learning Analytic Methodologies .......................................................... 1

Chapter 2: Examining the Influence of Undergraduates’ Achievement Goals on Metacognitive Behavior Sequences, and Achievement in Science ................................................................................................................. 9

1. Introduction ..................................................................................................................................... 9

2. Theoretical background .................................................................................................................... 12

   A. Achievement Goal Theory ........................................................................................................... 12

   B. Metacognitive monitoring ....................................................................................................... 17

3. Method .......................................................................................................................................... 23

   A. Participants ............................................................................................................................ 23

   B. Measures ............................................................................................................................. 23

   C. Procedure .......................................................................................................................... 25

4. Results .......................................................................................................................................... 27

viii
Chapter 3: A Latent Profile Analysis of Undergraduates’ Achievement Motivations and Metacognitive Behaviors, and their Relations to Achievement in Science

1. Introduction

2. Literature Review

   A. Self-Efficacy
   B. Achievement Goal Theory
   C. Expectancy-Value Theory
   D. An integrative view of multiple motivation types
   E. Person-centered approach
   F. Metacognition

3. Methods

   A. Participants and Procedures
B. Measure .................................................................................................................. 63

C. Data Analysis........................................................................................................... 66

4. Results....................................................................................................................... 69
   A. Motivation profiles ............................................................................................... 69
   B. Metacognition profiles......................................................................................... 76

5. Discussion............................................................................................................... 83
   A. Motivation profiles .............................................................................................. 83
   B. Motivation profiles by demographic information ............................................... 86
   C. Metacognitive behavior profiles .......................................................................... 87
   D. Relationships between motivation profiles and metacognition profiles............ 88

Chapter 4: Examining the Power of Multiple Data Sources in Predicting Academic Achievement in Undergraduate STEM Courses ........................................................................ 90

1. Introduction............................................................................................................ 90

2. Theoretical Background.......................................................................................... 94
   A. Data types used in predicting academic achievement ........................................ 94
   B. LMS Behavioral Data and Opportunities for Prediction and Intervention.... 100
   C. The current study ................................................................................................. 105

3. Methods................................................................................................................ 105
   A. Participants ......................................................................................................... 105
   B. Measures ............................................................................................................ 106
C. Analysis .......................................................................................................................... 111

4. Results............................................................................................................................. 113
   A. Sequential Patterns ..................................................................................................... 113
   B. Regression Analyses with individual resource types .............................................. 116
   C. Hierarchical Regression Analyses with multiple resource types.......................... 119

5. Discussion....................................................................................................................... 122
   A. What is the relative benefit of data types?................................................................. 122
   B. Differences in the predictive accuracy between theory-aligned vs. non-theory-
      aligned behavior ......................................................................................................... 128
   C. What patterns of theory-aligned behaviors emerge? ............................................... 129
   D. Different predictive power by course .......................................................................... 132

6. Implications...................................................................................................................... 133

Chapter 5: Synthesis, Conclusion and Implications .......................................................... 135

   1. Methodological advancement ..................................................................................... 135
   2. Theoretical contributions ........................................................................................... 136

   A. Examining the Influence of Undergraduates’ Achievement Goals on
      Metacognitive Behavior Sequences, and Achievement in Science........................... 136
   B. A Latent Profile Analysis of Undergraduates’ Achievement Motivations and
      Metacognitive Behaviors, and their Relations to Achievement in Science .............. 137
   C. Examining the Power of Multiple Data Sources in Predicting Academic
      Achievement in Undergraduate STEM Courses ......................................................... 138
3. Synthesis & Conclusion

Appendix

References

CURRICULUM VITAE
LIST OF TABLES

Table 1. Achievement goal framework (Elliot & McGregor, 2001) ......................................................... 14
Table 2. Descriptive statistics and group comparison .................................................................................. 28
Table 3. Result of multiple regression analyses of monitoring behaviors and exam scores ........ 33
Table 4. Result of T-test of Exam Scores .................................................................................................. 37
Table 5. Demographic information of the sample ..................................................................................... 63
Table 6. Omega coefficient (ω) and Cronbach's alpha (α) ..................................................................... 64
Table 7. Grand Means, Standard Deviations, and correlations of constructs ........................................ 69
Table 8. Latent Profile Fit Statistics for Models Based on the Ten Motivation Types ......................... 70
Table 9. Posterior Probabilities and Cross-probability of Motivation Profiles ................................ 72
Table 10. Comparison of Motivation Constructs by Profile ................................................................. 73
Table 11. Results of Multinomial Logic Regressions for the Effects of Predictors on Motivation Profile Membership ........................................................................................................ 75
Table 12. Latent Profile Fit Statistics for Models Based on the Three Metacognition Behaviors 76
Table 13. Posterior Probabilities and Cross-probability of metacognitive learning profiles ............... 78
Table 14. Contingency Table of motivation and metacognition profiles ............................................ 79
Table 15. Results of Multinomial Logistic Regression for the Effects of Motivation Profiles on Metacognition Profile Membership ......................................................................................... 80
Table 16. Items to measure demographic information ........................................................................... 106
Table 17. Nature of Exam 1 in Math and Biology .................................................................................. 108
Table 18. Features assigned to Resource Type ..................................................................................... 110
Table 19. Predictive Modeling Process ................................................................................................ 113
Table 20. Initial Items .................................................................................................................. 114
Table 21. Sequential Patterns in Biology and Math ................................................................. 115
Table 22. Resoue types of Independent Variables .................................................................. 116
Table 23. The result of Regression Analyses with Individual Resources in Biology a .......... 118
Table 24. The result of Regression Analyses with Individual Resources in Math a .............. 119
Table 25. Result of Hierarchical Regression Analysis in Biology a ........................................ 121
Table 26. The result of Hierarchical Regression Analysis in Math a ........................................ 122
Table 27. Factor Loadings ........................................................................................................ 142
Table 28. Factor Correlations ................................................................................................ 143
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data transformation for pattern mining</td>
</tr>
<tr>
<td>2</td>
<td>Elbow Chart (left) and Gap Statistic Method (right)</td>
</tr>
<tr>
<td>3</td>
<td>Scores by Achievement Goal Cluster. Red outline regions indicate discriminating features of clusters who differ in their endorsement of achievement goal items by subscales for Performance Approach and Performance Avoidance constructs. (MAP: mastery-approach; PAP: performance-approach; PAV: performance-avoidance)</td>
</tr>
<tr>
<td>5</td>
<td>Scores on exams by Achievement Goal Group</td>
</tr>
<tr>
<td>6</td>
<td>Use of monitoring self-assessment and performance by achievement goal profile. Frequency indicates the number of accesses of web-hosted self-assessment quizzes (top) or the number of visits to a table showing performance on scored assignments in the course (bottom). Red lines indicate weeks of course exams</td>
</tr>
<tr>
<td>7</td>
<td>Use of monitoring self-assessment and performance by final grade</td>
</tr>
<tr>
<td>8</td>
<td>Use of metacognitive monitoring patterns by achievement goals</td>
</tr>
<tr>
<td>9</td>
<td>Exam scores by use of the metacognitive monitoring pattern</td>
</tr>
<tr>
<td>10</td>
<td>Reorganization of Time Frame</td>
</tr>
<tr>
<td>11</td>
<td>Final model with four profiles. SELF = self-efficacy; MAP = mastery-approach; PAP = performance-approach; PAV = performance-avoidance; ATT_V = attainment value; INT_V = intrinsic value; UTI_V = utility value; EFF_C = effort cost; OPP_C = opportunity cost; PSY_C = psychological cost</td>
</tr>
<tr>
<td>12</td>
<td>Exam scores by motivation profiles</td>
</tr>
<tr>
<td>13</td>
<td>Final model with three metacognition profiles</td>
</tr>
</tbody>
</table>
Figure 14. Exam scores by Metacognition Profiles ............................................................... 78

Figure 15. Changes in Monitoring Learning over a semester .............................................. 81

Figure 16. Changes in Monitoring Performance over a Semester ........................................ 82

Figure 17. Changes in Planning over a Semester ................................................................. 83
Chapter 1: Exploring Relations between Motivation, Metacognition, and Academic Achievement through Variable-centered, Person-centered and Learning Analytic Methodologies

Motivational and metacognitive learning processes play critical roles in the self-regulated learning (SRL) process (Winne & Hadwin, 1998; Zimmerman & Campillo, 2003). Broadly, motivations refer to what moves people to act. Specifically, motivation is defined as “the process whereby goal-directed activities are instigated and sustained” (Schunk, Meece, & Pintrich, 2012, p.5). In this sense, motivations shape student engagement by regulating cognitive and metacognitive learning processes (Butler & Cartier, 2018). Metacognitive learning processes include monitoring one’s current state of knowledge, comparing it to one’s internal standards and controlling learning strategies, focusing on the optimization of learning through a cyclical and dynamical process (Ben-Eliyahu & Bernacki, 2015; Bernacki, 2018). Pragmatically, Sternberg (2017) considers learners’ motivation and metacognitive skills to be the most important components in translating learner’s abilities and skills into achievement.

Although studies have examined relations between motivation, metacognition, learning processes, and outcomes (Bernacki, Byrnes, & Cromley, 2012; Coutinho, 2008; Pellas, 2014; Vrugt & Oort, 2008), a majority of the research has measured target variables by survey and then examined their relations and effects based on a variable-centered approach that is concerned with a population of individuals. The approach aims to investigate (causal) relations between these variables at the group level (Bergman & Trost, 2006). Specifically, the variable-centered approach more focuses on main variable effects so might overlook subtle combined effects of variables. In this sense, there is a fundamental mismatch between holistic context in which existing factors interact with each other and linear models (Bergman & Magnusson, 1997). In
other words, there is no unique variable that is totally inconsistent with other variables. Therefore, methods based on the variable-centered approach have to consider interaction terms to capture the configurations of factors that jointly describe learners’ complex learning processes, sometimes resulting in the unacceptably complicated process of analysis and interpretation (Bauer & Shanahan, 2007). The unacceptably complicated process of analysis and interpretation

During learning, however, students are more likely to be influenced by multiple constructs of motivations (Conley, 2012) and employ several metacognitive strategies to manage their learning (Winne & Hadwin, 1998). Also, each student has the different preferences of motivation and metacognitive learning processes. In this case, a person-centered approach has a capability of appreciating interactions of multiple variables and nonlinear relationships between them (Bauer & Shanahan, 2007). This approach allows researcher to interpret result more easily and better reflect real educational phenomena. Therefore, in order to capture these profile information, it would be more appropriate to employ a person-centered approach in which the purpose is to identify heterogenous groups of individuals and figure out how these groups are related to outcomes such as academic achievement.

Most of the researchers who researched metacognition have used surveys (Wolters & Won, 2018) or think-aloud protocols (Greene, Robertson, & Costa, 2011) to measure metacognition learning processes in SRL models. However, these methods have some disadvantages. Specifically, the act of consciously responding to a survey or reporting one’s thoughts via a think-aloud distracts students from authentically engaging in a task (Biswa, Baker, & Paquette, 2018). Survey approaches suffer further from the timing of their reporting as they are often administered after learning. Students thus must remember their learning process
retrospectively so some of the important information might be lost (Winne & Jamieson-Noel, 2002). The aggregate nature of survey items further threatens the accuracy of reports as students must make summary judgments of their learning, which summarize events over time and context, and preclude more precise analysis of learning processes. Think–aloud protocols avoid these issues, but in addition to their taxing of students through concurrent reporting, they are expensive in terms of time costs to run subjects individually and transcribe accounts. These features preclude data collection at sufficient scale to collect data needed for complex analyses and require that data be collected in lab settings rather than authentic environments in longitudinal fashion (Biswas et al., 2018).

According to Winne and Hadwin (1998), metacognitive learning processes are associated with many other factors such as motivation and students’ characteristics, influencing one another. To fully understand such complicated learning processes, therefore, a cyclical, dynamic and sequenced collection of metacognitive behaviors is required (Bernacki, 2018). Learning management systems (LMS) provide one method to capture and record student learning activities taking place in the system, including attendance, grades, and use of hosted digital resources that are designed to support specific learning processes. The log data generated by the LMS allows researchers to take a closer look at metacognitive learning processes and track them with time-stamped data.

This data-intensive approach provides researchers with opportunities to better understand learning settings via multiple types and grain sizes of data such as input data (e.g., demographic information or motivation), process data (e.g., learning activities), and outcome data (e.g., test scores) collected over long learning tasks (Marsh, Pane, & Hamilton, 2006). The approach is
aligned with systematic data collection, analysis, interpretation and educational implication based on verifiable data (Mandinach, 2012).

Therefore, the goal of this dissertation is to use the affordances of log data to investigate the relationship between motivation and metacognition and their influence on academic achievement using multiple types of data. Further, to examine relations among the data via motivational and metacognitive profiles of the students, methods based on the person-centered approach will be employed. The person-centered approach takes into account the possibility that the sample might reflect multiple subpopulations characteristics by different sets of statistics (Magnusson, 2003). Therefore, the methods allow researchers to identify the best profile structure of multiple motivations and metacognitive learning processes. In addition to the investigation of the relations, how significantly the different types of predictors contribute to academic achievement will be examined, focusing on theory-aligned behavioral data designed based on metacognition components of the SRL model (Winne & Hadwin, 1998).

Specifically, the purpose of the first paper is to examine relations between achievement goals and metacognitive learning behaviors using a clustering analysis and visualization. In the research, achievement goal profiles are identified using the clustering analysis and their relations with metacognitive learning behaviors captured by an LMS are examined by employing time-based visualization and a pair of F-tests. Further, to investigate a more nuanced trace of student metacognitive processes by employing an emerging educational data mining method, sequence pattern mining. A majority of research on achievement goals assumed that people tend to pursue one type of goals motivation, in which students motivated with mastery goals are more likely to demonstrate adaptive outcomes, whereas those with performance are less likely to (Dweck, 1986; Dweck & Leggett, 1988). However, from the 2000s, researchers have focused on
demonstrating the pursuit of multiple achievement goals together, which is called the multiple goal perspective (Barron & Harackiewicz, 2001; Harackiewicz, Barron, & Elliot, 1998; Senko, 2016; Senko & Tropiano, 2016).

For research from the multiple goal perspective, a person-centered approach has some advantages over the variable-centered approach (Levy-Tossman, Kaplan, & Assor, 2007). The variable-centered approach more focuses on main variable effects so might overlook subtle combined effects of variables. Also, analyses based on the variable-centered approach assume the linearity of relation so might mask non-linear effects of variable combinations. Therefore, analyses based on the person-centered approach would be more appropriate methods to capture combined effects of multiple goals students pursue and to reflect their complex relations by each group than variable-centered based ones.

In addition to motivations, another important part of SRL, metacognitive learning processes are dynamic and cyclical with multiple sub-components including monitoring and controlling processes (Winne & Hadwin, 1998), so the processes should be examined based on behavioral patterns to measure the quality beyond the quantity of learning. Therefore, this research produces motivational profiles and investigates how the profiles are related to metacognitive learning patterns.

In the second paper, the process to identify motivational profiles employed in the first paper will be further extended by including multiple types of motivation and be elaborated by employing a latent profile analysis (LPA). Although many studies have examined the profiles of individual motivation such as achievement goals (Levy-Tossman et al., 2007; Litalien & Morin, 2017), a few studies have attempted to figure out motivational profiles based on multiple motivation theories (Conley, 2012). Additionally, to intersect with the emerging motivation
literature that considers multiple theories in the generation of learner profiles, a person-centered approach will be applied to metacognitive behaviors to capture the multidimensional nature of students’ metacognitive monitoring and control processes, using an LPA. Motivation has been found to influence metacognitive learning process to manage learning in the SRL process (Winne & Hadwin, 1998). However, most of the research has investigated relations between individual motivation and metacognitive learning process. Therefore, in this research, for better understanding relationships between motivation and metacognition from the perspective of profiles, multinomial logistic regression will be performed in which metacognition profiles will be used as a categorical dependent variable. Lastly, according to many studies on motivation, demographic information such as age, gender, and ethnicity was found to be significantly associated with motivation (Eccles et al., 1983; Dweck, 1986; Watt, 2004; Nicholls, 1990). Therefore, I will investigate how demographic information influence the membership of motivational profiles by including age, gender, ethnicity, parents’ schooling.

The third paper aims to consider the relative importance of capturing motivational and metacognitive processes as potential predictors of learning outcomes, and appraises them alongside both traditional prediction modeling approaches in higher education, and emergent methods arising from the field of educational data mining. Versions of this model will include demographic information (i.e., the current higher ed institutional approach), motivation (i.e., a predominant educational psychology approach), and data on student learning behavior (i.e., the emerging learning analytics approach). Accordingly, I intend to compare the relative benefits of collecting each variety for the purposes of informing a prediction model, and then to demonstrate further how the influence of theory-aligned feature design can divide behavioral data into theoretically aligned groups and improve their predictive power. Lastly, behavioral patterns
discovered through a sequence pattern mining technique beyond frequency will be included in
the model.

This dissertation provides insight regarding authentic relations between motivation,
metacognition, and academic achievement. Specifically, instructors can understand how multiple
types of motivations work together and the motivational profiles influence metacognitive
learning strategies. In courses, by examining motivational profiles, instructors can provide more
effective intervention with which students change their resolve their weak learning easier.
Additionally, through metacognitive behavioral patterns, students and instructors can improve
their understanding of metacognitive learning process. In this sense, in terms of the prediction
model, behavioral patterns might have more predictive power above the frequency of access to
contents.

Practically, by investigating each type of predictor from data resources including
demographic, motivation, and behavioral variables, findings from this dissertation can enable
researchers to prioritize development of prediction models to identify students who are more
likely to experience failure in courses. Additionally, instructors can figure out the importance of
interpreting variables through educational theories and in context through the comparison of
courses with differing instructional designs. Further, by appraising these results in light of
theory, instructors can take action to improve student’s learning outcomes by adjusting the
design of their courses.

In summary, study 1 and 2 focus on relations between motivation and metacognition
from a theoretical perspective. Previously, many studies examined achievement goals
individually based on the variable-centered approach using survey. In response, I identified
motivational profiles in Study 1 based on the person-centered approach by employing clustering
analysis with three types of achievement goals. According to Zimmerman (2012), multiple types of motivation work simultaneously in the SRL process; Study 2 extends Study 1 by applying a more objective, latent method to multiple motivation types. Therefore, based on the person-centered approach, clustering analysis and latent profile analysis can identify motivational profile and subsequently, I examined their relations with metacognition and academic achievement. Through the findings of Studies 1 and 2, instructors can gain insight into relations between motivation and metacognition in authentic contexts. In Study 3, I take a more practical perspective, and study the relative and combined predictive power of data resources that can predict achievement. Despite the availability of many types of data at university, there have been few studies that predict academic achievement using multiple sources. In developing prediction models, educational theories and contexts are largely ignored (Baker & Yacef, 2009). Therefore, by appraising traces of behaviors as they reflect events described in educational theories like metacognition and self-regulated learning, I demonstrate the importance of educational theory as a resource for improving the tracing of learning and accuracy of models that predict achievement.
Chapter 2: Examining the Influence of Undergraduates’ Achievement Goals on Metacognitive Behavior Sequences, and Achievement in Science

1. Introduction

The achievement goals that students hold for learning are theorized to guide their engagement in learning tasks and to have direct implications for both learning and behavior. (Graham & Weiner, 2011; Pintrich, 2003). Students’ goals have been shown to influence strategy value belief (Nolen & Haladyna, 1990), self-regulated learning (SRL) strategies (Ablard & Lipschultz, 1998; Bernacki et al., 2012), deep learning (Dupeyrat & Mariné, 2005; Nolen, 2003; Phan, 2010), cognitive engagement (Greene & Miller, 1996; Walker & Greene, 2009), and metacognitive strategies (Vrugt & Oort, 2008; Wolters, 2004; Wolters, Yu, & Pintrich, 1996).

According to achievement goal theory (Senko, 2016), students engage in a learning task, providing meanings to their behaviors. Students either wish to develop their competence and maximize their potential, or to demonstrate their competence and prove something to themselves. In the theory, achievement goals represent the purpose of or reasons, which consists of mastery- and performance goals according to the standard of competence (Elliot & McGregor, 2001). Students with strong mastery goals are theorized to engage in deeper cognitive and metacognitive learning strategies such as self-monitoring (Pintrich, 1999), while those with performance avoidance goals are theorized to tend towards shallower strategies such as surface processing (Coutinho & Neuman, 2008), and performance-approach oriented students’ tendencies are theorized to mix strategies (Elliot & Moller, 2003).

Much of prior research has examined achievement goals individually (i.e., mastery approach and avoidance, performance approach and avoidance; Coutinho & Neuman, 2008), but it is more typical for people to pursue multiple achievement goals together (Barron &
Harackiewicz, 2001; Harackiewicz, Barron, & Elliot, 1998). Originally, achievement theory held that different type of achievement goals separately influence a student’s achievement (Dweck, 1986). Specifically, mastery goals facilitate adaptive learning and health orientation through positive interest and strategies, whereas performance goals are considered to be concerned with maladaptive analysis. However, many studies revealed that performance goals better predicted academic achievement than mastery goals (Senko & Harackiewicz, 2005). This new pattern leads to the emergence of a multiple goal perspective in which mastery and performance goals can be beneficial together in their own way. This approach counters the mastery goal perspective.

Whereas research on achievement goals tends to focus on individual goals as guiding the learning process, overlooking motivational profiles characterized by goals, studies on metacognition tend to subsume a diversity of metacognitive processes under the more general rubric of metacognition, ignoring specific learning processes. Researchers typically assess metacognitive processes through reasonably brief self-report scales that ask general questions about students’ general tendency to engage in metacognitive processes and the frequency (Coutinho & Neuman, 2008; Ford, Smith, Weissbein, Gully, & Salas, 1998). Students tend to respond with limited precision to these instruments (Winne & Jamieson-Noel, 2003), and the metacognitive processes they report tend to be insufficiently precise to represent the many varieties of metacognitive monitoring – of progress towards goals, task performance, and judgments of learning, among others – and the appropriateness of their control strategies (Winne & Hadwin, 1998, 2008; Winne, 2010, 2011, 2018).

When learning is observed in ecologically valid contexts like university courses, the learning technologies available to students allow for timely, contextual assessment of learning goals through embedded questionnaires (Bernacki, Nokes-Malach, & Aleven, 2013), as well as unobtrusive assessment of learning processes (Aleven, Roll, McLaren, & Koedinger, 2010).
When instructors populate course sites on technology platforms like learning management systems, they build complex learning environments that afford students sufficient opportunity to access learning materials and to self-regulate their learning as they pursue an achievement goal. Over the course of a semester-long course where students complete many assignments and pursue mastery of a course’s learning objectives, a huge amount of data is generated in a learning management system (LMS) as it traces students’ use of learning resources. The application of these advanced learning technologies to educational tasks makes it possible for researchers to gain insight into student learning processes, and affords inquiry guided by both theory-driven and data-driven approaches such as cluster analysis of goal complex, and learning analytics (LA) and educational data mining (EDM) processes to understand metacognitive behavior (Gašević, Dawson, & Siemens, 2015; Papamitsiou & Economides, 2014). EDM is concerned with “developing, researching, and applying computerized methods to detect patterns in large collections of educational data that would otherwise be hard or impossible to analyze due to the enormous volume of data within which they exist” (Romero & Ventura, 2013, p. 12). According to the definitions introduced during the 1st International Conference on Learning Analytics and Knowledge (LAK), LA is “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and environments in which it occurs” (https://tekri.athabascau.ca/analytics/).

In this study, we revisit a commonly studied topic – the relations between achievement goals, metacognition, and achievement – but here leverage theory and analytics as tools to more validly represent students’ goal complexity and more extensively trace the metacognitive processes that goal complexes predict. We examine these relations in the context of a biology
course’s learning management system site and further explore how achievement goals and traces of metacognitive processes relate to students’ achievement in their college biology course.

Our specific aims of this study are to 1) employ cluster analysis, a common data-driven method used in motivation research, to reduce our data in ways that align to achievement goal theory (i.e., Barron & Harackiewicz, 2001) and represent students’ achievement goal complexes, and 2) employ a learning analytics approach to observe the specific metacognitive monitoring students conduct via their traces in the LMS. Once goal complexes are identified and students’ metacognitive processes were traced with precision, we then 3) examined how goal complexes related to metacognitive processes, and how processes align to levels of performance in the course. A final aim was to 4) employ an emerging, educational data mining method – sequence pattern mining – to develop a more nuanced trace of students’ metacognitive processes during biology learning and to explore its relations to goals and achievement.

2. Theoretical background

A. Achievement Goal Theory

An achievement goal theory explains students’ purposes or reasons for engaging in a learning task (Senko, 2016). Achievement goals theory started with a dichotomous model that was developed from the research in the late 1970s and early 1980s, where two qualitatively distinct goal types - mastery and performance - for achievement behavior were distinguished (Maehr & Nicholls, 1980). These goals were differentiated according to their focus on competence (Dweck, 1986). The former is concerned with intrapersonal values that aim to develop competence and task mastery, whereas the latter is concerned with differences from others, demonstrating competence relative to others (Elliot & Hulleman, 2017).
In the initial stage of the achievement goal theory, a mastery goal perspective is prevalent, in which mastery goals are considered superior to performance goals (Elliot & Dweck, 1988). Specifically, mastery goals lead to positive and adaptive process and outcomes, whereas performance goals relatively tend to produce negative and maladaptive results (Nicholls, 1984). Furthermore, according to Dweck and Leggett (1988), mastery goals align with a belief in which, intelligence is malleable and increased through effort, which is referred to as the *incremental theory of intelligence*. In contrast, performance goals are explained by a belief that intelligence is a fixed and uncontrolled trait, which is called *entity theory of intelligence*. Therefore, students with the stable concept of intelligence are more likely to focus on how performance is evaluated and how they compare with others while students holding incremental theory have more interests in self-assessment and self-improvement (Schunk et al., 2012).

Later, Elliot and Harackiewicz (1996) proposed the trichotomous achievement goal model by integrating the approach-avoidance distinction within performance goals, resulting in performance approach and performance avoidance. Whereas performance approach is theorized regarding striving to outperform others, performance-avoidance is theorized regarding striving to avoid appearing incompetent relative to others (VandeWalle, 1997). The incorporation of the approach-avoidance distinction into the previous dichotomous model makes it possible to explain why performance goals produced relatively inconsistent findings (Elliot & Hulleman, 2017). Mostly, approach version demonstrated more positive results than avoidant version (Vrugt & Oort, 2008; Wolters, 2004).

Finally, the fully crossed $2 \times 2$ model was proposed by applying the valence of approach and avoidance to mastery goals, resulting in mastery approach and mastery avoidance (Elliot & McGregor, 2001; see Table 1). Mastery approach is not different from mastery goals in
dichotomous and trichotomous models as previously this was portrayed in a positive light. In contrast, mastery avoidance was conceptualized in terms of striving to avoid misunderstanding or failure to master knowledge (Elliot, 1999).

Table 1. Achievement goal framework (Elliot & McGregor, 2001)

<table>
<thead>
<tr>
<th></th>
<th>Task-based or Intrapersonal</th>
<th>Normative and relative to others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desire for success</td>
<td>Mastery approach</td>
<td>Performance approach</td>
</tr>
<tr>
<td>Desire for avoiding failure</td>
<td>Mastery avoidance</td>
<td>Performance avoidance</td>
</tr>
</tbody>
</table>

Specifically, mastery approach focuses on the desire to learn, whereas mastery avoidance strives to avoid learning failure. In contrast, performance approach-oriented students desire to outperform others, and performance avoidance-oriented students tend to avoid appearing less talented by performing poorly relative to their peers (Senko, Hulleman, & Harackiewicz, 2011). According to Dweck (1986), however, compared with other achievement goals, mastery-avoidance has been relatively uncommon in previous literature (Ciani & Sheldon, 2010; Graham & Weiner, 2011), so the research focused on mastery approach, performance-approach, and performance-avoidance (Elliot & McGregor, 2001).

For a long time, within distinctions between mastery and performance goals, mastery goal theorists thought mastery and performance goals are incompatible with each other as they have opposing striving, so an increase in one goal should lead to a decrease in another goal (Ames, 1992; Darnon, Dompnier, Gilliéron, & Butera, 2010). However, many previous studies demonstrated that mastery and performance goals are positively correlated (Barron & Harackiewicz, 2001). Also, findings from studies on achievement goals theory sometimes showed performance goals were more influential on academic achievement (Harackiewicz et al., 1998) as well as task interest (Senko & Harackiewicz, 2005) than mastery goals. Even, in spite
of a number of studies showing the separation of performance approach and performance avoidance goals in factor analysis (Conroy, Elliot, & Hofer, 2003; Elliot & Church, 1997), both types of goals tend to exhibit strong positive correlation, which is opposed to the theory (Elliot & Murayama, 2008). As a result, multiple goals emerged to explain these unexpected results (Barron & Harackiewicz, 2001; DeShon, Kozlowski, Schmidt, Milner, & Wiechmann, 2004; Harackiewicz et al., 1998). This new perspective holds that people not only tend to pursue multiple goals together, but also can handle them effectively to attain benefits from each goal, which is a counter to the previous perspective such as mastery goal perspective (Senko, 2016).

Based on this multiple goal perspective, many researchers who study achievement goals have had interests in goal profiles that hold mastery and performance goals simultaneously to varying degree (Harackiewicz, Barron, Pintrich, Elliot, & Thrash, 2002). Barron and Harackiewicz (2001) pointed out the four patterns of interactions between mastery and performance goals result in advantages of the pursuit of multiple goals. In an additive goal pattern, mastery and performance goals have an independently positive effect on outcomes. In an interactive goal pattern, beyond the independent main effects, these goals have an interactive effect on outcomes. In other words, the level of performance goals might be associated with the effect of mastery goals on outcomes. In a specialized goal pattern, mastery and performance goals have an effect on different outcomes. For example, mastery goals might have a positive effect on persistence, whereas performance goals might predict final scores, but not persistence. Lastly, in a selective goal pattern, students are able to select a particular achievement goal depending on situation. For example, a student holds dominant mastery goals when preparing for an exam, but he or she might adopt performance goals when taking an actual exam. Therefore, students who pursue this achievement goal pattern might be able to demonstrate the best
performance as they shift between multiple goals to find the most relevant goal to a particular task.

In general, profiles that include mastery goals (i.e., dominant mastery goals or multiple goals holding both mastery and performance-approach goals) are shown to be more associated with adaptive learning than profiles that do not include dominant mastery goals (Dweck, 1988). In addition, students holding profiles including dominant performance goals demonstrate better performance than students holding the low level of both goals (Pintrich, 2000a; Levy-Tossman et al., 2007; Tuominen-Soini, Salmela-Aro, & Niemivirta, 2008). The effect of mastery goals is similar to a mastery oriented approach in which mastery goals adopt adaptive learning (Elliot & Dweck, 1988), while the role of performance goals in the studies on goal profiles was found to be different from that asumeed in the normative goal theory, demonstrating positive influence on outcomes. Methodologically, with growing interests in multiple goals perspective, cluster analysis and latent profile analysis (LPA) has been increasingly employed to figure out what achievement goal profiles leaners holds and how they are associated with outcomes such as academic achievement (Levy-Tossman et al., 2007; Madjar, Weinstock, & Kaplan, 2017; Pastor, Barron, Miller, & Davis, 2007; Tuominen-Soini, Salmela-Aro, & Niemivirta, 2008; Wang, Morin, Liu, & Chian, 2016). Tuominen-Soini and colleagues (2008) identified five types of achievement goals: 1) mastery-intrinsic, 2) mastery-extrinsic, 3) performance-approach, 4) performance-avoidance, and 5) avoidance. The result of the analysis showed six profiles fit the data best: 1) indifferent, 2) mastery-oriented, (3) success-oriented, (4) performance-oriented, (5) disengaged, and (6) avoidance-oriented. Regarding the relation of profiles with academic achievement (e.g., GPA), the success-oriented profile which holds the high level of mastery and performance goals demonstrated the highest GPA score. In addition, students in the
performance-oriented profile outperformed those in indifferent, disengaged, and avoidance-oriented profiles, showing that performance-approach goals positively influence academic achievement than less-goal orientation or avoidance orientation.

B. Metacognitive monitoring

Metacognition started from the research on metaprocesses in the early 1970s (Flavell, 1971). In general, metacognition is referred to as “thinking about thinking” or “knowing about knowing” (Papleontiou-louca, 2003). The role of metacognition is to optimize one’s cognitive actions in pursuit of learning goals (Griffin, Wiley, & Salas, 2013). In particular, metacognitive monitoring and metacognitive control are the main components of SRL (Winne & Hadwin, 1998, 2008; Winne, 2010, 2011; Zimmerman, 2008). Metacognitive monitoring refers to learners’ awareness of their learning (Pieschl, 2009; Winne & Hadwin, 1998). Metacognitive control refers to a process in which the present states of the learning activities are adjusted according to the products of metacognitive monitoring (Nelson, 1996).

According to Winne and Hadwin (1998), SRL consists of four loosely sequential and recursive phases: 1) defining the task, 2) goal setting and planning, 3) enacting study tactics and strategies, and 4) metacognitively adapting studying for the future. Throughout the first three phases, students who are skilled at self-regulating their leaning monitor information about how learning tactics and strategies are used, and the fit of internal and external conditions (Winne, 2018). In the last phase, learners make a substantial decision to change their learning tactic and strategies for the future task, which is achieved through at least three ways: revising goals, adapting plan or changing operations (Winne, 2010). Like this, the main components are metacognitive monitoring and control.
Nelson and Narens (1990) suggested two types of information, object-level, and meta-level information, in the model of metacognition. Information at the object-level is concerned with products of each SRL phase, including the definition of the task, goals and plans, tactics and strategies, and adaptations, whereas meta-level information consists of meta-information of object-level and cognitive operations that lead to change in object-level elements, which is called metacognitive control. In other words, the object-level information is associated with the products of learning and the meta-level information constitutes learner’s standard for the product (Winne & Hadwin, 1998).

Therefore, metacognitive monitoring is explained as a process where discrepancies between object-level (learning enacted) and meta-level information (learner’s standard) are identified. That is, the primary purpose of monitoring is to identify whether learning achievement corresponds to students’ existing standards that are criteria against which products they created are monitored, resulting in the cognitive evaluation (Winne, 2010; Winne & Hadwin, 1998). Based on the product made in the previous process, metacognitive monitoring, existing learning tactics or strategies are adjusted by comparison to student’s standard for learning, which is metacognitive control (Winne & Hadwin, 1998).

To date, many empirical studies on metacognitive monitoring have relied on self-report data (Griffin et al., 2013; Mokhtari & Reichard, 2002; Pintrich, Wolters, & Baxter, 2000). Self-report has the advantage of capturing student’s intention for the use of metacognition tactics and strategies (Pintrich, Smith, García, & McKeachie, 1993), but this type of data collection did not provide reliability enough for the research (Winne, 2005). The main reason for this issue is because learners cannot assess their learning process correctly while engaging in learning task (Winne, 2018; Winne & Jamieson-Noel, 2002). Also, another issue is that most of the self-report
is designed to assess general study tactics and strategies, but not focusing on specific contexts for which no item is assigned or at best just a few items within a large scale (Mokhtari & Reichard, 2002; Moore, Zabrucky, & Commander, 1997). Therefore, some researchers have criticized that such biased self-reports lead to inaccurate results and have limits in measuring actual learning behaviors (Cromley & Azevedo, 2006; Winne, 2005). In particular, the self-report method is vulnerable to analyzing temporal, sequential, or contextual features (Wolters & Won, 2018).

With regard to these issues, according to Bernacki (2018), time, granularity, and context should be taken into account to capture cognitive and metacognitive learning events in SRL processes fully. Firstly, the learning process in the SRL model is temporal. (Winne & Hadwin, 1998, 2008; Zimmerman, 2000a, 2008). Therefore, each learning event should be captured in combination with a time stamp. In addition, the log data stored temporally makes it possible to control granularity mentioned below. Next, the level of granularity can be varied depending on the unit of timeframe combined with learning event. For example, the data captured with a unit of seconds can be examined at a more fine-grained level than those with minutes. Also, according to research focuses, each learning event is restructured to represent higher-order learning processes by aggregating them by learning goal or feature. Lastly, contexts help to figure out why a learner does a certain action. In other words, contexts can be understood by a prior event that is examined and interpreted with log data in combination with a time stamp. Although most researchers agree with the importance of these factors, self-reports cannot exploit these advantages.

However, technology-based learning systems such as LMSs allows for more than the distribution of materials. Specifically, LMSs can capture most of the interactions among learners, instructors, and environments through the distributed form of learning activities, making it
possible to investigate student’s learning process in depth (Dabbagh, & Bannan-Ritland, 2005). In particular, LMS features as metacognitive tools to support SRL goal setting, self-monitoring, and time management would be used to test how students metacognitively engage in the learning process (Dabbagh & Kitsantas, 2013). Above all, by making it possible to capture learning events with time stamps, researchers are able to investigate cognitive and metacognitive learning process at divergent angles (Bernacki, 2018).

Like self-reports, the log data from LMS might have validation issues of whether the tool correctly reflect particular learning behaviors. This process allows researchers to obtain reliable data and leads to accurate analytical results with appropriate statistical power. Researchers are able to validate the log data by the time-stamped metadata based on a theory (Bernacki, 2018). For example, if the monitoring learning tools more frequently used before exams and monitoring performance tools used both right before and after exams, according to metacognition theory (Winne & Hadwin, 1998), it could be said that the tools trace particular learning behaviors accurately. Therefore, in this research the log data are validated through the visualization of temporal monitoring data. For more accurate validation, it is needed to collect additional data from students to supplement the log data (Bernacki, 2018).

In many SRL models, motivation and metacognitive learning processes are considered as important components of self-regulate learning (Butler & Cartier, 2018; Winne & Hadwin, 1998; Zimmerman & Campillo, 2003). This relation between motivation and metacognition was developed based on some assumptions about leaners. According to Pintrich (2000b), many SRL models assume that students actively can construct their goals and potentially monitor, control, and regulate their learning. In addition, the goals are used to determine whether the learning processes continue, or some changes are necessary. Further, the process of self-regulation based
on motivation and metacognition learning process and mediate between a learner’s contextual characteristics and academic achievement.

In this sense, learner’s goal orientation is theorized to influence how individuals plan, evaluate, and engage in the achievement-relevant task, in that students set goals that they want to achieve and develop standards that they want to meet, and in turn, monitor their learning process toward the goals or standards (Dweck, 1986; Winne & Hadwin, 1998). That is, goal setting is placed in front in the SRL process and influence the overall model (Zimmerman, 2002). Therefore, it is not surprising that setting appropriate goals is one of the most critical components for successful self-regulated SRL.

Achievement goals theory that explains the reasons why students pursue a specific task allows for the inclusion of motivation in SRL models by demonstrating how they are motivated to use learning strategies to master learning materials or outperform others (Schunk et al., 2012). Specifically, students motivated with mastery goals are more likely to search for metacognitive strategies to master a specific learning task and further self-regulate their learning process for learning progress (Ames, 1992). In contrast, students who have performance goals do not view learning as the purpose, but as the process to achieve their goals, specifically demonstrate their competence by outperforming others or succeeding with less effort (Senko, 2016). Therefore, the students focus on more performance evaluation than engagement in the task, which results in less interest in self-regulating learning by metacognitive strategies.

Given the interest in the roles of motivation and metacognition in the SRL process, there have been many studies on the relationships between achievement goals and metacognition (Vrugt & Oort, 2008; Wolters, 2004; Wolters, Yu, & Pintrich, 1996). Although most of the research showed that the mastery approach positively predicts metacognitive processes (e.g.,
Bernacki et al., 2012), the influence of performance approach was inconsistent (Coutinho, 2008; Coutinho & Neuman, 2008; Ford et al., 1998). Such prior investigations make use of a mix of self-reported and observed learning behaviors, which further limits the interpretations that can be made of the role that achievement goals have in metacognitive processes. We, therefore, examine the effect of achievement goals on metacognitive processes using rich logs of learning behaviors that can reflect monitoring of learning and performance. We further sequence individual learning events into patterns of metacognitive behaviors to understand how the type, frequency, and complexity of metacognitive processes might differ by the achievement goals that guide learning, and the level of achievement that results.

Specifically, according to a recent study (Bannert, Reimann, & Sonnenberg, 2014), it is more recommended to examine SRL processes more based on the event than trait or attitude. Increasingly, this interest leads to a focus on behavioral data and further, the advancement of technology allows for the availability of temporal data. Furthermore, recent research in SRL that emphasizes learning patterns as indicators of the quality of learning rather than using the frequency of accessing particular features (Bannert et al., 2014; Taub, Azevedo, Bouchet, & Khosravifar, 2014).

Four research questions guided our exploration:

1) Achievement goal profiles
   a. What achievement goal profiles emerge when students report their goals for an early undergraduate life science lecture course?
   b. How does academic achievement differ across motivational profiles obtained?

2) How differently do achievement goal profiles influence their metacognitive monitoring (learning via self-assessment, performance via checking grades)?
3) How much students’ metacognitive monitoring (learning via self-assessment, performance via checking grades) influence academic achievement?

4) Metacognitive behavioral sequences
   a. What sequences of metacognitive events emerge when logs are mined?
   b. How do these complex sequences of metacognitive processes influence academic achievement?

3. Method
   A. Participants
      We observed the data of 377 students from a large Southwestern university where ethnically diverse students study. They enrolled in a large face-to-face biology course in the 2015 spring semester, which was designed to provide fundamental knowledge needed for continued health-science education. 75% of the sample were females, and 42% were from an underrepresented minority group (26% Caucasian, 32% Asian, 22% Hispanic, 6% African American, and 16% Others).

   B. Measures
      At the beginning of the course, students were surveyed initially on the LMS to assess their achievement goals using 9-item Achievement Goals Questionnaire-Revised (AGQ-R) (Elliot & Murayama, 2008). Although the original version was comprised of 12 items for four sub-domains, three items for mastery-avoidance were excluded in the research as mastery orientation tend to show a similar influence regardless of the level of perceived competence (Elliot & Harackiewicz, 1996). For all items, students’ responses range from strongly disagree (1) to strongly agree (7). The internal consistency (Cronbach’s α) of the survey was .800 for mastery-approach (e.g., “My aim is to completely master the material presented in this
course”), .766 for performance-approach (e.g., “I am striving to do well compared to other students”), and .836 for performance-avoidance (e.g., “My goal is to avoid performing poorly compared to others”). Some students did not complete the survey (i.e., less than 1% missing data for each item), and I handled them by using mean values for the further analysis (e.g., clustering analysis).

In this research, for academic achievement, the scores of 4 exams were used. Students had exams in week 6, 10, and 14, and the last exam was provided in week 17 as a comprehensive exam. Each lecture exam consisted of a combination of question formats including multiple-choice, fill in the blank, short answer, and essay questions. The quizzes were administered almost every week throughout the semester.

Blackboard Learn, an LMS used in the university, captured all learning behaviors of students through learning-support tools provided by instructors in the course and then stored them in the log files of the LMS database. For metacognitive processes, monitoring learning took place through self-assessment and retrieval practice in ungraded practice quizzes. Monitoring refers to a process that confirms whether the cognitive products correspond to standards and goals students set at the initial phase of the SRL process (Winne, 2018). Therefore, students monitor their learning by comparing the quiz score and their goals until the result of the assessment is satisfactory, meaning more monitoring processes are not needed. “My Grades” allowed participants to monitor their performance by providing present scores obtained in the course. Similar to self-assessment quizzes, if the grades shown in My Grade do not meet student expectations, students will perform more learning behaviors to achieve their goals. Lastly, Splunk, data management software, enriched the log data generated by the LMS with metadata to
identify learning events (Bernacki, 2018). The software allowed researchers to extract data in flexible formats with various levels of timestamps using Splunk search language.

C. Procedure

K-means clustering analysis was performed as the preliminary cluster solution through an iterative process by minimizing the within-cluster variance and by maximizing the between-cluster variance to identify achievement goal profiles held by students in the course. During the analytical process, each object is assigned to center that is nearest by calculating their distance based on Euclidean distance. The data used for the analysis included the means of mastery approach, performance approach, and performance avoidance. SPSS 23 was used to perform the K-means clustering analysis with 20 iterations and the result of clustering membership was saved on the same file for the further analysis. One issue of clustering analysis is the lack of criteria for determining the best number of clusters (Hair, Black, Babin, Anderson, & Tatham, 2006). To determine the best number of clusters, the primary criterion is parsimony in which the smaller number of clusters is considered the better result. Additionally, multiple complementary methods were used to figure out how many clusters emerge. Firstly, an elbow chart shows a typical plot of an error measure, the within-cluster variation, and provides a point that decreases abruptly relative to previous changes in the slope, that is more likely to be the ideal number of clusters (Kassambara, 2017). Additionally, as a complementary method to the elbow chart, this study employed Gap statistic method, in which the total within-cluster variation for the different number of clusters is compared with their expected values under reference null distribution (Tibshirani, Walther, & Hastie, 2001). Secondly, the result is verified by a theory. In other words, the result of the clustering analysis was interpreted based on achievement goals theory to see if the sample was properly grouped and represent multiple goals well.
Lastly, the number of members in each cluster was examined. A cluster with too small sample could not be used for further analyses, and therefore, the sample should be examined in depth or excluded in the research.

Then, we examined how achievement goal profiles influence score on exams and quizzes using visualization and multivariate analysis of variance (MANOVA). Then, for each set of exams and quizzes, MANOVAs are conducted to examine how achievement goal profiles influence academic achievement throughout the semester. To figure out the relation of profiles and metacognitive learning strategies, use of self-assessment quizzes and My grades was tracked. To look at the effect of achievement goal profiles on metacognitive learning process across the semester, a pair of MANOVAs was conducted with subsequent F-tests to examine the effect over 16 weeks. Lastly, to investigate metacognitive leaning processes in detail, sequential pattern mining was implemented. For this process, the log data were pivoted to sequence events within a session using the syntax “FOR XML PATH” in MS-SQL, producing from the log blocked sequences of learning events per learner and session as shown in Figure 1.

![Figure 1. Data transformation for pattern mining](image)

In addition, we investigated the use of the metacognitive strategy found in the sequential pattern mining process and examined the effect of the strategy on all exams.
4. Results

A. Achievement Goal Profiles

![Image of elbow chart and gap statistic method](image_url)

Figure 2. Elbow Chart (left) and Gap Statistic Method (right)

The elbow chart and the result of Gap statistic methods are illustrated in Figure 2. The elbow plot demonstrated that the slope decreases from 3 and gets flat as the number of clusters increases. However, to determine the optimal number of clusters with more objective approaches through multiple methods, Gap statistic method was performed. The result of the method revealed that three clusters was found to be the optimal number of clusters.

Using scores on AGQ-R items, a $k$-means cluster analysis was conducted to identify the achievement goal profiles commonly adopted by students in the course. According to the results of elbow chart and Gap statistic method, a clustering analysis with three clusters was performed using SPSS 23. Each achievement goal profile represents 1) high mastery-approach and performance-approach endorsement ($n=142$), 2) high mastery goals compared to performance-approach and performance-avoidance goals ($n=40$), and 3) high endorsement of all goals ($n=195$). The mean scores for each item students obtained are illustrated in Figure 2.
Figure 3. Scores by Achievement Goal Cluster. Red outline regions indicate discriminating features of clusters who differ in their endorsement of achievement goal items by subscales for Performance Approach and Performance Avoidance constructs. (MAP: mastery-approach; PAP: performance-approach; PAV: performance-avoidance).

Table 2 shows descriptive statistics including means and standard deviations of each cluster and homogeneous groups identified by a series of ANOVAs with a post-hoc analysis. The High Goal Endorsement group has the greatest value of all achievement goals. In addition, the High Approach group had greater performance approach and performance avoidance than the High Mastery group, but mastery approach was similar in both of groups.

Table 2. Descriptive statistics and group comparison

<table>
<thead>
<tr>
<th></th>
<th>High Approach</th>
<th>High Mastery</th>
<th>High Goal Endorsement</th>
<th>Homogeneous clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Mastery Approach</td>
<td>6.53</td>
<td>0.63</td>
<td>6.53</td>
<td>0.78</td>
</tr>
<tr>
<td>Performance Approach</td>
<td>5.77</td>
<td>0.69</td>
<td>3.47</td>
<td>1.05</td>
</tr>
<tr>
<td>Performance Avoidance</td>
<td>4.63</td>
<td>0.92</td>
<td>2.84</td>
<td>1.29</td>
</tr>
</tbody>
</table>

*p<.001.
Note. For Homogeneous clusters, High Approach (1), High Mastery (2), High Goal Endorsement (3)

Figure 4 shows academic performance by achievement goal profiles through changes in scores of exams and quizzes throughout the semester. For both of exams and quizzes, mastery approach-oriented students performed better than other two groups, High approach and High goal endorsement group. In contrast, students in the High goals endorsement group obtained the
lowest scores for all exams and quizzes. In particular, for the first and final exams, the High mastery group outperformed the High goals endorsement group by five points.

To check whether the achievement goal profile significantly influences academic achievement across exams and quizzes after controlling for classification error, a multivariate analysis of covariance (MANCOVA) was carried out for exams scores. To accommodate classification error, distances from cluster centers were used as covariates in the analysis. The result demonstrated that the profile did not have a significant effect on exams across the semester, with $F(8, 742)=1.281$, $p>.05$, Wilk's $\Lambda=.973$, partial $\eta^2=.014$. However, subsequent analysis of variance (ANCOVA) for each exam revealed that the high mastery group significantly outperformed other groups at the beginning and end of the semester, showing $F(2, 373)=3.099$, $p<.05$, partial $\eta^2=.016$ for Exam_1, and $F(2, 373)=3.755$, $p<.05$, partial $\eta^2=.020$ for exam 4. Specifically, the result of a contrast analysis demonstrated students in the High-mastery group ($M=83.7$, $SD=13.9$ for Exam_1; $M=80.3$, $SD=14.3$ for Exam_4) showed significantly higher performance than those in high-goal endorsement group ($M=77.3$, $SD=12.6$ for Exam_1; $M=73.0$, $SD=14.8$ for Exam_4). However, the high-approach group ($M=79.2$, $SD=12.6$ for Exam_1; $M=76.0$, $SD=12.9$ for Exam_4) did not show any significant different performance from other groups.
B. Exploring Metacognitive monitoring behaviors

In addition to scores on exam and quizzes, use of self-assessment quizzes and checking grades was observed to examine metacognitive monitoring behaviors by achievement goal profiles.

I conducted an analogous visualization and analysis of metacognitive monitoring behaviors of students across achievement goal groups to examine these same behaviors as conducted by students with differing goal profiles. Students’ use of ungraded online quizzes were examined as a tool used for self-assessment and retrieval practice (i.e., metacognitive monitoring and control; Winne & Hadwin, 1998).

Visual evidence indicates that students in the High Mastery group used these quizzes more frequently overall, with particularly greater use in the week immediately prior to the first and second exam and the week of the final exam. For the first and second exam, mastery-oriented students also monitored their learning earlier than those in other two groups who more actively used the quizzes in the week of exams than students with high mastery approach. In contrast, in terms of checking grades, students in high goals endorsement were more apt to seek
feedback on their performance after exams whereas, mastery-oriented students less frequently accessed to the monitoring performance tool, which is in opposition to the self-assessment tool (Figure 5).

Then, I analyzed the data statistically. In general, however, the distribution of learning behaviors taking place in online systems tends not to be normal, where some students rarely use the system, resulting in highly skewed distribution. Therefore, it is not appropriate to analyze the data using ordinary least squares (OLS) techniques so, in this research, the variables of monitoring behavior were analyzed as count (Bernacki et al., 2012; Greene, Costa, & Dellinger, 2011). Additionally, to handle classification error, distances from cluster centers were analyzed as a covariate, and three achievement goal profiles were included using dummy coding with two variables in the analysis. The results of Poisson and negative binomial regression analyses with metacognitive learning behaviors were compared to determine the better regression analysis. Model fit indices of AIC, BIC, and SABIC were used to compare the quality of regression analyses depending distribution and negative binomial regression analyses produced the lower values of them, indicating the more appropriate distribution of monitoring behaviors.

The result of the regression revealed that in terms of self-assessment quizzes, students in the high-mastery group significantly more used the tool than those in the high goal endorsement group in week 9 and high-approach group in week 1. In contrast, students in the high-approach group showed statistically greater use of self-assessment quizzes than those in the high goal endorsement group in week 9 and those in the high mastery group in week 6, 11, and 16. Additionally, students with high goal endorsement also showed statistically more use of self-assessment quizzes than high mastery in week 6 and 16. Therefore, it was statistically confirmed that students motivated by high mastery approach prepared earlier than those in the high goal
endorsement group for the exam 2 through self-assessment quizzes whereas, students in high-approach and high goal endorsement demonstrate more self-assessment than those with high-mastery in the week of exam 1.

![Monitoring Learning and Self-Assessment](image)

**Figure 5. Use of monitoring self-assessment and performance by achievement goal profile.** Frequency indicates the number of accesses of web-hosted self-assessment quizzes (top) or the number of visits to a table showing performance on scored assignments in the course (bottom). Red lines indicate weeks of course exams.

Next, how metacognitive monitoring behaviors influence exam scores was examined by employing a series of multiple regression analyses for each exam. For the regression analyses, frequency of using self-assessment quizzes and checking grades before the exams was used as independent variables and exam score as a dependent variable (see Table 3). For exam 1, self-assessment in week 2 and checking grades in around exam 6 were found to be significant predictors of the exam score. For exam 2, self-assessment through quizzes one week before the exam influence positively the exam score. For exam 3 and 4, however, preparation through self-assessment quizzes the for exams in the week of the exams was positively related to exam
scores. Additionally, early checking grades one week before the exam 3 and 4 influence exam scores negatively.

Table 3. Result of multiple regression analyses of monitoring behaviors and exam scores

<table>
<thead>
<tr>
<th>Exam</th>
<th>Monitoring Behaviors</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>Exam_1</td>
<td>Self-assessment in week 2</td>
<td>0.260</td>
<td>0.115</td>
<td>0.123</td>
<td>2.266</td>
</tr>
<tr>
<td></td>
<td>Checking grades in week 6</td>
<td>0.538</td>
<td>0.242</td>
<td>0.135</td>
<td>2.223</td>
</tr>
<tr>
<td>Exam_2</td>
<td>Self-assessment in week 9</td>
<td>0.145</td>
<td>0.034</td>
<td>0.228</td>
<td>4.315</td>
</tr>
<tr>
<td></td>
<td>Checking grades in week 13</td>
<td>-1.288</td>
<td>0.577</td>
<td>-0.143</td>
<td>-2.231</td>
</tr>
<tr>
<td>Exam_3</td>
<td>Self-assessment in week 14</td>
<td>0.096</td>
<td>0.030</td>
<td>0.162</td>
<td>3.156</td>
</tr>
<tr>
<td></td>
<td>Checking grades in week 13</td>
<td>-1.288</td>
<td>0.577</td>
<td>-0.143</td>
<td>-2.231</td>
</tr>
<tr>
<td>Exam_4</td>
<td>Self-assessment in week 14</td>
<td>0.031</td>
<td>0.014</td>
<td>0.113</td>
<td>2.162</td>
</tr>
<tr>
<td></td>
<td>Checking grades in week 16</td>
<td>-0.800</td>
<td>0.393</td>
<td>-0.144</td>
<td>-2.033</td>
</tr>
<tr>
<td></td>
<td>Checking grades in week 17</td>
<td>0.575</td>
<td>0.165</td>
<td>0.239</td>
<td>3.485</td>
</tr>
</tbody>
</table>

*Note.* exam_1 in week 6, exam_2 in week 10, exam 3 in week 14, exam 4 in week 17

In terms of relations between monitoring behaviors and exam score, figure 6 provides visual evidence, showing the use of the monitoring tools, self-assessment quizzes and My grades by academic achievement (B or Better vs. B- or Worse). In this course, students with B or Better can take upper-level courses in their academic program. In terms of the use of both tools, students with high performance (B or Better) not only tended to assess their learning status but also checked their present grades more frequently than those with low performance (B- or Worse) throughout the semester. Specifically, there are four spikes that mean the tool is much more frequently used in the week than around other weeks and correspond to each exam. The extent of this use appears to differ across achievement groups where in particular, the B or Better group made the greater use of these monitoring tools immediately prior to or posterior to exams (red broken lines, Figure 6).
C. Mining Metacognitive Processes

We next sought to develop a richer understanding of students’ metacognitive processes by examining the patterns of such behaviors as they emerged within sessions of time spent on the course site. We first identified sequences of metacognitive events that occurred sufficiently frequently that they are a phenomenon of interest, and then examined how these frequently occurring patterns influenced student achievement. This pattern mining approach was employed to focus on monitoring learning via self-assessment quiz use and monitoring performance using the My Grades tool in the LMS.

The log of learning events was first pivoted to sequence events within a session, producing from the log blocked sequences of learning events per learner and session, which is a suitable format for pattern mining, in where all sets of learning activities are compared with each other. To find patterns reflecting more concise learning processes, navigational events such as ‘link to the content area’ and ‘content folder’ were excluded from the analysis. The result of the pattern mining analysis indicated that variations of repeated self-assessment in the pattern of

**Figure 6. Use of monitoring self-assessment and performance by final grade**

- **MONITORING LEARNING / SELF-ASSESSMENT**
- **MONITORING PERFORMANCE / CHECKING GRADES**

- **Exam_1**, **Exam_2**, **Exam_3**, **Exam_4**

- Frequency

- Week
MONITORING SELF-ASSESSMENT → MONITORING SELF-ASSESSMENT were the most frequently occurring during LMS sessions. Students commonly used two self-assessment quizzes consecutively, which can reflect retrieval practice (e.g., use of the same unit quiz repeatedly to rehearse knowledge and strengthen retrieval), and can also reflect metacognitive monitoring of one or more units of content. Repeated use of ungraded self-assessment quizzes spanned a single repetition of the learning event (i.e., MONITORING SELF-ASSESSMENT → MONITORING SELF-ASSESSMENT) most often, but a considerable number of patterns included three or more of such events (i.e., MONITORING SELF-ASSESSMENT → MONITORING SELF-ASSESSMENT → MONITORING SELF-ASSESSMENT→ …).

Upon mining these patterns, we next examined whether repeated self-assessment was more common amongst students in a specific achievement goal group. The frequency of repeated self-assessment is visualized in Figure 7. More students in the High Mastery group employed this pattern of metacognitive events than students in the High Performance and High-Goal endorsement groups, particularly during sessions in weeks before exams. Before each exam, approximately 50 percent of students in the mastery-oriented group used the monitoring self-assessment tool. In addition, High Mastery group students more often used the pattern one week before exam 1 and 2 than the week of the exams as well as in the week of the final exam. However, all groups used similarly for the exam 3 (i.e., week 14).

We further examined differences between students who employed repeated self-assessment within sessions compared to those who did not by conducting group comparisons per unit and corresponding exam. Visual analyses by achievement level also indicate tracks with this behavior; exam scores by use of the sequence were observed to examine the effect of the sequence on student achievement (Figure 8). For all exams, students who conducted a repeated
behavior pattern indicative of metacognitive monitoring and control outperformed those who did not.

![USE OF THE SEQUENCE BY ACHIEVEMENT GOAL PROFILE](image)

**Figure 7. Use of metacognitive monitoring patterns by achievement goals**

![EXAM SCORES BETWEEN REPEATED AND SELF-ASSOSR AND CONTROL GROUP](image)

**Figure 8. Exam scores by use of the metacognitive monitoring pattern**

In addition, Table 4 shows the result of t-test with mean differences between the repeated self-assessor and control group for all exams. On all exams, the group who repeatedly self-assessed their learning significantly outperformed those who did not. Of particular interest is the size of the effect observed for exam 3 and exam 4, which assesses mastery of challenging units.
of contents in the course. While mean scores confirm these were more difficult exams, (overall mean score = 69.7 for exam 3 and 74.9 for exam 4), the mean differences of 7.15 and 8.20 are much greater than other two exams and correspond to the medium effect size on performance, d = .409 and .610. We thus conclude that the behavior revealed through pattern mining may be a particularly important one for promoting learning in challenging tasks.

**Table 4. Result of T-test of Exam Scores**

<table>
<thead>
<tr>
<th></th>
<th>Use of the sequence</th>
<th>No use of the sequence</th>
<th>t</th>
<th>df</th>
<th>Sig.</th>
<th>Mean difference</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam 1</td>
<td>272</td>
<td>79.7</td>
<td>12.6</td>
<td>105</td>
<td>76.1</td>
<td>13.6</td>
<td>2.37</td>
</tr>
<tr>
<td>Exam 2</td>
<td>225</td>
<td>83.1</td>
<td>11.6</td>
<td>152</td>
<td>80.5</td>
<td>11.5</td>
<td>2.09</td>
</tr>
<tr>
<td>Exam 3</td>
<td>207</td>
<td>72.9</td>
<td>16.1</td>
<td>170</td>
<td>65.8</td>
<td>18.5</td>
<td>4.01</td>
</tr>
<tr>
<td>Exam 4</td>
<td>175</td>
<td>79.3</td>
<td>11.4</td>
<td>202</td>
<td>71.1</td>
<td>15.2</td>
<td>5.84</td>
</tr>
</tbody>
</table>

5. Discussion

A. Person-centered analysis of students’ multiple achievement goals

Based on the score of AGQ-R, participants were categorized into three groups who have multiple achievement goals. This result confirmed that multiple achievement goals define several different groups (Barron & Harackiewicz, 2001). Person-centered investigations can provide more authentic inferences than individual analysis of achievement goals based on the variable-centered approach, which tends to ignore the additional effects of other goals and their interactive effects of simultaneously pursuing strong mastery approach, performance approach and performance avoidance goals vs. only approach goals vs. only a mastery goal (Bergman & Lundh, 2015; Bergman & Trost, 2006).

Aligning to research on individual goals, the High Mastery group performed better compared to other groups for most quizzes and some exams. These result falls in line with many
studies indicating a mastery approach is positively associated with student achievement whereas, the effect of performance approach was found to be inconsistent (Coutinho & Neuman, 2008; Ford et al., 1998; Senko, 2016). Specifically, mastery approach is positively associated with mediating factors such as persistence (Simons, Dewitte, & Lens, 2004) or deep studying (Lee, Sheldon, & Turban, 2003) that exist between achievement goals and academic achievement. The current study provides some additional support for the idea that highly – and here, singularly – mastery-oriented learners are able to achieve better performance through adoption of a unique set of cognitive and metacognitive learning strategies.

In addition, figure 3 shows that students in the High-Goal endorsement group performed the most poorly on both of exams and quizzes even though they had significantly greater mastery approach and performance approach than other groups (see Table 2). From the result, we could hypothesize that performance avoidance determines the overall characteristics of achievement goal profiles and further lead to negative relations with academic achievement, overwhelming other approach goals. According to a study by Hulleman and his colleagues (2010), the effect of performance goals on academic achievement is determined by their focus. If they focus on appearing talented, the relation with achievement is negative, while the focus is on outperforming others, the relation is positive. In this study, items on performance goals are designed to ask about performance (e.g., My aim is to perform well relative to other students) so it is hypothesized that students with high performance approach are more likely to perform better. Therefore, it can be said the high performance avoidance is the strong determinant of the characteristic of High Goal endorsement despite high approach goals that influence positively scores on exams and quizzes.
Such a role of performance avoidance was also shown to be effective in determining the characteristics of High Approach and High Mastery group. They have the same level of mastery approach, but the different level of performance approach, in which High Approach have greater performance approach than High Mastery. However, considering the result where High Approach outperformed High Approach that has lower performance avoidance than itself,

In spite of the importance of avoidance goals, to date, considerable research has focused on approach goals by mastery goal theorists (Senko, 2016), as they are positively associated with learning process and outcomes such as cognitive strategy. However, from the perspective of multiple goals (Barron & Harackiewicz, 2001; Harackiewicz et al., 1998), people not only tend to hold multiple goals together, but also can handle them effectively to produce benefits from each goal. Therefore, it is necessary to take the pursuit of multiple goals into account based on the person-centered approach.

Lastly, the difference in scores on exams and quizzes by achievement goal group was significant just at the beginning of the course, disappearing afterward. The result might be explained by the change in achievement group over time throughout the semester. Some researchers considered motivation personality traits, but this assumption ignores the nature of motivation that learners can be motivated differently depending on time or context (Schunk et al., 2012). In this sense, a longitudinal study (e.g., latent transitional analysis (LTA)) based on the person-centered approach make it possible for researchers to examine changes in motivational profiles and their influence on outcome variables over time. Although recently researchers have paid considerable attention to person-centered research (Conley, 2012; Litalien et al., 2017), few studies have carried out a person-centered longitudinal study on motivation (Gillet, Morin, & Reeve, 2017; Martinent & Decret, 2015).
B. Exploring metacognitive monitoring behaviors

Students who performed well (i.e., with grades of B or better in the course) used both of metacognitive monitoring tools more frequently than those who performed poorly. This finding aligns to previous studies in which metacognition was found to serve as a mediating process between mastery goals and academic achievement (Ames & Archer, 1988), and a negatively associated process with the performance avoidance goals in the High Goals group (Schmidt & Ford, 2003).

Those in the High Goals group were apter than others to monitor their performance in the weeks following exams, which further illustrates that disentangling varieties of metacognitive monitoring might improve the precision with which we model their relationships with particular achievement goals. Performance avoidance-oriented people tend to focus on demonstrating that they are similarly competent in a task, so they have interests in performance.

Prior to exam 2 and the Final exam, mastery-oriented learners engaged in the greater use of self-assessment quizzes. Those with High Mastery goals also used the tools earlier than other two groups for exam 2. Metacognition has been positively associated with planning strategies (Dinsmore, Alexander, & Loughlin, 2008), which would support the inference that mastery-oriented students’ earlier and more frequent self-assessment practices derive from playfulness. According to Clarebout and her colleagues (2013), metacognitive skill concerns with determining when a tool would be used to attain benefits. Therefore, the result where High Mastery students used self-assessment quizzes more frequently earlier than other groups provides inferences that they are more skilled at using metacognitive learning strategy.

In addition, the finding that students in the High Mastery group, who had greater use of the self-assessment tool obtained higher final grades than other groups could be explained from the perspective of SRL. According to relevant literature (Griffin et al., 2013, Winne & Hadwin,
monitoring processes trigger control processes where existing learning tactics or strategies are adjusted by comparison to a student’s standard for learning based on discrepancies found during monitoring the learning. Therefore, students who assessed their learning were more likely to have more chances to find inefficient learning strategies and revise them. Eventually, this control process might lead students to the adjusted learning process and further, make it possible for learners to achieve high performance.

Mastery-oriented students are thought to be learners who respond well to negative feedback (which is available in self-assessment quizzes and can be used to plan future study) and who enjoy the development of intrapersonal competence through challenges (Elliot & Murayama, 2008; Senko et al., 2011). This orientation might make them more apt to utilize such a resource, and to benefit from doing so.

In contrast, performance avoidance is theorized to be associated with negative factors such as disorganization, low interest, and surface learning processing (Elliot et al., 1999; Elliot & Church, 1997; Elliot & Harackiewicz, 1996). Therefore, students in High Goal groups had no interest in self-assessment, but checked their grades more frequently than other groups, which is supported by the greatest performance avoidance in the group. Similarly, the High Mastery group who had the smallest level of performance avoidance used My grades the least, which illustrates that they are not interested in their place relative to others, but the development of their knowledge.

C. Mining metacognitive processes

Use of richly modeled sequences of metacognitive processes and subsequent pattern mining revealed stark differences among the metacognitive behaviors of those with different achievement goal profiles. In particular, earlier use of the sequence by High Mastery in
preparation for exams becomes more obvious, suggesting that mastery approach students who use the patterns are more apt to plan and monitor – and to do so earlier – than performance-oriented students who employ this practice. In other words, the result also demonstrated students in High Mastery are more likely to self-regulate their learning in that they tended to prepare for the exams by mastering the learning contents earlier than those in other groups (Winne & Hadwin, 1998). That is, the primary purpose of monitoring is to identify whether learning achievement corresponds to student’s existing standards that are criteria against which products students created are monitored, resulting in the cognitive evaluation (Winne, 2010; Winne & Hadwin, 1998; Winne & Nesbit, 2009). In this sense, we can say that mastery-oriented students tend more to monitor their learning through self-assessment quizzes.

Considering more use of the sequence by mastery students, they might find the self-assessment tool to be effective in monitoring their learning. Mastery approach students endorse that they wish to learn as much as possible and therefore, may iteratively increase their engagement in behaviors they perceive to be effective in helping them learn (Coutinho & Neuman, 2008).

In pattern mining analyses, repeated self-assessment within an LMS session led to a clear difference in academic achievement. A more nuanced treatment of these rich behavioral data was necessary to reveal this trend, which provides support for the use of data-driven approaches to explore learning behavior. That said, the behaviors that emerge should be considered in light of extant theory about the kinds of learning processes such behaviors may represent, and what implications they are theorized to have for learning. Here, monitoring is known to be more common amongst mastery-oriented learners (Butler, 1993) and to be associated with superior performance. This knowledge guided subsequent analysis and allowed us to align our findings to
further support and refine theoretical assumptions. In particular, the repeated-self assessment had
the greatest effect on exam 3, which assessed the most challenging unit in the course. As the
overall mean is lowest on this exam and the mean difference was the largest, it provides evidence
that SRL processes conducted during self-assessment may indeed be most impactful when
learning tasks are the most challenging; a central tenet of SRL theory.

Beyond the quantity of metacognition, it is also important to measure the quality of
metacognitive learning process, which could be achieved by detecting and investigating
meaningful patterns within the sequence of events log file (Veenman, 2013). Since
metacognitive processes have the dynamic nature, researchers should implement subsequent
analysis for the better understanding of such processes (Greene & Azevedo, 2010). The
understanding of metacognitive processes through the in-depth investigation provide
opportunities to figure out better relationships between mastery goals and academic achievement
because metacognition plays the role of a mediator among them (Coutinho, 2007; Mirzaei,
Phang, Sulaiman, Kashefi, & Ismail, 2012)

Additionally, the generation of data consisting of sequential events has to precede the
application of mining techniques such as sequential pattern mining. However, to date, the
majority of studies on metacognition have been conducted with data measured by off-line
measurement such as survey, which prevents researchers from digging in metacognition process
through cutting-edge analysis techniques. In this sense, the log data by an LMS allows
researchers to investigate metacognitive learning process closer than does off-line measurement.

6. Implications

A data-driven approach made it possible to better understand relations among
achievement goal profiles, metacognitive monitoring behaviors, and academic achievement. The
person-centered analysis provided authentic and generalizable groups and afforded observation of the learning behaviors of learners with typical combinations of goals.

The rich data that can be derived from LMS logs enabled a thorough tracing of learning events, which could be visualized for inspection of the timing, frequency, and differences in behaviors across groups. The inference drawn about what learning processes these behaviors reflect requires some validation, but such data are an asset to the refinement of SRL theories.

It was found that self-assessment is more common amongst the mastery-oriented, and that those who self-assess also perform better in a large lecture course. In particular, repeated use of self-assessment quizzes designed to support metacognitive monitoring produced a significant difference in exam performance. If undertaken with appropriate intentions and tactics, this strategy could be an effective way to improve student performance and ultimately result in better academic achievement. The evidence could support instructors’ tendency to encourage students to frequently self-assess through use of ungraded quizzes with feedback.

These analyses demonstrate how log data can capture the learning process in great detail, and how inferences can be made from behaviors and used to test assumptions related to SRL. Additional analyses using a frequent sampling of motivation and ongoing tracking of learning events will be needed to examine the complex and dynamic relations between processes that are proposed in SRL theories. For instance, future research in this project will make use of such data to track changes in achievement goals and motivations, shifts in metacognitive learning behaviors and changes in the relations between them as task conditions change over the course of learning. While these data may afford observation of such processes, modeling such complex models will continue to pose challenges, and require new conversations in the field about the
importance of combining rich data, the data-driven discovery of behavior patterns that emerge, and theoretical lenses that can be used to interpret them.
Chapter 3: A Latent Profile Analysis of Undergraduates’ Achievement Motivations and Metacognitive Behaviors, and their Relations to Achievement in Science

1. Introduction

Within the self-regulated learning (SRL) framework (Butler & Cartier, 2018; Winne & Hadwin, 1998; Zimmerman & Campillo, 2003), motivational factors provide learners with standards by which the product of their learning processes by cognition and metacognition is evaluated. According to Holy and Dent (2018), motivation is internal resources that promote or threaten the regulation of cognitive and metacognitive processing. That is, a student might perceive the same outcome of learning processing differently depending on the status of motivation. For example, students motivated by mastery goals are more likely to perceive errors as the natural part of a learning process. However, students with more performance goals accept them in a different way in which errors might be treated as a failure (Senko et al., 2011).

Metacognition processes based on monitoring and control strategies as key components in the SRL process contribute to academic performance by optimizing learning through comparison between learning products and the standard. If the learning outcomes are not satisfactory compared to standards learners expects, they could change learning strategies until they obtain desirable products. That is, students who utilize metacognitive learning strategies well can effectively figure out what is wrong and how it should be changed during learning. Increasingly, the iterative process leads to the achievement of academic standards.

Academic motivation has been theorized to influence achievement, effort, educational and vocational choices, interest, and persistence (Covington, 2000; Wigfield, Cambria, & Eccles, 2012). In particular, researchers have paid considerable attention to self-efficacy, achievement
goals and expectancy-value theories to examine how motivation relates to learning processes and outcomes (for self-efficacy see Zimmerman, 2000a; for expectancy-value see Eccles et al., 1983; for achievement goals see Elliot & Hulleman, 2017).

However, since a majority of research on motivation has studied each motivation type such as achievement goals theory or expectancy-value theory individually, little is known about how the combined components from multiple motivation models function (Conley, 2012; Liem, Lau, & Nie, 2008). According to literature (Barron & Harackiewicz, 2001; Harackiewicz, Barron, & Elliot, 1998), a motivation process is complex with the pursuit of multiple motivation constructs in real situations. Therefore, the integrated perspective of multiple motivation types is expected to explain better the complex roles as well as relations of motivation with predictors or outcomes.

However, the majority of studies on motivation have been conducted based on the traditional variable-centered approach in which researchers are interested in examining how specific variables are associated with other variables (Coutinho & Neuman, 2008; Laursen & Hoff, 2006). In addition, methodologically, the approach usually employs linear statistical methods across time to measure relationships among relevant variables (Bergman & Trost, 2006). In contrast, the person-centered approach pursues identifying homogeneous groups of individuals who have similar features within their group but function in a different way compared to those in other groups (Magnusson, 2003). In this sense, studies carried out based on the variable-centered approach have the innate drawback of overlooking the motivational subgroups of individuals rather than those taking the person-centered approach, which becomes the obstacle to providing interventions that are more personalized according to the characteristics of each group (Wang & Degol, 2013). In addition, demographic information such as age, gender,
and ethnicity was found to influence motivation (for gender see Eccles et al., 1983; Dweck, 1986; for age see Watt, 2004; Nicholls, 1990; and for ethnicity see Graham, Taylor, & Hudley, 1998). Since relations between demographic information and motivation have been examined primarily at the level of variables, it remains to be seen how learners with various personal characteristics tend to align to emergent motivational profiles (Schunk et al., 2012).

Lastly, many studies have researched relations between motivation and metacognition. However, similar to research on motivation, most of them they have relied on variable-centered methods examining how motivation constructs are associated with certain metacognitive behaviors was examined (Coutinho & Neuman, 2008; Vrugt & Oort, 2008). However, students are more likely to employ many metacognitive learning strategies to manage their learning, some of them might heavily rely on few learning processes though. Therefore, in addition to motivational profiles, it is necessary to figures out how multiple types of metacognitive learning processes take place together during learning based on the person-centered approach. Further, these relations between motivation and metacognition at the level of profiles can provide interesting information that helps instructors figure out a student’s SRL process.

The purpose of this research is to investigate what motivational and metacognition profiles emerge from the combination of different types of motivation and metacognitive strategies using latent profile analysis (LPA). Subsequently, this research examines relations between motivation and metacognition through the likelihood of profile membership. Lastly, this research aims to investigate motivational and metacognitive profiles influence academic achievement that is measured by exam scores.
2. Literature Review

A. Self-Efficacy

Bandura (1977) proposed social cognitive learning theory where a combined influence of external and internal factors regulate behaviors. Self-efficacy is defined as the judgment of capabilities to successfully perform a series of actions to achieve designated academic goals such as completing an assignment (Bandura, 1977, 1986; Schunk, 1989). Students’ self-beliefs about their perceived personal abilities play a critical role in developing the motivation to learn (Zimmerman, 2000a).

Self-efficacy has a cyclical relation with self-regulatory processes including goal setting, strategy use, self-monitoring, and self-judgment (Winne & Hadwin, 1988; Zimmerman, 2000a). Therefore, self-efficacy allows users to control their behaviors that leads to the achievement of tasks, so positively self-efficacious students are more likely to perform better, regulate their learning, and think more positively (Pajares & Miller, 1994; Schunk & DiBenedetto, 2014). In contrast, students who negatively perceive their capabilities tend to possess greater stress and to less use effective cognitive strategies, and when facing difficult tasks, they avoid or give it up (Bandura & Wood, 1989). Specifically, students’ self-efficacy positively influences a variety of learning processes and outcomes including uses of learning strategies (e.g., reviewing notes, organizing, and transforming), academic attainment, problem-solving performance, engagement, and work-related behaviors, while being negatively related to seeking adult helps (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996; Bernacki, Nokes-Malach, & Aleven, 2015; Hoffman & Spatariu, 2008; Stajkovic & Luthans, 1998; Pintrich & De Groot, 1990; Zimmerman & Martinez-Pons, 1990).

In terms of metacognitive behaviors, students with high self-efficacy are more likely to regulate their learning with more use of metacognitive strategies (Pintrich & De Groot, 1990).
According to a study by Pintrich (1999), a wide range of students from middle schools to colleges demonstrated positive relationships between self-efficacy and self-regulatory strategies including planning, monitoring, and regulating as well as performance including examinations, lab report, and final grades. Sniehotta, Scholz, and Schwarzer (2005) investigated relations among patients’ self-efficacy, action planning, action control, and physical exercise. The finding shows that self-efficacy significantly related to action planning that then influences actual exercise. However, students who believe that they have low capabilities are less likely to employ effective cognitive strategies such as constructing or selecting as they tend to attribute success to luck and failure to their abilities (Borkowski, 1988; Thomas & Rohwer, 1986). The majority of relevant research has revealed consistent results where high self-efficacy leads to students’ successful academic outcomes and more active learning behaviors. In addition, when learners are faced with complicated tasks that require spending a lot of time, self-efficacy plays a critical role to keep them engaged (Pintrich, 1999).

B. Achievement Goal Theory

The achievement goal theory has played a central role in the research on motivation for the past several decades, explaining how and why students engage in learning tasks with what reason or purpose (Elliot & Hulleman, 2017). The theory appeared in the 1980s and distinguished two qualitatively distinct goal types - mastery and performance – based on the definition of personal competence for achievement behaviors, either developing competence or demonstrating competence (Dweck, 1986; Maehr & Nicholls, 1980).

Mastery goals are concerned with a desire to develop competence by improving or learning as much as one can, while performance goals care about demonstrating competence by outperforming others or matching their success with less effort (Senko, 2016). In addition,
mastery goals define success with task-based standards, whereas performance goals use normative standards as criteria for judging success. Therefore, the premise underlying the achievement goal theory is that students who are mastery-goal oriented act in ways different from those whose purpose is to get the highest scores in their group (Conley, 2012). This approach has been called a mastery goal perspective in which the pursuit of only mastery goals is considered to be more beneficial than the pursuit at performance goals (Elliot & Dweck, 1988).

Whereas mastery goals have shown consistent positive associations with a variety of outcomes such as self-regulated behaviors and emotions (Bernacki et al., 2012), effects of performance-approach goals have been inconsistent depending on the focus of the goals. Specifically, according to the meta-analysis by Hulleman, Schrager, Bodmann, and Harackiewicz (2010), when focusing appearing talented, performance approach goals predicted negatively academic achievement, whereas they positively predicted academic achievement when focusing on outperforming others.

However, such inconsistent results of performance approach encouraged researchers to endorse a multiple goals perspective where a combination of mastery and performance goals should be considered adaptive (Harackiewicz, Barron, & Elliot, 1998). Barron and Harackiewicz (2001) have suggested four different hypotheses to explain how multiple goals yield positive effects. Four hypotheses include (a) additive goals where mastery and performance goals are independent, (b) interactive goals where mastery and performance goals create extra effects through their interaction, (c) specialized goals where each goal has their dedicated effects, and (d) selective goals where achievement goals vary depending on the situation.

In spite of the advantages of investigating multiple goals simultaneously rather than examining each achievement goal in isolation or including limited interactions at best, the
majority of research on achievement goals have employed a variable-centered approach that is
centered concern with examining the causal relationships between variables (Coutinho & Neuman,
2008; Elliot, & Church, 1997; Liem et al., 2008). However, after a multiple goals perspective
was suggested, some studies based on a person-centered approach have examined how learners
are motivated by the pursuit of multiple goals (Litalien, Morin, & McInerney, 2017; Pintrich,
2000a; Wilson, Zheng, C., Lemoine, Martin, & Tang, 2016). According to Pintrich (2000a), both
groups with high mastery and low-performance goals as well as high mastery and high
performance showed the most adaptive learning patterns including metacognitive learning
strategies. That is, this result supports a mastery goal perspective and a multiple goals
perspective.

C. Expectancy-Value Theory

Eccles’ expectancy-value theory (Eccles et al., 1983) has served as a comprehensive
framework for the studies of academic motivation. Specifically, the outcomes related to
expectancy and task values include academic performance, cognitive engagement, effort,
studies demonstrated that students who are highly confident about their ability and competence
related to certain tasks more tend to make an effort for and engage in them (Eccles et al., 1983;
Wigfield & Eccles, 2000).

Expectancy-value theory has two main components; 1) expectancy for success, and 2)
task value (Conley, 2012). Expectancy for success focuses on the general question “Can I do this
task?” while subjective task value beliefs focus on the general question “Why do I want to do
this task?” (Wigfield et al., 2016). Accordingly, expectancy is defined as beliefs about the future
outcomes (Roese & Sherman, 2007), whereas value refers to psychological experiences when
being attracted to an object or activities (Higgins, 2007). Such expectancy for success and the value they place on the task are theorized to predict their choices, persistence, and achievement outcomes. (Eccles & Wigfield, 2002; Wigfield & Eccles, 2000).

Subjective task value is composed of intrinsic (or interest), attainment (or importance), and utility (or usefulness) value, as well as cost (defined by effort, opportunity, and psychological costs; Eccles et al., 1983). Intrinsic value is the enjoyment individuals get from performing the task. When children intrinsically value an activity, they often become deeply engaged in it and can persist at it for a long time (Wigfield & Eccles, 1992, 2000). In this sense, it can be said that mathematics has high intrinsic value to mathematicians in that they feel like that solving math problem is interesting, so they highly engage in the process. Attainment value is defined as the importance to the self of doing well on a task. The value provides individuals with an opportunity to confirm the central aspects of their self-schemas (Wigfield & Eccles, 1992). For example, if doing well at the Olympics is one of the most important values for athletes, the event has high attainment value, and a good athlete is the central aspect of their self-schemas (Schunk et al., 2012). Utility value refers to how the task relates to future goals and can be seen as capturing more extrinsic reasons for doing the task. For example, if a student who has a plan to pursue a doctoral program in chemistry is more likely to have higher utility value but lower intrinsic value of organic chemistry courses at the school than others who have different plans. Therefore, individual learners have different motivation profiles based on expectancy-value depending on their situation such as goals, interests, environment, and so on.

Cost refers to the accumulated negative aspects in engaging in the task, including, the amount of time and effort required for the task, forgoing of engagement in other activities (e.g., gainful employment, leisure activities), and the anticipated emotional states (e.g., performance
anxiety). In other words, cost value could be defined as what individuals have to give up by doing a task (Wigfield & Cambria, 2010). Therefore, cost is shown to be highly associated with a choice among outcomes in that one choice results in the giving up of other options (Eccles et al., 1983). According to Eccles (2005), value and cost operate together to determine individuals’ achievement motivation. If learners perceive more positive value than cost toward a task, they would be more engaged in a task. In this sense, a person-centered approach can shed light on interactive effects of value and cost better than traditional methodologies.

In spite of such important findings, so far cost has been the least studied component of task values. Specifically, some studies combined cost with other task value components into one score on average (Buehl & Alexander, 2005), and there even was a case in which cost was excluded from task value components (Andersen & Cross, 2014; Chow, Eccles, & Salmela-Aro, 2012). However, recently important studies on cost have been increasingly performed (Wigfield, Rosenzweig, & Eccles, 2017). Therefore, this study includes three types of costs (i.e., psychological, effort, and opportunity) as independent components based on the person-centered approach.

D. An integrative view of multiple motivation types

Motivation could be defined as a motive to engage in a particular task (Weiner, 1985). In education, a basic idea underlying motivation is that the likelihood of learning behaviors is a result of how much individuals believe in their capabilities (Zimmerman, 2000a), values toward particular tasks (Eccles et al., 1983), and goals they pursue in a given situation (Elliot & Dweck, 1988). Like this, there are several dominant motivational theories to explain what makes learning, but how they are interrelated in predicting academic achievement is unclear. In other words, research on several types of motivations demonstrated different relations between them.
Specifically, although Eccles and colleagues’ original expectancy-value model (Eccles et al., 1983) proposed that goals influence expectancy and values that further predict academic achievement, some studies demonstrated different relations from the theory. Specifically, competence belief or values predict goal orientations (Chouinard, Karsenti, & Roy, 2007; Greene, DeBacker, Ravindran, & Krows, 1999). In addition, the expectancy-value belief mediates relations between goals and academic achievement (Plante, O’Keefe, & Théorêt, 2013). These complex relations could be explained the complex nature of motivation, in which motivation has multiple concepts: 1) behavioral antecedents, 2) processes during task engagement, and 3) outcomes (Hulleman, Durik, Schweigert, & Harackiewicz, 2008). Therefore, it is not surprising that the role and direction of motivational causal links could change and be interrelated rather than have fixed relations.

In practice, expectancy-value and achievement goals contribute to complicated learning processes simultaneously (Conley, 2012). For example, students who value a calculus course for their future may not engage learning if they doubt capabilities to learn math in the course or they have more interests in showing better performance to others. In contrast, although they think the course requires more efforts than those expected to obtain desirable grades or the course seems to be not interesting, they who have high competence belief and pursue the mastery of calculus may engage in learning. These behavioral patterns by partially overlapping motivational states (e.g., positive expectancy-value & performance goals) are more likely to be less distinct than ones by similarly directed motivations (e.g., positive expectancy-value & mastery-goals). Also, students might have different motivational profiles depending on their characteristics (Wang et al., 2016). Therefore, it is important to identify divergent motivational profiles that exist within subgroups of the population based on the combination of multiple types of motivations.
The interactive effect of these motivations on SRL are well described together in self-motivational belief of self-regulatory process developed based on social cognitive theory (Zimmerman, 2002). When goals in the SRL process are proximal, specific, and appropriately challenging, learners’ sense of self-efficacy would be increased (Schunk, 1990). Also, self-efficacy is influenced differently depending on the type of goals, suggesting the higher level under mastery-goals than performance goals (Elliott & Dweck 1988). In terms of achievement goals and values, mastery goals are theorized to be more beneficial to academic outcomes than performance goals. Whereas students with performance goals are more focused on the product of learning, mastery goals enjoy the process of learning (Flum & Kaplan, 2006). Therefore, students with high performance goals might not concern the value of tasks (e.g., intrinsic value). Specifically, students motivated by mastery goals are more likely to have positive task values in that the goals allow students to have more interest in the mastery of task itself and enjoy challenging task considering errors are a natural learning process (Bong, 2001; Eccles & Wigfield, 2002). Eventually, the interactions of these motivational belief lead to the self-regulation of learning through the proper use of learning strategies (Zimmerman & Campillo, 2003).

There have been some attempts to investigate different types of motivation together from a perspective of an integrative analysis (Braten, Samuelstuen, & Strømsø, 2004; Conley, 2012; Hulleman et al., 2008; Pintrich, 1999; Pintrich & De Groot, 1990). Pintrich (1999) demonstrated an association between motivation and self-regulated learning with a wide range of samples, students from middle school to university. Motivation in the research includes self-efficacy, task value, and goal orientation. However, he investigated multiple motivation types individually using correlation and regression, so he did not look at combined effects. Hulleman and his
colleagues (2008) investigated relations among achievement goals, task values, performance, and interest in both contexts of a college classroom and a high school sports camp. The finding revealed that initial interest and mastery goals were found to be predictors of subsequent interest and task value played a role as a mediator between these variables. In addition, actual performance (e.g., final course grades or coach ratings) was predicted by performance approach goals and utility value. However, most of the studies examined an effect of each motivational construct exclusively based on a variable-centered approach, ignoring a combination of different motivational perspectives.

Unlike other studies, Conley (2012) employed a person-centered approach in examining combined pathways of achievement goals, competence belief, and task values using a clustering analysis. Seven clusters were produced based on eight types of motivation measures. This research employed a clustering analysis that produced seven groups based on achievement goals. The result suggested that an integrated approach with multiple motivations better explained how motivation influence learning outcomes. However, the method subjectively chose the number of clusters for the best grouping, latent profile analysis provides a more objective approach with model fit indices including AICs, BICs, LMRs, and entropy (Pastor et al., 2007). In addition, the research treated cost as one structure, ignoring its subconstructs: a) effort cost, b) opportunity cost, and c) psychological cost.

E. Person-centered approach

Modern psychology emerged at the end of the 19th century, the mainstream of research interests is a measurement of individual variables and reporting results at the individual level (Danziger, 1990). Psychology fields including basic and applied psychology were scientifically advanced with interests at the aggregated level, but unfortunately, individual characteristics in
most experiments were hidden (Bergman & Lundh, 2015). Due to this issue, increasing researchers in developmental psychology experienced problem-method mismatch (Bergman & Vargha, 2013), and then made researchers paid considerable attention to an individual level as well as a person-centered approach (Bergman & Lundh, 2015).

The main difference between variable-centered and person-centered approaches is the research target (Block, 1971). The variable-centered approach focuses on population or groups of individuals, whereas the person-centered approach is more interested in individuals. Mäkikangas and Kinnunen (2016) pointed out that there are theoretical and methodological differences between person-centered and variable-centered approach.

Theoretically, in the variable-centered approach, theories are developed by considering the basic concepts as variables whose developmental importance usually is indicated by statements about (causal) relations between these variables (Bergman & Trost, 2006). Therefore, this approach focuses on the relations among variables at the group level.

The person-centered approach employs a holistic-interactionistic view as a theoretical framework where things process as wholes that cannot be discomposed into independent small parts, and the whole thing has greater value than when it is split into the separate parts (Magnusson, 1988). In terms of the framework, there are two propositions. First, all individuals should fall into a unique cluster based on their characteristics. Another one is that each cluster should be able to be characterized by its distinct patterns of constructs. For these propositions, individuals should meet the following two assumptions (Magnusson, 2003). All individuals do not have the same pattern for distinct clustering. In addition, the number of clustering based on patterns is restricted, meaning some students have the similar pattern of components.
Methodologically, since the variable-centered approach is more concerned with associations among variables (Bergman, Magnusson, & El-Khoury, 2003; Mäkikangas & Kinnunen, 2016), means comparison methods such as t-test or analysis of variance (ANOVA), correlation, regression, and factor analysis are commonly used methods in the approach (Magnusson, 2003). These analytic models allow researchers to answer questions about the relative importance of predictor variables in explaining variance in outcome variables (Bergman et al., 2003).

In contrast, the purpose of the person-centered approach is to identify groups of individuals who function in a similar way to others within a given group, but function differently relative to other clusters (Magnusson, 2003). The approach is based on the assumption that the population is heterogeneous as to how variables function on outcomes (Laursen & Hoff, 2006). In other words, the association among variables is heterogeneity across the sample, but homogenous between members within subgroups. Therefore, the interest of the person-centered approach is in discovering the factors that characterize the group of individuals as particular combinations of motivational variables better explain individuals than each variable (Hayenga & Corpus, 2010; Mäkikangas & Kinnunen 2016). To identify interesting motivation profiles, most of the studies on motivation based on a person-centered approach have employed clustering analysis (Conley, 2012; Karabenick, 2003) or latent profile analysis (Chen & Usher, 2013; Pastor et al., 2007; Tuominen-Soini et al., 2008).

Chen and Usher (2013) investigated what sources made self-efficacy. The sources include mastery experiences, vicarious experiences, verbal and social persuasions, and physiological and affective states. Then they examined relationships between profiles and self-efficacy and the result revealed mastery experience is the most important source of self-efficacy.
Also, it is found that an additive benefit of multiple sources influences self-efficacy most positively by interacting with one another. Lastly, as predictors, gender did not predict the membership of profiles, whereas the implicit theory of ability and grade level did.

F. Metacognition

Metacognition is defined as the knowledge about input and output information as well as operations that work on the information (Winne, 2011). A metacognitive learning process is considered one of the most important components for successful SRL in which thought, feeling, and actions are self-generated for planning and adaption to the attainment of designated goals (Butler & Cartier, 2018; Winne, 2018; Winne & Hadwin, 1998; Zimmerman, 2000b). Therefore, lack of the metacognitive awareness of personal learning issues might lead to their deficiency in learning (Zimmerman, 2002). Specifically, successful SRL could be achieved through the appropriate operations of processes including planning, monitoring, strategy use, handling of task difficulty and demands (Greene & Azevedo, 2009).

Planning refers to activities that help learners plan the operation of their cognitive strategies for learning and facilitate prior knowledge for better organization of tasks (Pintrich, 1999). According to Greene and Azevedo (2009), plan generation, sub-goal setting, activating prior knowledge, and recycling goals in working memory are sub-processes in the phase of planning. During the process, students need to create systematic methods for successful problem solving, which increase their performance by utilizing cognition effectively and controlling emotions (Zimmerman & Campillo, 2003). Highly self-regulated learners tend to spend more time planning their learning in which they analyze the task ahead (Winne, 2018).

Monitoring is defined as a process that identifies whether the outcomes of cognitive process correspond to standard and the considered to be an essential aspect of SRL (Winne,
Through the process, students check their understanding of designated tasks, and then depending on the discrepancy between current status and desired status they can determine whether current learning strategy is appropriate in achieving learning goals. In the phase of self-evaluation, outcomes of monitoring are compared with standards developed based on task condition and cognitive condition including motivation. Therefore, there is a strong association between motivation and metacognition learning process (Winne & Hadwin, 1988).

In order to examine metacognitive behaviors, a majority of the research on metacognition has carried out surveys (Winne & Perry, 2000) or think-aloud protocols (Greene & Azevedo, 2007; Azevedo, Cromley, & Seibert, 2004). In particular, since SRL was developed, many kinds of surveys such as Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al., 1993) have been developed and widely used to measure SRL including metacognition. However, for these methods, students might reconstruct their memories of their metacognition processes (Winne, 2002). Therefore, this research investigates log data to examine the dynamic and complex features of metacognitive learning processes in the SRL model (Bernacki, 2018). This approach provides the best way to capture complex self-regulated learning. Further, by applying a person-centered approach, combinations of multiple metacognitive learning behaviors are examined. Therefore, following research questions can be addressed through this research.

1) Motivation profiles

   a. What motivational profiles emerge from the combination of self-efficacy, achievement goals, and expectancy-value? How differently do motivations contribute to the profiles?
b. How much can one’s demographic information including age, gender, ethnicity, and first-generation predict the likelihood of memberships of motivational profiles?

c. Difference in academic achievement by memberships of motivation profiles
   i. Do scores on overall exams differ as a function of motivation profile membership?
   ii. Do scores on each exam differ as a function of motivation profile membership?

2) Metacognitive behavior profiles
   a. What metacognitive profiles emerge from the combination of monitoring and planning behaviors? How differently do motivations contribute to the profiles?
   b. How does one’s motivational profile predict one’s metacognitive profile?
   c. Difference in academic achievement by memberships of metacognition profiles
      i. Do scores on overall exams differ as a function of metacognition profile membership?
      ii. Do scores on each exam differ as a function of metacognition profile membership?

3. Methods

   A. Participants and Procedures

   Participants of 1326 undergraduate students from a large Southwestern university where ethnically diverse students study were analyzed in the study. They enrolled in a large face-to-face biology course from fall 2014 to fall 2016, which was designed to provide fundamental knowledge needed for continued health-science education. Specifically, 71.3% of the students
were females, and 45.5% were from an underrepresented minority group. In addition, 44.6% of the sample were from first-generation students, and an average age of the sample was 21.3 years old (see Table 5). At the beginning of each semester, students completed the motivation survey to measure demographic information and several types of motivation including self-efficacy, achievement goals, task value, and cost, which was administered through the LMS. Student learning activities taking place in the system throughout the semesters from fall 2014 to fall 2016 were captured and then stored in the database.

**Table 5. Demographic information of the sample**

<table>
<thead>
<tr>
<th></th>
<th>Gender</th>
<th>Ethnicity</th>
<th>Age</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Underrepresented</td>
<td>Non-underrepresented</td>
<td>Mean</td>
</tr>
<tr>
<td>2014 Fall</td>
<td>363</td>
<td>104</td>
<td>259</td>
<td>169</td>
<td>194</td>
</tr>
<tr>
<td>2015 Spring</td>
<td>249</td>
<td>69</td>
<td>180</td>
<td>118</td>
<td>131</td>
</tr>
<tr>
<td>2015 Fall</td>
<td>196</td>
<td>68</td>
<td>128</td>
<td>86</td>
<td>110</td>
</tr>
<tr>
<td>2016 Spring</td>
<td>254</td>
<td>73</td>
<td>181</td>
<td>107</td>
<td>147</td>
</tr>
<tr>
<td>2016 Fall</td>
<td>264</td>
<td>66</td>
<td>198</td>
<td>123</td>
<td>141</td>
</tr>
<tr>
<td>Total</td>
<td>1326</td>
<td>380</td>
<td>946</td>
<td>603</td>
<td>723</td>
</tr>
</tbody>
</table>

B. Measure

i. Motivation

**Self-efficacy.** We developed five items based on the academic self-efficacy scales from the patterns of adaptive learning scales (PALS; Midgley et al., 2000). All items designed to be course-oriented. One example item read, “I'm certain I can master the skills taught in class this year.” Higher scores indicate higher self-efficacy for a course.

**Achievement goals.** For achievement goals, 9-item Achievement Goals Questionnaire-Revised (AGQ-R) was used (Elliot & Murayama, 2008). The questionnaire has nine items, respectively three items for each goal (e.g., for mastery-approach “My goal is to learn as much as
possible”, for performance-approach “I am striving to do well compared to other students”, and for performance-avoidance “My goal is to avoid performing poorly compared to others”).

**Value.** Task values were measured with adapted items developed by Eccles and Wigfield (1995), which was designed for STEM college students. Four items were assigned for each value (e.g., for attainment value “Is the amount of effort it will take to do well in your STEM courses worthwhile to you?”, for intrinsic value “Learning the material covered in my STEM courses is enjoyable.” and for utility value “How useful is what you learn your STEM courses for your daily life outside school?”).

**Cost.** We adapted 12 items assessing three types of cost (effort, opportunity, and psychology) for STEM courses (Perez, Cromley, & Kaplan, 2014). a) effort cost “Considering what I want to do with my life, taking STEM courses is just not worth the effort.”, b) opportunity cost “I worry about losing track of some valuable friendships if I'm taking a lot of STEM courses and my friends are not.”, and c) psychological cost “My self-esteem would suffer if I tried in my STEM courses and was unsuccessful.”

To examine internal consistency reliabilities of scale scores reported through the survey, I looked at Omega coefficient (ω) and Cronbach’s alpha (α) (see Table 6).

**Table 6. Omega coefficient (ω) and Cronbach's alpha (α)**

<table>
<thead>
<tr>
<th></th>
<th>SELF</th>
<th>MAP</th>
<th>PAP</th>
<th>PAV</th>
<th>ATT_V</th>
<th>INT_V</th>
<th>UTI_V</th>
<th>EFF_C</th>
<th>OPP_C</th>
<th>PSY_C</th>
</tr>
</thead>
<tbody>
<tr>
<td>ω</td>
<td>0.887</td>
<td>0.891</td>
<td>0.751</td>
<td>0.688</td>
<td>0.809</td>
<td>0.906</td>
<td>0.818</td>
<td>0.725</td>
<td>0.800</td>
<td>0.667</td>
</tr>
<tr>
<td>α</td>
<td>0.889</td>
<td>0.754</td>
<td>0.737</td>
<td>0.837</td>
<td>0.747</td>
<td>0.91</td>
<td>0.806</td>
<td>0.798</td>
<td>0.892</td>
<td>0.821</td>
</tr>
</tbody>
</table>

*Note.* SELF: self-efficacy; MAP: mastery-approach; PAP: performance-approach; PAV: performance-avoidance; ATT_V: attainment value; INT_V: intrinsic value; UTI_V: utility value; EFF_C: effort cost; OPP_C: opportunity cost; PSY_C: psychological cost.

**ii. Metacognition Behaviors**

Blackboard Learn, an LMS used in the university, captured all learning behaviors of students through learning-support tools provided by instructors in the course and then stored
them in the log files of the LMS database. In terms of metacognition processes, students used a syllabus, calendar, and lecture schedule for their learning plan over the semesters. These documents provide topics to be learned based on a specific date so that students make a plan ahead. In addition to the overall schedule, instructors provided specific review paper and blueprints for each exam. Review papers include summarized key concepts and important questions to ask main points, and blueprints show the distribution of questions depending on the topic. Therefore, these supportive materials help students make a study plan more effectively in preparation for exams.

In terms of a monitoring process, to help student effectively monitor their progress, two types of features were provided: 1) monitoring learning and 2) monitoring performance. According to Winne (2004), students would have a hard time controlling their learning behaviors when monitoring their learning progress with incongruent or invalid standards. Therefore, the provision of such monitoring tools allows for more effective monitoring and control processes than arbitrary monitoring processes. Firstly, for monitoring learning processes, self-assessment quizzes that are ungraded provide students with opportunities not only to monitor their mastery of learning materials but also to practice for an upcoming exam. Therefore, students were able to use self-assessment quizzes until they think they master contents relevant to the exam. In addition to monitoring learning, “My Grades” allowed students to monitor their performance by checking grades. They figure out their ultimate score and a relative place to others as this tool provides an average score as well as the median.

ii. Exam Scores

In this course, four exams were administered throughout each semester. All exams were administered online in the LMS and consist of a combination of question formats; multiple-choice, fill in the blank, short answer, and essay questions may be utilized. These exams were
based primarily upon the lecture material presented in class. Specifically, the exam 1 through 9, and exam 3: chapter 10 through 12), but the last exam was a comprehensive exam to cover all chapters learned in the course. After the completion of each exam, scores on them were calculated and posted by instructors in the system.

C. **Data Analysis**

Splunk, a data management software, was used to collect data from the LMS server. This tool also enriches the log data generated by the LMS with metadata such as a course name to identify learning events (Bernacki, 2018). The software allows researchers to extract data in flexible formats with various levels of timestamps using the Splunk search language.

Data analysis was conducted using M-plus 6.1 maximum likelihood estimation with robust standard error (MLR). Compared to maximum likelihood estimation (ML), MLR calculates differently the values of chi-square for model test and standard errors for parameters, so is robust against the violation of assumptions such as the unmodeled heterogeneity (Hox, Maas, & Brinkhuis, 2010). Although LPA usually are performed based on aggregated scale scores, the analysis in some studies starts from a preliminary analysis using confirmatory factor analysis (CFA) on the items of questionnaires for some advantages (Litalien et al., 2017; Liu, Wang, Tan, Koh, & Ee, 2009). According to Rhemtulla, Brosseau-Liard, and Savalei (2012), however, it could be problematic to apply continuous normal theory ML to categorical variables measured on Likert scales in that this approach might lead to biased parameter estimates. The findings of their research suggested that for the data set with more than 5 categories the results of normal theory ML is similar to those of methods for categorical variables such as cat-LS. Therefore, in
this research in which items to measure motivation have six or seven categories, ML with robust standard error was used.

The CFA was conducted to validate the psychometric properties of our measures and estimate factor scores from which a latent profile analysis would be carried out. Factor scores were saved using “SAVE=fscore” in Mplus, by which the graphical representation of profiles can be readily interpreted (Litalien et al., 2017). Also, although using factor scores from a preliminary measurement model does not provide complete control for measurement error, giving more weight to more reliable items makes it possible to provide partial control for measurement error (Morin, Boudrias, Marsh, Madore, & Desrumaux, 2016).

This CFA includes ten correlated factors: 1) self-efficacy, 2) mastery-approach, 3) performance-approach, 4) performance-avoidance, 5) attainment value, 6) intrinsic value, 7) utility value, 8) effort cost, 9) opportunity cost, and 10) psychological cost. CFA models were assessed by the Tucker-Lewis index (TLI), the comparative fit index (CFI), and the root mean square error of approximation (RMSEA). According to Marsh, Hau, and Grayson, (2005), values more than .90 of TLI and CFI, and smaller than .06 indicate excellent model fit.

Based on the factor scores of 10 types of motivation estimated from the CFA, latent profile analysis was conducted to identify motivational profiles. Models were estimated based on 5,000 random sets of start values, 100 iterations per random start, and the 200 best solutions retained for final stage optimization. To identify the optimal number of motivation profiles, a series of statistical indicators were used including the consistent AIC (CAIC), the Bayesian information criterion (BIC), the sample-size adjusted BIC (ABIC), the adjusted Lo–Mendell–Rubin likelihood ratio test (aLMR), the bootstrap likelihood ratio test (BLRT), and Entropy (Geiser, 2013; Morin & Wang, 2016). Lower values on AIC, CAIC, BIC, and ABIC indicate
better model fits while aLRT and BLRT are used to compare the model with k profiles with a model with the k-1 profile(s) and a significant result indicates the k profile model is superior to the k-1 profile model. Entropy ranges from 0 to 1; a higher value suggests a more accurate classification.

Next, multinomial logistic regression was employed to estimate how much demographic information influences the likelihood of membership in motivational profiles. The demographic information includes age, gender, whether a student is first-generation or not, and whether a student is underrepresented or not. As predictors, age was treated as a continuous variable and gender, underrepresented, and the first generation were analyzed as binary variables.

Another LPA was conducted to identify a model with the optimal number of metacognition profile based on three metacognition behavior indicators: 1) monitoring learning through self-assessment quizzes, 2) monitoring performance by checking grades, and 3) planning with blueprints and review papers. Similar to the LPA for the identification of motivation profiles, the optimal number of metacognition profiles was identified by model fit indicators.

To investigate relations between motivational and metacognitive learning profiles, another multinomial logistic regression was conducted. Four motivation profiles identified from the first LPA were used as categories of a nominal independent variable, and three types of metacognition profiles were used as categories of a nominal dependent variable.

In addition, time series line graphs were used to examine different patterns of each metacognitive process behaviors by motivation profiles over a semester. Since the day of each exam was slightly different by semester, the representation of time was reorganized based on each exam (see Figure 9). 1) grade-checking and planning week – within seven days after each
exam, 2) no-press period - between seven days after and before each exam, and 3) cram week - within 7 days before each exam.

**Figure 9. Reorganization of Time Frame**

Lastly, to examine how motivation and metacognition profile influence academic achievement, MANOVAs were conducted with four exam scores (see Figure 9). If the result of the analysis is significant, a post hoc test would be performed to identify significant difference in scores among profiles.

### 4. Results

#### A. Motivation profiles

The means, standard deviations and correlation coefficients of motivational constructs used in the analysis are shown in Table 7.

**Table 7. Grand Means, Standard Deviations, and correlations of constructs.**

<table>
<thead>
<tr>
<th></th>
<th>SELF</th>
<th>MAP</th>
<th>PAP</th>
<th>PAV</th>
<th>ATT_V</th>
<th>INT_V</th>
<th>UTI_V</th>
<th>EFF_C</th>
<th>OPP_C</th>
<th>PSY_C</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELF</td>
<td>.432**</td>
<td>.127**</td>
<td>.028</td>
<td>.400**</td>
<td>.407**</td>
<td>.336**</td>
<td>-.314**</td>
<td>-.215**</td>
<td>-.238**</td>
<td></td>
</tr>
<tr>
<td>MAP</td>
<td>.282**</td>
<td>.131**</td>
<td>.162**</td>
<td>.075**</td>
<td>.108**</td>
<td>.000</td>
<td>.019</td>
<td>.104**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAP</td>
<td>.689**</td>
<td>.128**</td>
<td>0.017</td>
<td>.079**</td>
<td>0.033</td>
<td>0.023</td>
<td>.141**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAV</td>
<td>.591**</td>
<td>.594**</td>
<td>-.421**</td>
<td>-.176**</td>
<td>-.036</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATT_V</td>
<td>.625**</td>
<td>-.373**</td>
<td>-.142**</td>
<td>-.181**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INT_V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTI_V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFF_C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.416**</td>
<td>.220**</td>
</tr>
<tr>
<td>OPP_C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.432**</td>
<td></td>
</tr>
<tr>
<td>PSY_C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>5.14</td>
<td>6.64</td>
<td>6.06</td>
<td>5.80</td>
<td>5.25</td>
<td>4.70</td>
<td>5.03</td>
<td>2.37</td>
<td>2.14</td>
<td>3.67</td>
</tr>
<tr>
<td>SD</td>
<td>0.67</td>
<td>0.54</td>
<td>1.09</td>
<td>1.45</td>
<td>0.59</td>
<td>0.77</td>
<td>0.76</td>
<td>1.01</td>
<td>1.11</td>
<td>1.21</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01.

**Note.** SELF: self-efficacy; MAP: mastery-approach; PAP: performance-approach; PAV: performance-avoidance; ATT_V: attainment value; INT_V: intrinsic value; UTI_V: utility value; EFF_C: effort cost; OPP_C: opportunity cost; PSY_C: psychological cost.
The preliminary CFA estimated factor scores, showing acceptable model fit indicators (TLI = .958; CFI = .965; RMSEA = .035; \( \chi^2 \)(591) = 1542.60; p < .001; For factor loadings and factor correlation, see Table 26 and Table 27 in Appendix). The fit indices of a series of LPA models are reported in Table 8. I started with a solution with one profile as the minimum possible, and then extended through seven profiles based on model improvement. Specifically, the values of AIC, CAIC, BIC, and SABIC decreased as the number of profiles increased. The aLMR remained significant, demonstrating that solution with four profiles was superior to one with three profiles, and the value of Entropy reached a peak at the model with four profiles. Additionally, from a solution with five profiles, a group with relatively small sample size (i.e., n=37, 2.8% of the sample) to the other groups appeared. Importantly, this small group represented similar motivational patterns to another group, demonstrating high self-efficacy, goals, and values, but low cost. The profile continued coming up even when running other models with more than 5. Lastly, the value of Entropy decreased substantially from a solution with six profiles to one with seven profiles. Based on the values of model fit indices and the distinct representation of motivational patterns, therefore, the model with four profiles was shown to be fully satisfactory.

### Table 8. Latent Profile Fit Statistics for Models Based on the Ten Motivation Types

<table>
<thead>
<tr>
<th>Model</th>
<th>LL</th>
<th>Scaling</th>
<th>#FP</th>
<th>AIC</th>
<th>CAIC</th>
<th>BIC</th>
<th>SABIC</th>
<th>aLMR</th>
<th>BLRT</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 profile</td>
<td>-13907.2</td>
<td>1.303</td>
<td>20</td>
<td>27854.42</td>
<td>27896.85</td>
<td>27928.22</td>
<td>27894.69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 profiles</td>
<td>-12301.4</td>
<td>1.567</td>
<td>31</td>
<td>24664.74</td>
<td>24730.60</td>
<td>24825.62</td>
<td>24727.15</td>
<td>&lt;.05</td>
<td>&lt;.05</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>3 profiles</td>
<td>-11704.8</td>
<td>1.939</td>
<td>42</td>
<td>23493.60</td>
<td>23582.75</td>
<td>23711.58</td>
<td>23578.16</td>
<td>.166</td>
<td>&lt;.05</td>
<td>.863</td>
</tr>
<tr>
<td>4 profiles</td>
<td>-11130.1</td>
<td>1.671</td>
<td>53</td>
<td>22366.20</td>
<td>22478.69</td>
<td>22641.27</td>
<td>22472.91</td>
<td>&lt;.05</td>
<td>&lt;.05</td>
<td>.899</td>
</tr>
<tr>
<td>5 profiles</td>
<td>-10820.0</td>
<td>2.302</td>
<td>64</td>
<td>21768.07</td>
<td>21903.84</td>
<td>22100.23</td>
<td>21896.93</td>
<td>.699</td>
<td>&lt;.05</td>
<td>.894</td>
</tr>
<tr>
<td>6 profiles</td>
<td>-10547.0</td>
<td>1.834</td>
<td>75</td>
<td>21244.00</td>
<td>21403.19</td>
<td>21633.24</td>
<td>21395.00</td>
<td>.231</td>
<td>&lt;.05</td>
<td>.893</td>
</tr>
<tr>
<td>7 profiles</td>
<td>-10338.0</td>
<td>1.776</td>
<td>86</td>
<td>20847.91</td>
<td>21030.54</td>
<td>21294.25</td>
<td>21021.06</td>
<td>.122</td>
<td>&lt;.05</td>
<td>.877</td>
</tr>
</tbody>
</table>
The profiles are illustrated in Figure 10. Profile 1 (n = 120; 9.1% of the sample) is characterized by low self-efficacy and achievement goals, and values with high cost. In particular, the values for most motivation types except for performance goals are greater than those in other profiles. That is, this “High cost oriented” profile consists of students who feel that they give up anything while learning. Profile 2 (n = 500; 37.7% of the sample) is characterized by the moderate level of all motivations., students in this profile show moderate self-efficacy, achievement goals, value and cost, which close to 0. This “moderately motivated” profile describes students who are slightly interested in doing better than others and feel that learning leads to giving up somethings. Profile 3 (n = 556; 41.9% of the sample) is characterized by the highest values of self-efficacy, achievement goals, and values among the profiles. In other words, these students believe in their capabilities to study and have their distinct achievement goals of learning in both directions, mastery and performance. In addition, the students highly valued their learning in multiple ways. The “High goals and values oriented” profile represents students who have positive attitudes toward learning and are clearly explain why they would study. Lastly, Profile 4 (n = 150; 11.3% of the sample) is characterized by low performance goals and cost. Students in the profile show similar motivation patterns to profile 3, but performance goals, performance-approach and -avoidance are found to be the lowest among the profiles. In particular, the value of performance-avoidance is highly low compared to other motivation types. Lastly, the level of psychological is shown to be the lowest compared to that in other profiles. This “Low performance goals and cost (pure mastery learners)” demonstrates students who have no interest in outperforming other students and no concern that others would do better than them. In addition, they have no negative attitude toward learning.
Figure 10. Final model with four profiles. SELF = self-efficacy; MAP = mastery-approach; PAP = performance-approach; PAV = performance-avoidance; ATT_V = attainment value; INT_V = intrinsic value; UTI_V = utility value; EFF_C = effort cost; OPP_C = opportunity cost; PSY_C = psychological cost.

Table 9 shows the posterior probability that students belongs to the assigned profile, but not other profiles. Posterior probabilities should be greater than 70% to ensure that students appropriately belong to assigned profiles (Stanley, Kellermanns, & Zellweger, 2017). The values of motivation profiles demonstrate greater than 90%, which means students well fall into target profiles. A posterior probability across profiles refer to the average of posterior probabilities for each profile.

Table 9. Posterior Probabilities and Cross-probability of Motivation Profiles

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Profile 1</th>
<th>Profile 2</th>
<th>Profile 3</th>
<th>Profile 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profile 1</td>
<td>120</td>
<td>93.3%</td>
<td>1.5%</td>
<td>4.6%</td>
<td>0.6%</td>
</tr>
<tr>
<td>Profile 2</td>
<td>500</td>
<td>4.5%</td>
<td>94.6%</td>
<td>0.0%</td>
<td>0.9%</td>
</tr>
<tr>
<td>Profile 3</td>
<td>556</td>
<td>4.5%</td>
<td>0.0%</td>
<td>94.6%</td>
<td>0.9%</td>
</tr>
<tr>
<td>Profile 4</td>
<td>150</td>
<td>1.9%</td>
<td>0.4%</td>
<td>2.7%</td>
<td>95.0%</td>
</tr>
<tr>
<td>Across Profiles</td>
<td>1326</td>
<td>9.2%</td>
<td>37.7%</td>
<td>41.7%</td>
<td>11.4%</td>
</tr>
</tbody>
</table>

Note. Profile 1 = High cost oriented; Profile 2 = Moderate motivated; Profile 3 = High goals and values; Profile 4 = Low performance goals and cost.
To examine differences in each motivational construct by profile, post-hoc comparisons were conducted (see Table 10). All profiles have significantly different levels of mastery-approach, performance-approach, performance-avoidance, attainment value, intrinsic value, and utility value. Profiles 1 and 2 have a similar level of self-efficacy, effort cost, and opportunity cost, while profiles 3 and 4 have opportunity cost and psychological cost.

Table 10. Comparison of Motivation Constructs by Profile

<table>
<thead>
<tr>
<th>Construct</th>
<th>Profile_1 (n=120)</th>
<th>Profile_2 (n=500)</th>
<th>Profile_3 (n=556)</th>
<th>Profile_4 (n=150)</th>
<th>Post hoc comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELF</td>
<td>- .761</td>
<td>- .225</td>
<td>.295</td>
<td>.264</td>
<td>1=2&lt;4&lt;3</td>
</tr>
<tr>
<td>MAP</td>
<td>- .772</td>
<td>- .108</td>
<td>.245</td>
<td>.070</td>
<td>1&lt;2&lt;4&lt;3</td>
</tr>
<tr>
<td>PAP</td>
<td>- .630</td>
<td>.127</td>
<td>.321</td>
<td>-1.111</td>
<td>4&lt;1&lt;2&lt;3</td>
</tr>
<tr>
<td>PAV</td>
<td>- .809</td>
<td>.260</td>
<td>.473</td>
<td>-1.975</td>
<td>4&lt;1&lt;2&lt;3</td>
</tr>
<tr>
<td>ATT_V</td>
<td>- .985</td>
<td>- .278</td>
<td>.419</td>
<td>.164</td>
<td>1&lt;2&lt;4&lt;3</td>
</tr>
<tr>
<td>INT_V</td>
<td>- .961</td>
<td>- .362</td>
<td>.441</td>
<td>.340</td>
<td>1&lt;2&lt;4&lt;3</td>
</tr>
<tr>
<td>UTI_V</td>
<td>-1.050</td>
<td>- .311</td>
<td>.458</td>
<td>.177</td>
<td>1&lt;2&lt;4&lt;3</td>
</tr>
<tr>
<td>EFF_C</td>
<td>.900</td>
<td>.416</td>
<td>- .465</td>
<td>- .383</td>
<td>3&lt;4&lt;2=1</td>
</tr>
<tr>
<td>OPP_C</td>
<td>.573</td>
<td>.357</td>
<td>- .361</td>
<td>- .309</td>
<td>4=3&lt;2=1</td>
</tr>
<tr>
<td>PSY_C</td>
<td>.362</td>
<td>.300</td>
<td>- .225</td>
<td>- .456</td>
<td>4=3&lt;2&lt;1</td>
</tr>
</tbody>
</table>

As a subsequent analysis, a multivariate analysis of covariance (MANCOVA) was conducted to examine how motivation profiles influence academic achievement based on exam scores controlling for the posterior probabilities to accommodate classification error. The result shows that the membership motivation profiles significantly influence academic achievement (F(12, 3381) = 3.370, p < .001, λ = .969, partial \( \eta^2 = .010 \)). The following figure 11 illustrates differences in scores on each exam. Specifically, for all exam, the lowest scores were found in the High-cost group, whereas the Low performance and cost group showed the highest academic achievement. Also, students in High goals and values and the Low performance and cost performed significantly better than the other two groups, High cost and Low performance goals, throughout the semester.
The result of the multinomial logistic regression investigating relationships between demographic information and the likelihood of memberships of each motivation profile are reported in Table 11. The predictor of first-generation (first-generation=1 & no first-generation=0) was not significant for all comparison, so was not reported (i.e., for profile 1 vs 2, \(B=-.070, .05>p\); for profile 1 vs 3, \(B=.061, .05>p\); for profile 1 vs 4, \(B=.033, .05>p\); for profile 2 vs 3, \(B=.131, .05>p\); for profile 2 vs 4, \(B=.104, .05>p\); for profile 3 vs 4, \(B=-.028, .05>p\)).

Gender and ethnicity were included in the analysis as categorical variable (male=0 & female=1; underrepresented=1 & non-underrepresented=0), whereas age was treated as a continuous variable.
Table 11. Results of Multinomial Logic Regressions for the Effects of Predictors on Motivation Profile Membership

<table>
<thead>
<tr>
<th></th>
<th>Profile 1 vs. 4</th>
<th>Profile 2 vs. 4</th>
<th>Profile 3 vs. 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef. (SE) OR</td>
<td>Coef. (SE) OR</td>
<td>Coef. (SE) OR</td>
</tr>
<tr>
<td>Age</td>
<td>-.073 (.031) .930*</td>
<td>-.071 (.021) .932**</td>
<td>-.017 (.018) .983</td>
</tr>
<tr>
<td>Under-represented</td>
<td>-.311 (.255) .733</td>
<td>-.250 (.193) .779</td>
<td>.144 (.190) 1.115</td>
</tr>
<tr>
<td>Gender</td>
<td>-.073 (.257) .930</td>
<td>.441 (.200) 1.554*</td>
<td>0.470 (.196) 1.599*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Profile 1 vs. 3</th>
<th>Profile 2 vs. 3</th>
<th>Profile 1 vs. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coef. (SE) OR</td>
<td>Coef. (SE) OR</td>
<td>Coef. (SE) OR</td>
<td>Coef. (SE) OR</td>
</tr>
<tr>
<td>Age</td>
<td>-.055 (.028) .946*</td>
<td>-.053 (.016) .948**</td>
<td>-.002 (.029) .998</td>
</tr>
<tr>
<td>Under-represented</td>
<td>-.455 (.210) .635*</td>
<td>-.394 (.128) .674**</td>
<td>-.060 (.213) .941</td>
</tr>
<tr>
<td>Gender</td>
<td>-.542 (.215) .581*</td>
<td>-.029 (.142) .972</td>
<td>-.514 (.217) .598*</td>
</tr>
</tbody>
</table>

*Note. Profile 1 = High cost oriented; Profile 2 = Moderately motivated; Profile 3 = High goals and values; Profile 4 = Low performance goals and cost. *p < .05. **p < .01.

As students get older, they were more likely to be in the profile High goals and values and Low performance goals than profiles in the High cost oriented and Moderately motivated profiles, which means older students tend to have higher self-efficacy, mastery-approach, and values, but low cost than younger ones. In terms of ethnicity, underrepresented ethnic minority students showed the higher likelihood of being in the High goals and values group than High cost oriented and Moderately motivated groups. This result suggests that compared to Caucasian and Asian students, underrepresented ethnic minority students tended to be more motivated by higher goals and values than high cost or high performance goals. Lastly, male students are more likely to be members of High cost oriented group than Moderately motivated and High goals and values groups, but they are also more likely to be members of Low performance goals group than High cost oriented and Moderately motivated groups. This result suggests that male students are more likely to be motivated by higher performance goals.
B. Metacognition profiles

To investigate the patterns of metacognition behaviors, another profile analysis was conducted with metacognition learning behaviors including monitoring learning through self-assessment quizzes, monitoring performance by checking grades, and planning using review papers or blueprints. Model fit indices of a series of LPA models including one through seven profiles examined (see Table 12). The values AIC, CAIC, BIC, and SABIC decreased as the number of profiles increases and aLMR remained significant up to the model with five profiles. However, one of the profiles in the four-profile and five-profile models respectively described only 5.2% (n = 69) and 2.9% (n = 39) of the sample. Additionally, comparing solutions with three and four profiles, I found, in the solution with four profiles, both seemed to be similar and only one difference between them is the small different value of monitoring learning behaviors, which theoretically makes less sense. Therefore, considering the representation of profiles based on theories and the availability of further analyses, the model with three profiles was considered optimal.

Table 12. Latent Profile Fit Statistics for Models Based on the Three Metacognition Behaviors

<table>
<thead>
<tr>
<th>Model</th>
<th>LL</th>
<th>Scaling</th>
<th>#FP</th>
<th>AIC</th>
<th>CAIC</th>
<th>BIC</th>
<th>SABIC</th>
<th>aLMR</th>
<th>BLRT</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 profile</td>
<td>-21115.3</td>
<td>1.348</td>
<td>6</td>
<td>42242.60</td>
<td>42255.34</td>
<td>42273.74</td>
<td>42254.68</td>
<td></td>
<td></td>
<td>.947</td>
</tr>
<tr>
<td>2 profiles</td>
<td>-20782.8</td>
<td>1.396</td>
<td>10</td>
<td>41585.57</td>
<td>41606.83</td>
<td>41637.47</td>
<td>41605.71</td>
<td>&lt;.05</td>
<td>&lt;.05</td>
<td>.899</td>
</tr>
<tr>
<td>3 profiles</td>
<td>-20610.9</td>
<td>1.481</td>
<td>14</td>
<td>41249.83</td>
<td>41279.52</td>
<td>41322.49</td>
<td>41278.02</td>
<td>&lt;.05</td>
<td>&lt;.05</td>
<td>.898</td>
</tr>
<tr>
<td>4 profiles</td>
<td>-20484.2</td>
<td>1.407</td>
<td>18</td>
<td>41004.39</td>
<td>41042.61</td>
<td>41097.81</td>
<td>41040.63</td>
<td>&lt;.05</td>
<td>&lt;.05</td>
<td>.913</td>
</tr>
<tr>
<td>5 profiles</td>
<td>-20403.5</td>
<td>1.357</td>
<td>22</td>
<td>40850.92</td>
<td>40897.7</td>
<td>40965.10</td>
<td>40895.21</td>
<td>&lt;.05</td>
<td>&lt;.05</td>
<td>.913</td>
</tr>
<tr>
<td>6 profiles</td>
<td>-20354.5</td>
<td>1.443</td>
<td>26</td>
<td>40760.93</td>
<td>40816.19</td>
<td>40895.87</td>
<td>40813.28</td>
<td>.060</td>
<td>&lt;.05</td>
<td>.904</td>
</tr>
<tr>
<td>7 profiles</td>
<td>-20294.7</td>
<td>1.851</td>
<td>30</td>
<td>40649.46</td>
<td>40713.08</td>
<td>40805.16</td>
<td>40709.86</td>
<td>.591</td>
<td>&lt;.05</td>
<td>.917</td>
</tr>
</tbody>
</table>
Figure 12 illustrates metacognition behaviors in the final model with three metacognition profiles. We labeled each profile respectively as infrequent metacognitive processing (n=994; 75.0% of the sample), checking performance and planning (n=199; 15.0% of the sample), and self-assessment (n=133; 10.0% of the sample) based on use patterns of metacognitive behaviors. The interesting point is that students demonstrated a clear preference for a distinct subset of resources design to support metacognitive activities. As shown in Figure 12, students in the self-assessment group engaged in monitoring their learning, whereas students in the checking performance and planning group showed high checking performance and planning but a much lower level of self-assessment. Therefore, it can be assumed that only a few students use each metacognition behaviors at the similar level. Upon inspection of individual student data, only 11 students, 0.8% of the sample, demonstrated frequent use (>1SD) of all tools designed to support metacognition. Furthermore, they showed better performance than other groups, although there was no significant difference due to small samples.

![Figure 12. Final model with three metacognition profiles.](image)

Like motivation profiles, posterior probabilities associated with each metacognitive learning profile and cross-probability are shown in Table 13. The posterior probabilities of each profiles demonstrate greater than 90%, which means students are clearly classified into profiles
without the high probability of belonging to more than one profiles. Additionally, posterior probabilities across profiles are shown to be similar to the actual proportion of students in each profile.

Table 13. Posterior Probabilities and Cross-probability of metacognitive learning profiles.

<table>
<thead>
<tr>
<th>Profile</th>
<th>n</th>
<th>Profile 1</th>
<th>Profile 2</th>
<th>Profile 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profile 1</td>
<td>994</td>
<td>97.0%</td>
<td>2.1%</td>
<td>0.9%</td>
</tr>
<tr>
<td>Profile 2</td>
<td>199</td>
<td>9.3%</td>
<td>90.2%</td>
<td>0.6%</td>
</tr>
<tr>
<td>Profile 3</td>
<td>133</td>
<td>5.4%</td>
<td>1.6%</td>
<td>93.0%</td>
</tr>
<tr>
<td>Across Profiles</td>
<td>1326</td>
<td>74.6%</td>
<td>15.3%</td>
<td>10.1%</td>
</tr>
</tbody>
</table>

*Note. Profile 1 = Infrequent metacognitive processing; Profile 2 = Checking performance and planning; Profile 3 = Self-assessment.*

Figure 13 shows differences in scores on each exam by metacognition profile. A MANCOVA showed that metacognition profile influences overall academic achievement controlling for the posterior probabilities to accommodate classification error (F(8, 2558) = 5.709, p < .001, λ = .965, partial η² = .018). Also, students in Checking performance and planning and Self-assessment groups showed similar academic achievement on each exam, whereas students in the Infrequent metacognitive processing group showed significantly lower academic achievement than other two groups.
Table 14 shows count and expected count of motivation and metacognition profiles. Specifically, students in the High cost-oriented group showed less checking performance and planning than those in the Low performance and cost.

But interestingly, they self-assessed their learning more than those in the Low-values and High-goals and value groups. Another noteworthy point is that the relatively small number of students in the Low performance and cost belongs to the Weak metacognition group than other motivation profiles. Lastly, High goals and values students also showed slightly more checking performance and planning and self-assessment than expected (see Table 14).

**Table 14. Contingency Table of motivation and metacognition profiles**

<table>
<thead>
<tr>
<th>Motivation Profiles</th>
<th>Metacognition Profiles</th>
<th>Weak Metacognition</th>
<th>Checking performance and planning</th>
<th>Self-assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Cost-oriented</td>
<td>Count</td>
<td>96</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Expected Count</td>
<td>90.0</td>
<td>18.0</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>% of Expected</td>
<td>106.7%</td>
<td>55.5%</td>
<td>116.3%</td>
</tr>
<tr>
<td>Low values</td>
<td>Count</td>
<td>383</td>
<td>70</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>Expected Count</td>
<td>374.8</td>
<td>75.0</td>
<td>50.2</td>
</tr>
<tr>
<td></td>
<td>% of Expected</td>
<td>102.2%</td>
<td>93.3%</td>
<td>93.7%</td>
</tr>
<tr>
<td>High goals and values</td>
<td>Count</td>
<td>417</td>
<td>84</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Expected Count</td>
<td>416.8</td>
<td>83.4</td>
<td>55.8</td>
</tr>
<tr>
<td></td>
<td>% of Expected</td>
<td>100.1%</td>
<td>100.7%</td>
<td>98.6%</td>
</tr>
<tr>
<td>Low performance and cost</td>
<td>Count</td>
<td>98</td>
<td>35</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Expected Count</td>
<td>112.4</td>
<td>22.5</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td>% of Expected</td>
<td>87.2%</td>
<td>155.5%</td>
<td>113.0%</td>
</tr>
</tbody>
</table>

To figure out the relations between the four motivation profiles and the three metacognition profiles, multinomial logistic regression was carried out controlling for posterior probabilities to accommodate classification errors. Since the nominal predictor, motivation profiles, has four categories, each category had to be used as a reference group to examine odds ratios between all categories along with a change in a reference category of the dependent
variable with three categories. All significant results emerged when the *low performance goals and cost* group (profile 4) was used as a reference category (see Table 15).

Compared to the *low performance goals and cost* group, students in the *High cost* group were more likely to be a member of *self-assessment* group than *checking performance and planning* as well as of a member of an *infrequent metacognitive process* than *checking performance and planning*. In other words, when the motivation profile of students changes from high goals and value to high cost, they are more likely to do no metacognitive process or self-assess their learning than checking grades and planning together.

In addition, students in *high performance and goals* and *high goals and values* groups relative to the *low performance goals and low cost* group were more likely to be members of the *infrequent metacognitive process* than *checking performance and planning*. This result demonstrates that students in the *low performance goal and cost* group are more likely to check grades and make a learning plan than infrequent metacognitive process compared to other groups including the *high cost* group.

**Table 15. Results of Multinomial Logistic Regression for the Effects of Motivation Profiles on Metacognition Profile Membership**

<table>
<thead>
<tr>
<th>Reference Motivation Profiles</th>
<th>Metacognition Profiles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Profile 1 (Infrequent) vs. 2 (Checking performance and planning)</td>
</tr>
<tr>
<td>Low performance goals and cost</td>
<td>Coef (SE)</td>
</tr>
<tr>
<td>High Cost</td>
<td>1.23 (.39)</td>
</tr>
<tr>
<td>Moderately motivated</td>
<td>.68 (.24)</td>
</tr>
<tr>
<td>High goals and values</td>
<td>.57 (.23)</td>
</tr>
<tr>
<td>High goals and values</td>
<td>.66 (.35)</td>
</tr>
<tr>
<td>Moderately motivated</td>
<td>.10 (.18)</td>
</tr>
<tr>
<td>Moderately motivated</td>
<td>.60 (.36)</td>
</tr>
</tbody>
</table>
Next, changes in specific metacognition behaviors by motivation profile were examined. The following three figures (14, 15, & 16) demonstrate a change in each metacognition behavior including monitoring learning, monitoring performance, and planning over a semester. In terms of monitoring learning, all students show metacognition behaviors that are more active from one week before each exam than between exams. However, students in the High-cost group used less self-assessment than other groups in preparation for exam 2 and exam 3. Also, what is of interest is that for one week after each exam, students rarely assessed their learning through self-assessment quizzes.

**Figure 14. Changes in Monitoring Learning over a semester**

The study handles five consecutive semesters so there are small discrepancies in timing. For example, in 2014 fall, exam 1 was provided on day 30, whereas on day 34 in day 2015 spring. Therefore, for the improved comparison, time windows were calculated based on exact days in each semester. Time was divided into three categories, checking and preparation (within 6 days after each exam; Labeled as 1), no-press period (between 7 days after and before each exam; Labeled as 2), and cram week (before 6 days each exam; Labeled as 3). Also, to examine how the existence of exam influence metacognition behaviors, time windows was calculated based on days*exam. For example, “Learning 1.3” demonstrated how many times students monitor their learning for one week in the preparation for exam 1. Also, Performance 2.1 showed how many times students check their grades for one week after exam 1, not exam 2.
In addition, students in the *high cost* group least checked their grades compared to other profiles throughout a semester, whereas students in the *low performance and cost* more frequently used the tool. Unlike the self-assessment tool, “My grades” to check grades was used more constantly regardless of exams. Interestingly, all students showed a great deal of interest in their grades before a final exam, even including the *high cost* profile.

![Figure 15. Changes in Monitoring Performance over a Semester](image)

Lastly, in terms of planning, just students in the *Low performance and cost* group started planning their learning earlier than other groups. After then, the high cost group more frequently accessed planning-related materials by exam 1. During the rest of the semester, although the groups of high cost and low performance goals and cost slightly less used the features, in general, use patterns by each group were shown to be similar.
Figure 16. Changes in Planning over a Semester

5. Discussion

A. Motivation profiles

The result of motivational profiles is explained based on two approaches to achievement goals, a mastery goal perspective and a multiple goals perspective. According to the mastery goal perspective, high mastery and low performance goals should produce the most adaptive learning, whereas in the multiple goals perspective, high mastery and high performance goals should lead to the best outcomes (Senko, 2016).

This study employed an LPA to identify naturally occurring motivational profiles and produced a four-profile solution. The groups of moderately motivated and low performance goals and cost are explained by a mastery goal perspective where high mastery and low performance have the most adaptive learning pattern (Pintrich, 2000). In this theory, mastery goals orient students to adaptive outcomes such as high self-efficacy, task value, and metacognitive learning strategies, whereas performance goals are shown to be maladaptive regarding learning strategy use and performance (Dweck, 1986; Dweck & Leggett, 1988; Schunk et al., 2012).
Specifically, *low performance goals* showed a mastery-oriented pattern that is characterized by high mastery goals and low performance goals, specifically achievement and avoidance. In addition to mastery goals, students in this profile are motivated by higher self-efficacy and values. Therefore, that this group demonstrates the highest academic achievement is consistent with previous studies in which self-efficacy and task values are positively related to performance (Hulleman et al., 2008; Pintrich, 1999). Considering the students has low performance goals, students in the profile could be considered pure learners who focus mastery of learning contents without interests in demonstrating competence to others. Therefore, they can use more effective learning strategies such as deep learning, note-taking, and seeking more information (Bernacki et al., 2012; Dupeyrat & Mariné, 2005). In addition, according to Elliot (1999), who proposed the incorporation of approach-avoidance distinction to performance goals, performance approach demonstrates more adaptive patterns than performance avoidance. Therefore, a perspective of achievement goal profiles, the profile of the *low performance goals* is highly oriented to predict positive outcomes.

In contrast, the profile of *Moderately motivated* demonstrates a helpless pattern, showing the pursuit of performance goals and low self-efficacy (Dweck, 1986). The low self-efficacy also might lead to low mastery approach and insufficient deep learning processes, which might result in low performance (Phan, 2010). According to Bernacki and his colleagues (2012), performance avoidance is negatively related to learning behaviors, and therefore, these students might not study in an effective way. Also, performance approach is also associated with shallow strategies such as rote learning or memorization (Dupeyrat & Mariné, 2005; Phan, 2010).

Unlike the two profiles above, motivational patterns in high cost and high goals and *values* show similar patterns of mastery goals and performance goals. Therefore, they could be
explained by the multiple goal perspective where mastery goal and performance goals are positively correlated, and sometimes performance goals are more beneficial than mastery goals (Harackiewicz et al., 1998). Barron and Harackiewicz (2001) suggested four hypotheses in explaining positive combined effects of mastery goals and performance goals.

Barron and Harackiewicz (2001) have suggested four different hypotheses to explain how multiple goals yield positive effects. Four hypotheses include (a) additive goals where mastery and performance goals are independent, (b) interactive goals where mastery and performance goals interact, (c) specialized goals where each goal has their effects, and (d) selective goals where achievement goals vary depending on the situation.

The profile of high goals and values shows the multiple goals pattern. Also, this group is motivated by high self-efficacy and values. Therefore, from a perspective of quantity, this group is the most motivated (Van den Broeck, Lens, De Witte, & Van Coillie, 2013). Also, according to Vrugt and Oort (2008), similar to mastery goals, performance approach also positively related to various types of learning strategies such as deep cognitive strategies and metacognitive strategies. However, students in the low performance goal group demonstrated better performance than those in high goals and values group. Based on the post-hoc comparisons (see Table 4), we suspect that performance avoidance might prevent students from using effective learning strategies such as deep strategies and metacognition process (Dupeyrat & Mariné, 2005; Vrugt & Oort, 2008).

Lastly, the profile of high cost shows the highest level of all kinds of cost with low value of the rest. Cost refers to what ones have to give up by learning something (Wigfield & Cambria, 2010). For example, if one feels cost more than their expected value, they do not want to pursue the task and to engage in tasks (Eccles, 2005). Therefore, the high level of cost might lead to
negative emotions and disengagement in learning, which further influences academic achievement. Also, this group is similar to the status of amotivation in self-determination theory in which students are not willing to engage in learning, and they are also not more motivated intrinsically or extrinsically (Ryan & Deci, 2000). Therefore, this low motivation and maladaptive patterns led students to the lowest performance.

**B. Motivation profiles by demographic information**

The result revealed that older students are more likely to be in profiles 3 and 4 than profiles 1 and 2, which means they are motivated by higher self-efficacy, mastery-approach, and values, and lower cost than younger ones. It could be explained that learners with more experience are more likely to be motivated positively. However, this result is seemingly contradicted by relevant literature in which with age, students increasingly have poor performance in school, which lead them to perceive lower values of the subject to protect their self-worth (Wigfield & Eccles, 1992; Skaalvik, 1997). The incremental view of the ability tends to change into more entity-like view of ability (Dweck, 2006). However, most of the research on age difference in motivation has been conducted with children such as elementary students (Dweck & Leggett, 1988; Eccles, Wigfield, Harold, & Blumenfeld, 1993). Therefore, further research is needed with more appropriate samples such as students in higher education.

Regarding ethnicity, underrepresented students showed a higher likelihood of being in the group characterized by higher values (profile 3). However, this result is not consistent with most of the previous research in where underrepresented students such as Hispanic or African Americans do not value their learning tasks in the same manner as Caucasian or Asian students (Graham et al., 1998; Mickelson, 1990). According to Hines (2003), students would have a higher value on tasks through the evidence of successes of people similar to them, which means
that in general minority students have relatively fewer opportunities for the experience due to small members. However, the university in this research is one of the most ethnically diverse schools. Therefore, the underrepresented students in the study might less or even not experience disadvantages previous research mentioned.

From the result of gender differences in the membership of profiles, it was shown that females were more likely to be motivated by higher performance goals than males, meaning they have more desire to outperform others. According to Dweck (1986), female students tend to avoid challenging task and a great deal of debilitation when experiencing failure. Therefore, it could be suggested females are more likely to performance goals than males. However, the results of studies on gender differences have been inconsistent, and some studies demonstrated there is no significant difference in achievement goals (Harackiewicz, Barron, Tauer, & Elliot, 2002). In addition, subject domains and age might influence gender differences in achievement goals as values (Wigfield & Eccles, 1992; 2000).

C. Metacognitive behavior profiles

Most of the students, 75% of the sample, fell into the infrequent metacognitive process group, meaning that many students do not use metacognitive learning process effectively to manage their learning. There may be some reasons for the low rate of using metacognition supportive tools. Firstly, some students might not realize that the metacognition supportive tools were provided intentionally by instructors in the LMS. Second, students might not know the effectiveness of metacognition learning strategies. Therefore, they noticed there are some metacognition supportive tools provided by instructors in the system, but they might not have thought that would be helpful for their academic achievement.
According to Winne and Jamieson-Noel (2002), a successful self-regulation process requires equivalence between the perceived use of learning strategies and actual use of them, but in general, students have a positive bias in which they are overconfident about their learning. In this sense, many of students in the infrequent metacognitive process group might think they used metacognition supportive tools enough by planning and monitoring their learning. The more serious problem is that learners who overestimated their leaning would not notice learning problems and not repair them (Winne, 2004).

In terms of the self-assessment group, some users interested in self-assessment quizzes use this tool repeatedly until they think the learning tasks seems to be mastered (Butler, 1993). Therefore, some students show a metacognition process pattern focusing on self-assessment quizzes without other learning strategies. This pattern leads to another profile, checking grades and planning, who also demonstrated the biased use of metacognitive process tools. Academic achievement by metacognition profile indicates that use of metacognition, that is biased, improve learning outcomes in students in the infrequent metacognitive process group show the lowest achievement. However, a few students, 0.8% of the sample, used all types of metacognition supportive tools, and they demonstrated the highest performance on all exams. Therefore, based on this result, we hypothesize that the use of comprehensive metacognitive process would be more helpful than biased use.

D. Relationships between motivation profiles and metacognition profiles

Compared to the low performance and cost group, students in the high cost group are more likely to use self-assessment quizzes than checking grades and planning. This result is contradicted by general motivation theories in which learners who are low motivated demonstrate less use of effective learning strategies and less engagement in learning than those
who are highly motivated (Ryan & Deci, 2000). In terms of this result, we can think of motivation dynamics in that some researchers considered motivation personality traits, but this assumption ignores the nature of motivation that learners can be motivated differently depending on time or context (Schunk et al., 2012). Although few studies attempted to investigate changes in motivational profiles over time, they focused on the singular motivation theory such as self-determination theory (Gillet et al., 2017; Martinent & Decret, 2015) or achievement goals (Litalien et al., 2017). Therefore, for the better understanding of changes in motivation, it is needed to conduct longitudinal studies based on the integrative perspective of multiple motivational theories.

Also, little research on cost has been conducted (Andersen & Cross, 2014; Barron & Hulleman, 2015; Buehl & Alexander, 2005; Chow, Eccles, & Salmela-Aro, 2012), so previous research is not enough to explain this issue. In particular, three types of cost (i.e., effort, opportunity, and psychological) has rarely been studied (Wigfield et al., 2017). Therefore, further research on specific types of cost is necessary to explain how motivational profiles influence academic achievement and use of learning strategies including metacognition.

Next, compared to the motivation profile of low performance goals and value, other groups are less likely to use of all metacognitive supportive tools than checking grades and planning. Specifically, the least motivated profile, high cost, demonstrated a much more likelihood of being in the infrequent metacognitive process group. Overall, mastery-oriented learners more tend to employ adaptive behavior pattern including the use of effective learning strategies than helpless oriented (Dweck & Leggett, 1988), amotivated (Ryan & Deci, 2000), as well as high mastery and high performance goal oriented individuals (Meece & Holt, 1993).
Chapter 4: Examining the Power of Multiple Data Sources in Predicting Academic Achievement in Undergraduate STEM Courses

1. Introduction

The high attrition rate and slow progress through degree programs among undergraduate students in Science, Technology, Engineering, and Mathematics (STEM) disciplines are serious threats to the health of the modern workforce (Dai & Cromley, 2014). These rates are considerably worse for underrepresented students (Kokkelenberg & Sinha, 2010). The large, introductory STEM courses at the beginning of students’ degree programs comprise a challenging stage in degree pursuit and serve as gatekeeper courses where many students fail to meet requirements for future courses. This leads to a decrease in future enrollment (Atkinson, Hugo, Lundgren, Shapiro, & Thomas, 2007). If retention rates are to be improved, it is important that students enrolled in STEM courses do not experience failure that might lead to withdrawal or dropping out. There has been a growing interest in the use of timely deployment of prediction algorithms that can identify students likely to perform poorly, and efforts to help students improve academic outcomes are ubiquitous in higher education (Pistilli, Willis, & Campbell, 2014; Pritchard & Wilson, 2003; Zajacova, Lynch, & Espenshade, 2005).

Therefore, in order to predict students’ success and failure in school, researchers have attempted to develop a prediction model using many different types of variables that derive from different data sources. Firstly, students’ demographic information has been studied as an important predictor of academic success. According to Tinto (1975, 1993), individual characteristics are shown to be important predictors of academic success. Based on Tinto’s work, many researchers have examined demographic information such as age, gender, and ethnicity, and their relationship to achievement (Dennis, Phinney, & Chuateco, 2005; Petty, 2014;
Pritchard & Wilson, 2003). In particular, the information collected before college enrollment has allowed researchers to predict students’ academic success earlier than other types of data that accrue at the university (e.g. transcripts).

Secondly, in the field of educational psychology, many studies have investigated relations between motivation and achievement. This approach informs motivation theories but also misses an opportunity to use data to identify poor performers and provide interventions in a timely manner. Specifically, researchers in the fields have interests in exploring relations between specific motivational constructs and academic outcomes such as performance or decision (Hulleman et al., 2010). From a workforce development perspective, it is also important to figure out how well students are doing in the course and whether they can get desirable outcomes to pursue their major program as early as possible to provide them with a chance to change their wrong learning strategies. Motivational data may contribute to such a solution.

Researchers who eschew educational psychology theories and focus instead on educational technology are prevalent in the field of educational data mining (Romero, Ventura, Pechenizky, & Baker, 2010). Their research on student activities indicates that, in traditional face-to-face classrooms, it is not easy to track student learning and identify students at-risk (Macfadyen & Dawson, 2010). In these courses, little data are collected as most learning happens in the classroom and with print media. Further complicating matters, assessment is largely through summative exams that are administered through written tests or assignments that evaluate the level of students’ understanding once or twice in a semester. With little behavioral data and sparse assessment data, figuring out student learning processes and providing personalized intervention are challenging (Coates, 2005). When struggling students are identified, the intervention provided by an instructor based on the summative exam are often
superficial and are administered too late in the term to produce changes in students’ behaviors before they achieve an undesirable, summative performance (Pistilli et al., 2014).

Unlike traditional classrooms where a student’s learning behaviors are monitored only via paper-based assessment methods, courses that employ a learning management system (LMS) are able to capture and store students’ learning behaviors that take place within this system, generating a huge amount of data (Macfadyen & Dawson, 2010; Márquez-Vera, Cano, Romero, & Ventura, 2013). These log data captured by systems allows researchers and instructors to discover students at risk by tracing student-initiated uses of the technology and its tools precisely and extensively over time (Romero & Ventura, 2013). In particular, this type of research has received great attention by researchers in educational data mining (EDM) and learning analytics (LA). EDM is concerned with “developing, researching, and applying computerized methods to detect patterns in large collections of educational data that would otherwise be hard or impossible to analyze due to the enormous volume of data within which they exist” (Romero & Ventura, 2013, p. 12). According to the definitions introduced during the 1st International Conference on Learning Analytics and Knowledge (LAK), LA is defined as “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and environments in which it occurs” (https://tekri.athabascau.ca/analytics/).

The inclusion of behavioral data by an LMS adds values beyond demographics and motivation, but most often these data are not aligned to educational theories (Romero, Espejo, Zafra, Romero, & Ventura, 2013). Many previous studies on the prediction of student achievement have been interested in developing methodology and algorithms, while they overlooked educational theories and contexts (Baker & Yacef, 2009). However, this approach is
a missed opportunity, as unaligned treatment of these data relegate them to representing “clicks” on content, which generally measures activities but does not describe how learners engage with course content. A preliminary, collaborative approach between instructors and learning scientists can inform the predictive modeling process and allow for consideration of educational theories in labeling data, interpreting events, and discussing findings with instructors in accessible terms. This process might make prediction models more precise than atheoretical models, and can make it easier for users such as teachers to understand learning phenomena. Such alignment to theory can also make results more interpretable and a more thoughtful selection of interventions could be provided to students according to their need. Therefore, in the present research, the behavioral data I consider are divided into two types by their relationship to educational theory. I examine theory-aligned variables and non-theory-based variables and compare their predictive power in the prediction model.

Finally, self-regulated learning theorists (e.g., Winne & Hadwin, 1998) describe metacognitive skills as procedural knowledge used to regulate learning processes. They suggest that metacognitive monitoring and control process elicits each other within the SRL framework (Winne & Hadwin, 1998), and thus should be viewed as a process composed of multiple events within a sequence. The data by most methods used to measure metacognition are either collected apart from the learning task and in aggregate over long periods of self regulation (e.g., self-report), or at such a fine level of detail that few subjects can be considered and intensive coding and labeling must be conducted before analysis can begin (e.g., and think-aloud protocol). Eventually, these latter “event” methods (Winne & Perry 2001) capture segmented information by which researcher can identify events, but even then, researchers might have difficulty sampling a sufficient volume of instances where sequences contain theorized event
combinations, limiting the capture of metacognitive learning processes. Therefore, the recent research in SRL that emphasizes the investigation of learning as patterns of events rather than just quantity based on the frequency of access to features, which enable researchers to capture the whole process of the learning and better understand metacognitive learning process (Bannert et al., 2014; Winne, 2014).

Though prior researchers have drawn on multiple data sources in prediction models, few studies have investigated these variables together in a prediction model. Fewer still systematically compare the accuracy achieved by models with different data set from multiple sources. To date, there has been no intentional effort to compare the effectiveness of each data type in predicting student achievement, nor sufficiently rich data models that can examine the power of combining these data sources to predict student achievement. This study examines how well individual and combined data sources can explain variance in students’ academic achievement through a systematic comparison of prediction models. Analyses further investigate the difference in predictive powers of variables across multiple course contexts – where learning topics, objectives, materials, and the assessment of student knowledge all differ – to examine the robustness of predictive power across early biology and mathematics courses.

2. Theoretical Background

A. Data types used in predicting academic achievement

Research on the prediction of academic achievement including GPA, performance, and success or retention has been conducted in divergent fields. However, each field has focused on different sets of predictors on their target outcomes. Past research on relations to achievement metrics is summarized below.

i. Student Demographics
Demographic information is one of the most commonly used predictors of researchers in higher education (Pritchard & Wilson, 2003; Zajacova et al., 2005). In general, Caucasian students are less likely to drop out of college than African-American or Hispanic students (Liu & Liu, 1999). Regarding age, older students have higher GPAs than younger students (Zajacova et al., 2005). However, these results sometimes are different or non-significant depending on subject or environment. In a research conducted in a nursing department (Walker, 2016), age was not significantly related to retention in the first semester. Also, there was no difference in academic performance between Caucasian students and others (Morris, 2016). Usually, their purpose of the research is to identify influential factors on academic success or failure such as dropping out and in turn, to prevent them by providing interventions in a timely manner.

Recently, much of research in EDM or LA included demographic characteristics such as gender, age, and ethnicity in developing prediction models (Dejaeger, K., Goethals, Giangreco, Mola, & Baesens, 2012; Guruler, Istanbullu, & Karahasanc, 2010). Many types of demographic information were found to be significant predictors of performance and choices in STEM disciplines (Moakler & Kim, 2014; Wang, 2013). Although demographic information has been used alone in the EDM and LA fields (Dejaeger et al., 2012; Guruler et al., 2010), sometimes, studies includes additional data types such as behavioral data in the prediction modeling with data mining techniques (e.g., Jayaprakash, Moody, Lauría, Regan, & Baron, 2014).

According to the relevant literature, students from the most deprived areas performed less well than more affluent students. Asian and black students performed less well than white students. Female students performed better than their male counterparts. Contrasting with past research, though school performance was positively associated with entry grades, students from
low performing schools were more likely to achieve the highest degree classifications (Thiele, Singleton, Pope, & Stanistreet, 2016).

Guruler and his colleagues (2010) investigated how the various types of demographic information influence students’ grade point averages (GPA) using a decision tree. The predictors included registered information (e.g., registered city, etc.), high school information (e.g., diploma degree, etc.), family’s living conditions. In the case in which dependent variable was whether GPA was more than 2.0 (=1) or not (=0), the number of the years at school was the most important factor in predicting a student’s academic achievement, and the second most effective variable was receiving grants to pay tuition fee.

ii. Students’ Achievement Motivation

Motivation is commonly studied as fundamental to learning and a predictor of achievement by researchers in educational psychology (Bong, 2001; Harackiewicz et al., 2000). Motivation is a fundamental component of the learning process. In education, motivation is positively associated with engagement, performance, persistence, and choice, and therefore, researchers have tried to determine how motivation influences them in a variety of educational settings (Schunk et al., 2012). In other words, motivation provides learners with energy, direction, and volition needed to achieve academic learning goals (Martin, 2012).

Therefore, many researchers interested in motivation usually aim to investigate relationships between motivation and academic achievement to figure out how motivation influence academic performance such as scores on exams or quizzes. Similar to demographic information, in which multiple variables such as age, gender, ethnicity, and parents’ schooling are analyzed in the prediction process, there have been many motivation theories to explain students’ different types of motivation to learn. Among them, self-efficacy (Bandura, 1977), achievement goals (Elliot & McGregor, 2001), and expectancy-value (Eccles et al., 1983) are the
most widely studied types of motivation in education. In particular, these types of motivations were theorized as important components of motivational beliefs in the self-regulation process (Zimmerman, 2002).

Self-efficacy refers to an individual’s confidence in one’s capability to successfully achieve designated tasks (Zimmerman, 2000a). Students with high self-efficacy are more likely to engage in learning processes and demonstrated better performance (Schunk & DiBenedetto, 2014). Achievement goals are defined as the reason why students engage in learning and are divided by how to deal with competence. Mastery goals focus on developing competence with mastery of learning contents whereas performance goals are concerned with demonstrating talents that are overwhelming against others (Elliot & Murayama, 2008). In general, mastery-approach is found to significantly influence on academic achievement, whereas results of performance-approach have not been inconsistent (Senko, 2016). Subjective task value has four components: 1) intrinsic or interest value, 2) attainment value, 3) utility value, and 4) cost (Eccles et al., 1983; Wigfield & Eccles, 2000). Intrinsic or interest value is the enjoyment individuals get from performing the task, or the subjective interest they have in the subject. Attainment value is the importance to the self of doing well on a task. It is linked with identity and confirming or disconfirming salient aspects of the self. Utility value is how the task relates to future goals and can be seen as capturing more-extrinsic reasons for doing the task. Cost refers to the accumulated negative aspects of engaging in the task, including anticipated emotional states (e.g., performance anxiety), the amount of time and effort required for the task, and the forgoing of engagement in other activities (e.g., gainful employment).

Many studies examine relations between motivation, strategic and metacognitive learning processes, and performance outcomes. For example, an oft-cited study by Pintrich and De Groot
(1990) observed that self efficacy and intrinsic value demonstrated significantly positive relations with cognitive engagement and academic performance. Studies are sufficiently numerous that meta-analyses have been conducted for some motivational theories. The result of the meta-analysis conducted based on 68 studies (Cellar et al., 2011) demonstrated a corrected correlation coefficient ($\rho$) to relationships between achievement goals and task performance including GPA. The coefficient between mastery-approach and task performance was .13, whereas the value of performance-approach was .06.

In addition to relations with academic achievement, there have been many studies that investigated how motivation influence cognitive/metacognitive learning process (Pintrich & De Groot, 1990), in which most of the variables measured by a survey that is a commonly used method in most of the psychological studies as well as other fields (e.g., MSLQ; Pintrich et al., 1993). However, the development of technology has made it possible to capture the variables in different ways in which the researchers can capture and figure out learning processes beyond the previous single-point measurement (Bernacki et al., 2012).

iii. **Student behavior**

With the dramatic increase in the number of students in higher education and their diversity, efficient maintenance and utilization of the data have been one of the primary goals of higher education institutions (Clancy & Goastellec, 2007). The development of learning technology allows the institutions to capture and store the information in a well-organized way and use them to predict academic achievement (Romero et al., 2013). Additionally, this type of the data is easier to access and collect than other types that are difficult to measure (e.g., study habit) or restrictedly accessible due to some reasons such as personal privacy (e.g., prior academic achievement) (Nistor & Neubauer, 2010). In this sense, behavioral data have received considerable attention by researchers after the development of computer supporting learning
systems such as an LMS. Specifically, the information about learning behaviors allows instructors and researchers to track students’ learning processes and examine their changes. Recently, the accessibility of these data has afforded pragmatic attempts to use the data in near real time to inform prediction algorithms that use early behavior to predict achievement (Hu, Lo, & Shih, 2014; Jayaprakash et al., 2014).

This development of learning technology leads to an increase in the amount of data including behavioral data, requiring researchers to employ the new methodology to handle the data. Accordingly, from a methodology perspective, the emergence of EDM and LA provides researchers with various ways to utilize the behavioral data in education (Baker & Inventado, 2014). Many studies conducted by researchers from EDM and LA are concerned with building a prediction model based on the log data by a system such as tutoring systems or LMSs (Jayaprakash et al., 2014). In particular, researchers in emerging fields (e.g., EDM) are sometimes more interested in developing better algorithms by comparing the predictive power of algorithms.

However, these studies focused on the development of better algorithms sometimes miss educational perspectives. In large part, the data used for the studies include activities, and can be interpreted as measuring interactions or engagement at most. This process without the educational consideration is an impoverished approach that limits the way data can be aggregated in ways that better reflect the learning processes. Additionally, studies without the consideration of educational theories and contexts might lead researchers to have troubles interpreting the results and providing effective intervention based on them (Rogers, Gašević, & Dawson, 2015).
B. LMS Behavioral Data and Opportunities for Prediction and Intervention

Among the many technologies available to learners at university, the LMS has become a ubiquitous tool in higher education. Instructors use the platform for communicating with students, conducting critical assessment tasks, and sharing digital resources students can use for learning. When instructors and students make use of LMS features, the system captures a trace of each event in a log file. The trace data thus allow researchers to understand better learning behaviors of students as they provide a rich, fine-grained, and accurate record of students’ actions (Nistor & Neubauer, 2010).

The data collected through log data by an LMS provide some advantages in predicting student achievement. First, the method makes it easier to manage the data and get a large amount of information including frequency, time, and patterns consisting of a series of activities including reading, writing, posting, and taking exams (Black, Dawson, & Priem, 2008). LMSs such as Blackboard Learn to capture and store learning activities of students with time stamps at a fine-grained level, allowing researchers to track a variety of user actions and to take a look at the data in various angles (Krumm, Waddington, Lonn, & Teasley, 2014). Much of the interesting information in this kind of research based on the sophisticated data could not be obtained by other means.

Second, this method is not invasive (Macfadyen & Dawson, 2010). Usually, variables such as motivation or metacognition related to learning are measured by a survey that might distort the data by a mind that tends to overestimate or underestimate student abilities. Similarly, an observation also might influence student learning behaviors as they perceive someone observes their behaviors. Therefore, they tend to demonstrate their classroom activities differently compared to normal situations. However, an LMS can capture students’ natural
learning behaviors because they do not notice their activities are recorded and stored in the system. Therefore, their behaviors are totally the same as the usual.

Third, the log data automatically generated by systems allow researchers to examine students learning behaviors in real time. This advantage of access to the data makes it possible for researchers to build a prediction model and provide in-time interventions based on the real-time data. With the immediacy of the data, researchers and instructors can keep students from having the experience of failure (Marsh et al., 2006). In this sense, it is difficult to obtain this advantage from traditional methods such as a survey.

i. Collection of LMS data using ubiquitous university software

Not only are LMS data ubiquitous on campus, but so are tools that are necessary for information technology (IT) units to host it and serve end users. For example, IT operations units maintain an analysis & support group that monitors the health of university servers and the software they host, and user traffic on those platforms. IT services also include help desks to assist users who encounter trouble with hardware, software, and interactions with the learning materials hosted on an LMS. These kinds of analyses and services require a robust data platform with capabilities to not only log data, but also to enrich these data by labeling common events with values from look up tables, generating alerts when student behaviors represent a known issue (e.g., a hacked account, a log in that might represent cheating, or a known ineffective learning behavior). This provides a unique opportunity to collect raw LMS event data and to leverage IT tools that afford real-time collection, enrich data to reflect theory-aligned learning events (i.e., using lookup tables), and when warranted, to alert students, instructors or learning support units about events students conduct that are likely to lead to poor outcomes.

ii. The importance of enriching raw LMS data
In addition to the insight that can be gained about past learning events, logs of learning events can be combined with achievement data in order to identify (un)productive patterns of events and predict the achievement of future students based on their behavioral match to prior students who achieved certain levels of performance (Jayaprakash et al., 2014; Macfadyen & Dawson, 2010; Romero et al., 2013). Therefore, prediction models developed based on early learning activities in class can be a solution to keep students at-risk from experiencing negative outcomes.

The log data generated by LMS has been used as effective predictors in considerable research on the prediction of achievement (Jayaprakash et al., 2014; Macfadyen & Dawson, 2010; Romero et al., 2013; You, 2016). Romero and his colleagues (2013) used basic learning behaviors logged in the system such as the number of posts and total time to the assignment. These predictors were analyzed by multiple algorithms, and then the researchers found the best algorithm to predict academic achievement, and specifically, the result revealed that the best model predicted the final marks of students with approximately 65% of accuracy. Similarly, You (2016) used LMS log data that captured self-regulated learning behaviors to predict course achievement that measured by combining course exam and final course scores. The log data included study time, session time, late submission, and the number of the message. The result of a hierarchical regression analysis revealed multiple behavioral variables explained 58.1% of the course achievement variance ($R^2=.581$).

However, in developing prediction models of academic achievement, many studies overlooked educational theories and contexts (Baker & Yacef, 2009). From a perspective of learning analytics, interpretation and contextualization of the data are important factors in understanding and improving learning (Agudo-Peregrina et al., 2014). Therefore, the
consideration of educational theories and contexts lead to the better understanding of the learning process by educationally interpreting the data (Xing, Guo, Petakovic, & Goggins, 2015). Specifically, Xing and his colleagues (2015) built a prediction model to identify final performance based on participation in an online course. For a practical and understandable prediction model, they employed activity theory to theoretically quantify students’ online participation, which includes seven interaction components: 1) Subject, 2) Object, 3) Tools, 4) Division of Labor, 5) Community, 6) Rules, and 7) Outcome. According to these components, all learning behaviors were categorized and quantified, and included in modeling processes using Genetic Programming (GP) that is an evolutionary computation technique (for more information, see Koza, 1992). The finding demonstrated 80.2% of overall prediction accuracy and 89.5% of at-risk prediction, which is better results than other similar studies that used learning activities (e.g., Agudo-Peregrina et al., 2014).

According to Rogers and his colleagues (2015), the consideration of educational theories and contexts requires a kind of reappraisal of the cognitive tasks that a course imposes, the kinds of resources provided to students, and their utility for affording opportunities to employ learning strategies. Through theory, we can interpret the results of educational research through such a theoretical lens and understand them better. In this sense, we classify students use of monitoring tools designed based to afford metacognitive learning processes like self assessment and grade checkin through the lens of SRL models (e.g., Winne & Hadwin, 1998). According to Winne and Hadwin (1998), metacognitive learning processes are considered important components in the SRL process in that they manage and regulate students’ learning. Therefore, the result of monitoring behaviors could be understood and interpreted more comprehensively within the framework of the SRL model. This is distinct from navigational behaviors and more general
counts of clicks in learning environments. If an inference cannot be made about the kind of (meta)cognitive process a student undertakes when using a digital resource, these are left unclassified and treated as non-theory-aligned events.

Additionally, designing course based on educational theories allows researchers to take action more effectively. In other words, research on learning where events are described more precisely through theory can generate actionable findings (Rogers et al., 2015). Many practitioner including instructor have challenges utilizing the findings of prediction studies to improve students’ performance in that there is no consideration of educational theory and contexts. This means that the result could not be generalized. In this sense, I investigated the importance of educational theory by providing two types of LMS behavioral data, one designed based on metacognitive learning theory (Winne & Hadwin, 1998), called theory-aligned behaviors, and another provided as default by a system, that is called non-theory-aligned behaviors. When an event that reflects a known learning behavior such as metacognitive monitoring through self-assessment is shown to predict achievement, instructors can understand how they might promote such behavior and make design changes. When events corresponding to an event that does not suggest a single cognitive process (e.g. access a content folder), it is more difficult to determine how to respond when such an event is predictive of achievement.

iii. Complex modeling of event data to represent learning processes.

In addition to logging more precisely described learning events, log data are uniquely suited to capture learning behavior patterns using learner and session identifiers, and timestamps to order them in temporal space. This feature of the log data leads to the application of sequential pattern mining in education, allowing for the investigation of learning processes (Reimann, Markauskaite, & Bannert, 2014). The purpose of this technique is to identify sequential patterns relevant to a target variable based on behavioral data (Bannert et al., 2014). Recently, the focus
of the research on SRL processes has moved from an aptitude perspective to an event perspective (Winne, 2010). From the former perspective, SRL processes are explained by trait or internal mental states such as motivational or metacognitive constructs, but this approach has a limit in capturing the dynamic feature of the learning behaviors. In other words, the direct trace of learner actions is more appropriate in explaining SRL processes than static measures by their interpretation. Therefore, recently, the understanding of SRL processes by events has received more attention than the attitude-based approach in that learning process can be changed easily by the contexts of learners during learning. Accordingly, in this research, sequential patterns based on metacognitive learning behaviors were found, and their contribution to predictive power was examined to see if dynamic sequential patterns predict academic achievement beyond the frequency of access to metacognition supporting tools.

C. The current study

The following are the research questions to be addressed in this research.
1. With what degree of accuracy can models composed solely of demographic, motivation, or behavioral data predict academic achievement?
2. Are there differences in the predictive accuracy of LMS activity models achieved by theory-aligned vs. non-theory-aligned behavioral data?
3. What patterns of theory-aligned behaviors emerge as predictors in behavior-based models?
4. How does the accuracy of predictions and features of predictive models differ across courses with differing instructional design features?

3. Methods

A. Participants

For this research, the sample was collected from two courses, introductory Mathematics and Biology. Students of 448 and 1326 were collected respectively from Mathematics and Biology in 2014 fall through 2016 fall semesters. Both courses were introductory face-to-face
courses designed for freshmen students in preparation for the pursuit of their major. The sample respectively included 48.9% and 72.9% females, 54.1% and 56.8% first-generation students, and 41.6% and 44.5% of underrepresented students in mathematics and biology courses.

B. Measures

Demographic information. Demographic information was collected by a questionnaire administered in week 1 of each course online through the LMS. For the high rate of participation in the questionnaire, students who completed the questionnaire became eligible for additional points. The questionnaire had a variety of items regarding personal information including age, ethnicity, and gender. After the completion of the data collection, most of the data were recorded except for age which was used as a continuous variable. Males were coded as 1 and females as 0, and Caucasian and Asian were coded as 0 and others as 1 which refers to an under-represented minority group. In terms of parents’ educational level, students with parents who had a Bachelor’s degree or above were coded 0 and others as 1 that refers to a first-generation student.

Table 16. Items to measure demographic information

<table>
<thead>
<tr>
<th>Measure</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Please indicate your gender</td>
</tr>
<tr>
<td></td>
<td>☐ Male</td>
</tr>
<tr>
<td></td>
<td>☐ Female</td>
</tr>
<tr>
<td>Gender</td>
<td>How old are you in years?</td>
</tr>
<tr>
<td></td>
<td>(                     )</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>What is your ethnicity? (select all that apply)</td>
</tr>
<tr>
<td></td>
<td>☐ Caucasian</td>
</tr>
<tr>
<td></td>
<td>☐ African American</td>
</tr>
<tr>
<td></td>
<td>☐ Asian American</td>
</tr>
<tr>
<td></td>
<td>☐ Latino/Hispanic</td>
</tr>
<tr>
<td></td>
<td>☐ Other</td>
</tr>
<tr>
<td>First Generation</td>
<td>What is the highest level of education completed by a parent?</td>
</tr>
<tr>
<td></td>
<td>☐ High School or less</td>
</tr>
<tr>
<td></td>
<td>☐ Some college</td>
</tr>
<tr>
<td></td>
<td>☐ Bachelor’s Degree</td>
</tr>
<tr>
<td></td>
<td>☐ Master’s Degree</td>
</tr>
<tr>
<td></td>
<td>☐ Professional degree (medical, dental, law, educational, etc. like a MD, DDS, JD, DPharm, Ed.D.)</td>
</tr>
<tr>
<td></td>
<td>☐ Doctoral Degree (Ph.D.)</td>
</tr>
</tbody>
</table>

Note. ☐ Radio button; ☐ Checkbox
Motivation. In this research, three types of motivation were measured: 1) self-efficacy, 2) achievement goals, and 3) value and cost. A survey was administered at the beginning of the semester to measure student motivation, specifically over week 1 and week 2. Five items with six-point Likert scale (1 = Strongly disagree to 6 = Strongly agree) for self-efficacy were developed based on patterns of adaptive learning scales (PALS; Midgley et al., 2000). Cronbach’s α of the item scales were .90 in Mathematics and .89 in Biology.

For achievement goals, the survey included nine items with seven-point Likert scale (1 = Strongly disagree to 7 = Strongly agree) adapted from Achievement Goals Questionnaire-Revised (AGQ-R) (Elliot & Murayama, 2008). Three items were assigned three sub-scale: mastery-approach (e.g., “My goal is to learn as much as possible”), performance-approach (e.g., “I am striving to do well compared to other students”), and performance-avoidance (e.g., “I am striving to avoid performing worse than others”). Cronbach’s α for each scale ranged from .75 to .88 in Mathematics and from .75 to .84 in Biology.

For value and cost, 24 items with six-point Likert scale (1 = Not at all worthwhile to 6 = Very worthwhile) were adapted respectively based on research by Eccles and Wigfield (1995) and Perez and his colleagues (2014). Four items were assigned for each value (e.g., attainment, intrinsic, and utility value) and cost (e.g., effort, opportunity, and psychological cost). Cronbach’s α for each scale ranged from .72 to .92 in Mathematics and from .76 to .92 in Biology.
Table 17. Nature of Exam 1 in Math and Biology

<table>
<thead>
<tr>
<th>Week of Exam 1</th>
<th>Biology</th>
<th>Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Week 5 or Week 6</td>
<td>Week 4 or Week 5</td>
</tr>
<tr>
<td>Chapters covered by Exam 1</td>
<td>1. An introduction to Anatomy and Physiology</td>
<td>1. Functions and Their Representations</td>
</tr>
<tr>
<td>Objective</td>
<td>2. The Chemical Level of Organization</td>
<td>2. A Catalog of Essential Functions</td>
</tr>
<tr>
<td></td>
<td>3. The Cellular Level of Organization</td>
<td>3. Limit of a Function</td>
</tr>
<tr>
<td></td>
<td>4. The Tissue Level of Organization</td>
<td>4. Calculating Limits</td>
</tr>
<tr>
<td></td>
<td>5. The Integumentary System</td>
<td>5. Continuity</td>
</tr>
<tr>
<td></td>
<td>- To Review anatomical terminology as well as the basic organization of homeostatic regulatory mechanisms, cells and tissues</td>
<td>- To Understand the definition of a function and how it relates to the vertical line test</td>
</tr>
<tr>
<td></td>
<td>- To Find the limit of a function with various techniques</td>
<td></td>
</tr>
</tbody>
</table>

**Behavioral data.** Although both Mathematics and Biology are introductory courses to provide foundational knowledge for their fields, they are designed differently in various ways. Specifically, the first exam was administered in week 5 or week 6 in Biology, but in week 4 or week 5 in Math. Therefore, student activity data within the weeks before exam 1 will be included as potential predictors in order to obtain sufficient accuracy of prediction (i.e., multiple weeks for events to accrue). More information on the first exam by course is summarized in Table 17.

One complication is induced by the number of weeks in a semester before the first exam. The timing for the exam differed across courses, and as such events will be collapsed into three periods that align with students’ study habits. The first period is the first week of the semester, which is labeled a “checking and planning” week when students make a learning plan for the upcoming unit. The second is a “no-press” period in which students can study without any press (e.g., preparation for exams) at their pace because some weeks remain until the day of an exam. The last period is one week before exam 1 when students have to engage in preparing for exam 1 using multiple tools in the LMS to support their learning, which is called a “cram” week (see Figure 9 from Chapter 3).
The university LMS, Blackboard Learn, captures and records student use of materials hosted on course sites. When enriched with sufficient metadata, these machine data can be made to describe learning events conducted by students (Dominguez, Bernacki, & Uesbeck, 2016). According to Marsh and her colleagues (2006), raw data should be organized and combined with a human understanding of the situation to provide better insight into students’ learning. Moreover, the data at the content level would increase the likelihood of overfitting because the use of the content is strongly linked to each class. Therefore, content items need to be grouped into “resource type” based on the metadata incorporated in Splunk, software for searching and analyzing data (see Table 18).

In addition, behavioral data will be divided into two types, theory-aligned and non-theory-aligned variables. That is, behavioral data could be further enriched through theory guided feature engineering to make more informed classifications of LMS activities. Regarding the theory-aligned behavioral data, the LMS provided multiple tools to measure metacognitive learning process. According to Winne and Hadwin (1998, 2008), metacognitive learning processes are based on monitoring and control behaviors. Through self-assessment quizzes, students monitor the mastery of their learning, while they monitor a performance level by checking their grade using “My Grades.” In addition, by organizing their study session, they check the learning process, and their learning progress is monitored by self-scoring progress towards course learning objectives. Lastly, students make a learning plan using review papers or blueprints. Through the tools, metacognitive learning behaviors to manage their learning are captured and recorded, allowing researchers to track changes in the use of them.
Table 18. Features assigned to Resource Type

<table>
<thead>
<tr>
<th>Variable type</th>
<th>Resource Type</th>
<th>Features</th>
<th>Average # of features per course</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Biology</td>
</tr>
<tr>
<td>Non-theory-aligned</td>
<td>Communication</td>
<td>Email, Message, etc.</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Environmental Structure</td>
<td>LMS tools (Help, course navigation)</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td>Link to External Website</td>
<td>Link to outside resources</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>Lecture notes</td>
<td>Class notes (posted by instructor)</td>
<td>19.0</td>
</tr>
<tr>
<td></td>
<td>Policy</td>
<td>Policy and procedure documents</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>Content Folder</td>
<td>Subfolders within content areas</td>
<td>22.2</td>
</tr>
<tr>
<td></td>
<td>Link to Content Area</td>
<td>Main menu links to content areas</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td><strong>Sub Total</strong></td>
<td></td>
<td><strong>65.7</strong></td>
</tr>
<tr>
<td>Theory-aligned</td>
<td>Monitoring Learning</td>
<td>Self-assessment quizzes</td>
<td>20.4</td>
</tr>
<tr>
<td></td>
<td>Monitoring Performance</td>
<td>Synopsis of a students’ grades</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>Monitoring Process</td>
<td>Tool to organize a study session</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Monitoring Progress</td>
<td>Tool to self-score progress towards course learning objectives</td>
<td>15.2</td>
</tr>
<tr>
<td></td>
<td>Planning</td>
<td>Syllabus, Course calendar, schedule, exam guides</td>
<td>29.4</td>
</tr>
<tr>
<td></td>
<td><strong>Sub Total</strong></td>
<td></td>
<td><strong>67.8</strong></td>
</tr>
</tbody>
</table>

As shown in Table 18, there is no big difference in the total number of non-theory-aligned behaviors between two courses, while even more theory-aligned behaviors are provided in Biology than Math. In addition, some metacognitive learning tools, Monitoring Learning, Monitoring Process, And Monitoring Progress were not provided by instructors across semester in Math. Therefore, among theory-aligned behaviors, just Monitoring Performance and Planning were included in the further analytical process for Math.

In addition, regarding non-theory-aligned behaviors, Content Folder and Link To Content Area have no educational information in that both of features play a role as a gate or a bridge to get to actual contents. Therefore, these two variables were also excluded in further analyses.
**Academic achievement.** Academic achievement was measured using final course grades. The letter grades were converted to numeric values according to the school policy that ranges from 4.0 (A) to 0.0 (F).

C. Analysis

Sequence pattern mining will be employed to find the interesting patterns of learning behaviors designed based on metacognitive process (Baker & Yacef, 2009). In this research, I employed Apriori Algorithm, the best known sequential pattern mining algorithm, wherein the algorithm work through two main steps: 1) generate frequent behavior sets, and 2) generate confident sequential patterns from the frequent behavior sets (Liu, 2011). Based on this approach, in this research, frequent behavior sets were generated, which have transaction support that is above minimum support, and then the sequential rules with above minimum confidence among the behavior sets were found.

Here, the support of behavioral sequence A refers to the proportion of sequences that contain A as a subsequence, whereas the confidence of sequence A is the percentage of sequences that contain behaviors a as well as b, and estimation of probability Pr(b|a). Therefore, the values of support and confidence can be calculated by Eq. (1) and Eq. (2) (Najafabadi, Mahrin, Chuprat, & Sarkan, 2017).

\[
\text{Support} (a \rightarrow b) = \frac{\text{Number of transactions which contain } a \text{ and } b}{\text{Number of all transactions in the database}} \tag{1}
\]

\[
\text{Confidence} (a \rightarrow b) = \frac{\text{Number of transactions which contain } a \text{ and } b}{\text{Number of transactions which contain } a} \tag{2}
\]

For example, let us assume there are three sequences, \([a, b, c, d]\), \([a, b, d, e]\), \([a, c, d, e]\) and \([b, c, d, e]\). In this sets of the sequences, \([a, c]\) is the subsequence of \([a, b, c, d]\) and \([a, c, d, e]\), so the support of \([a, c]\) is \(= 2 / 4\) (sequences that contain \([a, c]\) / all sequences) \(= 50\%\). The
confidence of \([a, c]\) can be calculated by dividing 2 (sequences that contain \([a, b]\)) by 3 (sequences that contain \([a]\)), and therefore, the value is 66.6\%. Based on these rules, this process to find all subsequences is repeated until no more frequent sequences are found (Liu, 2011).

According to studies by Winne and Hadwin (1998, 2008), metacognitive learning process should be considered as procedural knowledge used to self-regulate learning so that the investigation of patterns of events can explain how the learning process is regulated. In this research, therefore, Therefore, the most frequently used behavioral patterns were found based on theory-aligned behaviors that rely on metacognitive learning processes in the SRL model (Taub et al., 2014).

After finding sequential patterns, prediction models of academic achievement using a set of regression analyses are with multiple resource types of variables was developed (see Table 22). In terms of the predictive modeling process, I first conducted a set of multiple regression analysis with individual resource types where Model 1 to Model 5 respectively included demographic information, motivation, non-theory-aligned behaviors, theory-aligned behaviors, and behavioral patterns. Then, from Model 6, hierarchical regression analyses were performed by adding resource types to a previous model at each step. It should be noted that for Model 6, the previous model was not Model 5, but Model 1. For example, for Model 6, motivation was added to Model 1, which included demographic information, and then, for Model 7, non-theory-aligned behaviors were added to Model 6. This process proceeded until Model 9 was developed where behavioral patterns are added to Model 8 that includes demographics, motivation, non-theory-aligned, and theory-aligned behaviors. After the completion of the modeling process, all prediction models were compared based on various criteria including R-square \((R^2)\) that demonstrates the amount of variance in a dependent variable explained by independent variables.
The process of the prediction modeling with input data types is summarized in Table 19. For all regression analyses, significant variables were included in the prediction models through a forward selection method.

Lastly, to investigate the difference in predictive powers of variables associated with different course designs, the prediction models were applied to Math and Biology respectively. As shown in Table 18, non-theory-aligned features were shown to be the same for Math and Biology, whereas more theory-aligned features designed based on metacognition theory (Winne & Hadwin, 1998) were included in Biology than Math. Therefore, the effectiveness of theory-aligned features by course design was evaluated by comparing changes in variance between the two courses.

Table 19. Predictive Modeling Process

<table>
<thead>
<tr>
<th>Model</th>
<th>Input data types</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Demographics</td>
<td>Motivation</td>
<td>Non-theory-aligned</td>
<td>Theory-aligned</td>
<td>Behavioral patterns</td>
</tr>
<tr>
<td>Model 1</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 4</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 5</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Model 6</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 7</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Model 8</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Model 9</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Note. All model will be applied to Mathematics and Biology individually.

4. Results

A. Sequential Patterns

From the transformed data, the sequential pattern mining identified 84,254 sessions (i.e., a unit of the transaction) in Biology and 11,004 in Math. Initial behaviors in the first round to
generate candidates include Monitoring Learning, Monitoring Performance, Monitoring Progress, Monitoring Process, and Planning in Biology, while only Monitoring Performance and Planning in Math. The support of each initial item across courses is summarized in Table 20. In Math, Planning is found to be the most frequently included behavior in sessions (36.62%), while Monitoring Learning (58.76%) in Biology.

Table 20. Initial Items

<table>
<thead>
<tr>
<th>Items</th>
<th>Math Count</th>
<th>Support</th>
<th>Biology Count</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monitoring Learning</td>
<td>49511</td>
<td>58.76%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monitoring Performance</td>
<td>964</td>
<td>8.73%</td>
<td>3817</td>
<td>4.53%</td>
</tr>
<tr>
<td>Monitoring Progress</td>
<td>-</td>
<td>-</td>
<td>2768</td>
<td>3.29%</td>
</tr>
<tr>
<td>Monitoring Process</td>
<td>-</td>
<td>-</td>
<td>651</td>
<td>0.77%</td>
</tr>
<tr>
<td>Planning</td>
<td>4044</td>
<td>36.62%</td>
<td>12930</td>
<td>15.35%</td>
</tr>
</tbody>
</table>

To discover the most frequently used sequential patterns, the frequent sequences were generated based on initial items. Many studies demonstrate that the use of metacognitive learning strategies is positively associated with academic achievement (Vrugt & Oort, 2008; Winne and Hadwin, 1998; Zimmerman, 2002). Accordingly, in this research, the sequence pattern mining focused on finding the behavioral patterns the most frequently used by students. Since tools to support metacognitive learning were provided differently by course, the different values of support and confidence were applied to each course. Specifically, on average, 67.8 contents to support metacognitive learning in Biology were provided, whereas 8.3 in Math were used across semesters (for the course design, see Table 18). Accordingly, 5.0% of the minimum support and 10.0% of minimum confidence for Biology, and 1.0% and 40% for Math were used to identify the frequently used sequential patterns. Based on the support and confidence, two frequent sequential patterns were found in each course (see Table 19). The sequential patterns include the
repeated monitoring of learning status using self-assessment quizzes in Biology, which means that students more frequently used self-assessment quizzes to monitor their learning status than other metacognitive tools such as planning. In addition, the patterns of two or three times access to Planning contents were found to be the most frequently used in Math.

Table 21. Sequential Patterns in Biology and Math

<table>
<thead>
<tr>
<th>Course</th>
<th>Pattern</th>
<th>Support</th>
<th>Confidence*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biology</td>
<td>Monitoring Learning → Monitoring Learning</td>
<td>16.0%</td>
<td>27.18%</td>
</tr>
<tr>
<td>Biology</td>
<td>Monitoring Learning → Monitoring Learning → Monitoring Learning</td>
<td>7.97%</td>
<td>13.56%</td>
</tr>
<tr>
<td>Math</td>
<td>Planning → Planning</td>
<td>2.13%</td>
<td>80.48%</td>
</tr>
<tr>
<td>Math</td>
<td>Planning → Planning → Planning</td>
<td>1.11%</td>
<td>42.12%</td>
</tr>
</tbody>
</table>

* Lowest one among possible confidences calculated in the behavioral pattern.
B. Regression Analyses with individual resource types

Table 22. Resource types of Independent Variables

<table>
<thead>
<tr>
<th>Demographic information</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Age</td>
</tr>
<tr>
<td>• Gender</td>
</tr>
<tr>
<td>• Ethnicity (dichotomous to underrepresented group vs. majority)</td>
</tr>
<tr>
<td>• First-generation college student</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Self-Efficacy</td>
</tr>
<tr>
<td>• Mastery-Approach</td>
</tr>
<tr>
<td>• Performance-Approach</td>
</tr>
<tr>
<td>• Performance-Avoidance</td>
</tr>
<tr>
<td>• Attainment Value</td>
</tr>
<tr>
<td>• Intrinsic Value</td>
</tr>
<tr>
<td>• Utility Value</td>
</tr>
<tr>
<td>• Effort Cost</td>
</tr>
<tr>
<td>• Opportunity Cost</td>
</tr>
<tr>
<td>• Psychological Cost</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Behavioral Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Non-theory-aligned</td>
</tr>
<tr>
<td>o Content Folder</td>
</tr>
<tr>
<td>o Environmental Structure</td>
</tr>
<tr>
<td>o Link to Content Area</td>
</tr>
<tr>
<td>o Lecture notes</td>
</tr>
<tr>
<td>o Policy</td>
</tr>
<tr>
<td>• Theory-aligned</td>
</tr>
<tr>
<td>o Monitoring Learning</td>
</tr>
<tr>
<td>o Monitoring Performance</td>
</tr>
<tr>
<td>o Monitoring Process</td>
</tr>
<tr>
<td>o Monitoring Progress</td>
</tr>
<tr>
<td>o Planning</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Behavioral Patterns discovered in Sequence Pattern Mining</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Math</td>
</tr>
<tr>
<td>o Planning – Planning</td>
</tr>
<tr>
<td>o Planning – Planning – Planning</td>
</tr>
<tr>
<td>• Biology</td>
</tr>
<tr>
<td>o Monitoring Learning – Monitoring Learning</td>
</tr>
<tr>
<td>o Monitoring Learning – Monitoring Learning – Monitoring Learning</td>
</tr>
</tbody>
</table>

First, a set of simple linear regression analyses were conducted with individual data resources in each course; 1) Model 1: demographic, 2) Model 2: motivation, 3) Model 3: Non-theory-aligned behaviors, 4) Model 4: Theory-aligned behaviors, and 5) Model 5: Behavioral patterns. In Biology, some predictors from each resource type significantly predicted the
dependent variable, showing the different amount of $R^2$ (see Table 23). Specifically, motivation and theory-aligned behaviors explained more variance than others (i.e., 5% of variance explained by motivation and 6% by theory-aligned behaviors). Along with the greater amount of explained variance, these types of variables included more significant variables, demonstrating six variables of motivation and seven of theory-aligned behaviors. In addition to these types, a behavior pattern, three-time repeated monitoring learning, also significantly explained 3% of the variance. Among demographic information, ethnicity and sex were found to be significantly related to academic achievement with $R^2=.02$ and use of environmental structure in the first week is only variable of non-theory-aligned behaviors showing the smallest explained variance ($R^2=.01$). In summary, variables of all resource types significantly predicted academic achievement, but the amount of the explained variance was different by resource type.
Next, some models including individual data sources explained some variance of academic achievement, but others do not in Math (see Table 24). Specifically, predictor variables of demographic information (i.e., ethnicity and age) explained 3% of the variance of the dependent variable, and motivation variables including performance approach, effort cost, and opportunity cost, account for 5% of the variance that is little more than demographic information. However, other types of variables, behaviors (both of non-theory-aligned and
theory-aligned) and behavioral patterns, have no significant predictors to predict academic achievement.

Table 24. The result of Regression Analyses with Individual Resources in Math $^a$

<table>
<thead>
<tr>
<th>Predictor variables</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• URM b</td>
<td>-.40***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>• Age</td>
<td>-.03*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Motivation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Performance Approach</td>
<td>.10**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>• Effort Cost</td>
<td>-.22***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>• Opportunity Cost</td>
<td>.17**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Non-theory-aligned</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Theory-aligned</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Behavioral Patterns</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.03</td>
<td>.05</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>.03</td>
<td>.04</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$F$</td>
<td>8.50***</td>
<td>7.42***</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* p<.05, ** p<.01, *** p<.001.

$^a$ This table includes only significant predictor variables

$^b$ Under-represented Minority Status (Caucasian & Asian=0, Others=1)

C. Hierarchical Regression Analyses with multiple resource types

Additionally, I carried out a set of hierarchical linear regression analyses with all data types in biology and math. The result for biology demonstrated that each resource type made a significant improvement of $R^2$ (see Table 25). Specifically, in Model 1 with variables of demographic information, ethnicity and Age were significant predictors of academic achievement, accounting for 3% of the variance. In Model 6, six predictor variables (i.e., performance approach & avoidance, Attainment & Intrinsic value, Effort & Opportunity cost) were found to significantly improve the model over Model 1, demonstrating the greatest increase in $R^2$ ($\Delta R^2=6\%$). In Model 7, two variables (Environmental structure and Communication tool in the first week) of non-theory-aligned behaviors were significant predictors of academic achievement. These predictor variables explained additional 1% of the variance on top of
demographic information and motivation. Model 8, with theory-aligned behaviors, was improved by five significant predictor variables (Monitoring Learning, Progress and Performance). With four types of resources, Model 8 explained 13% of the variance, showing 4% of an increase in R² by theory-aligned behaviors. Unlike models with individual resource types, Monitoring performance in the first week and Monitoring progress in the last week were not significant any longer. Lastly, in Model 9, the inclusion of the three-time repeated monitoring learning behavior was found to be significantly improvement of the model over earlier models. Specifically, compared to Model 8 with demographic information, motivation, and behaviors, Model 9 with the additional behavioral pattern demonstrated the improvement of 0.1 of the explained variance, which is the smallest increase in R² together with that of non-theory-aligned behaviors.
Table 25. Result of Hierarchical Regression Analysis in Biology a

<table>
<thead>
<tr>
<th>Predictor variables</th>
<th>Model 1</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
<th>Model 9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• URM b</td>
<td>-.29***</td>
<td>-.38**</td>
<td>-.032***</td>
<td>-.29***</td>
<td>-.29***</td>
</tr>
<tr>
<td>• Sex</td>
<td>.25**</td>
<td>-.04**</td>
<td>.023**</td>
<td>.25**</td>
<td>.24**</td>
</tr>
<tr>
<td><strong>Motivation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Performance Approach</td>
<td>.13**</td>
<td>014**</td>
<td>.13**</td>
<td>.13**</td>
<td></td>
</tr>
<tr>
<td>• Performance Avoidance</td>
<td>-.14***</td>
<td>-.015***</td>
<td>-.13***</td>
<td>-.14***</td>
<td></td>
</tr>
<tr>
<td>• Attainment Value</td>
<td>.17*</td>
<td>.017*</td>
<td>.19**</td>
<td>.19**</td>
<td></td>
</tr>
<tr>
<td>• Intrinsic Value</td>
<td>.11*</td>
<td>.010</td>
<td>.08</td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td>• Effort Cost</td>
<td>-.12**</td>
<td>-.012**</td>
<td>-.12**</td>
<td>-.12**</td>
<td></td>
</tr>
<tr>
<td>• Opportunity Cost</td>
<td>.11**</td>
<td>.011**</td>
<td>.11***</td>
<td>.11***</td>
<td></td>
</tr>
<tr>
<td><strong>Non-theory-aligned</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Environmental Structure (1) c</td>
<td></td>
<td>.10***</td>
<td>.08**</td>
<td>.08**</td>
<td></td>
</tr>
<tr>
<td>• Communication Tool (1)</td>
<td>.08*</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Theory-aligned</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Monitoring Progress (1)</td>
<td></td>
<td>.04*</td>
<td>.04*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Monitoring Learning (2)</td>
<td></td>
<td>.00**</td>
<td>.00*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Monitoring Performance (2)</td>
<td></td>
<td>.04*</td>
<td>.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Monitoring Progress (2)</td>
<td></td>
<td>.03**</td>
<td>.02*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Monitoring Learning (3)</td>
<td></td>
<td>.00**</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Behavioral Patterns</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Repeated Monitoring Learning&lt;3&gt; d</td>
<td></td>
<td></td>
<td></td>
<td>.14**</td>
<td></td>
</tr>
</tbody>
</table>

|                  | .02     | .08     | .09     | .13     | .14     |
| R²               |         |         |         |         |         |
| ΔR²              | -       | .06     | .01     | .04     | .01     |
| Adjusted R²      | .02     | .07     | .09     | .12     | .12     |
| F                | 16.02***| 13.55***| 13.26***| 12.61***| 12.58***|

* p<.05, ** p<.01, *** p<.001.

a This table includes only significant predictor variables

b Under-represented Minority Status (Caucasian & Asian=0, Others=1)

c (1) “checking and planning” week, (2) “no-press” week, (3) “cram” week (see Figure 9 in Chapter 3).

d Three-time repeated self-assessment quizzes

The result of the regression analysis in Math demonstrated adding some types of variables improved the explained variance of the dependent variable (see Table 26). Specifically, the inclusion of motivation variables including performance approach, effort cost, and opportunity cost on top of demographic information allowed for an increase in 5% of the variance. In addition to motivation, monitoring learning in last one week of the first exam, a predictor of theory-aligned behaviors, newly came up and significantly contribute to the
improvement of R-square ($\Delta R^2 = .01$). Although one variable of theory-aligned behaviors, monitoring learning, was shown to significantly predict academic achievement, none of non-theory-aligned variable was significantly related to the dependent variable. In other word, after controlling for demographic information and motivation, any non-theory-aligned behaviors did not make a significant contribution to R-square. In addition, any behavioral patterns discovered sequential pattern mining based on theory-aligned behaviors were not found to significantly predict.

Table 26. The result of Hierarchical Regression Analysis in Math

<table>
<thead>
<tr>
<th>Predictor variables</th>
<th>Model 1</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
<th>Model 9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• URM b</td>
<td>-.40***</td>
<td>-.38**</td>
<td>-.38**</td>
<td>-.35**</td>
<td>-.345**</td>
</tr>
<tr>
<td>• Age</td>
<td>-.03*</td>
<td>-.04**</td>
<td>-.04**</td>
<td>-.04**</td>
<td>-.042**</td>
</tr>
<tr>
<td><strong>Motivation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Performance Approach</td>
<td>.10**</td>
<td>.10**</td>
<td>.09**</td>
<td>.094**</td>
<td></td>
</tr>
<tr>
<td>• Effort Cost</td>
<td>-.23***</td>
<td>-.23***</td>
<td>-.22***</td>
<td>-.221***</td>
<td></td>
</tr>
<tr>
<td>• Opportunity Cost</td>
<td>.16**</td>
<td>.16**</td>
<td>.16**</td>
<td>.158**</td>
<td></td>
</tr>
<tr>
<td><strong>Non-theory-aligned</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Monitoring Learning (3) c</td>
<td></td>
<td></td>
<td></td>
<td>.03**</td>
<td>.026**</td>
</tr>
<tr>
<td><strong>Theory-aligned</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral Patterns</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>.03</td>
<td>.08</td>
<td>.08</td>
<td>.09</td>
<td>.09</td>
</tr>
<tr>
<td>$\Delta R^2$</td>
<td>-</td>
<td>.05</td>
<td>-</td>
<td>.01</td>
<td>-</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>.03</td>
<td>.07</td>
<td>.07</td>
<td>.08</td>
<td>.08</td>
</tr>
<tr>
<td>$F$</td>
<td>8.50***</td>
<td>8.04***</td>
<td>8.04***</td>
<td>8.07***</td>
<td>8.07***</td>
</tr>
</tbody>
</table>

* p<.05, ** p<.01, *** p<.001.

a This table includes only significant predictor variables

b Under-represented Minority Status (Caucasian & Asian=0, Others=1)

c (1) “checking and planning” week, (2) “no-press” week, (3) “cram” week (see Figure 9 in Chapter 3).

5. Discussion

A. What is the relative benefit of data types?

The results of regression analyses with individual resource types demonstrated that predictor variables of demographic information and motivation accounted for the variance in both courses. In addition to them, (both of non-theory-aligned and theory-aligned) behaviors and
behavioral pattern (repeated self-assessment) were shown to significantly predict academic achievement in Biology.

i. **Demographic information**

In terms of the demographic information, the finding is in line with many studies predicting students’ academic achievement, in which demographic information such as sex, age, ethnicity and socio-economic status (SES) was a significant predictor of academic achievement (Thiele et al., 2016). Specifically, the amount of $R^2$ was .03 in Math and .02 in Biology, which is not much greater than other predictor variables. Additionally, significant variables were different by course. Similarly, in a study to predict students’ withdraw (Woodfield, 2017), different variables of demographic were found to be significant in biology, business, education, and psychology courses (i.e., gender was significant in psychology and business courses, whereas ethnicity in only biology). However, in spite of the difference in specific predictor variables, it is worth noting that demographic information was significant in both courses.

Therefore, the finding of this research revealed that demographic could be used across courses together with motivation. Unlike other resource types, demographic information has some advantages. First, demographic information is one of the data resource types that researchers obtain before students enter school, meaning this type of the data has potential to predict academic achievement without the actual learning-based data (Woodfield, 2017). Second, the demographic information of students rarely changes over time. Therefore, the resource type can produce a consistent predictive power from the beginning of the course. Third, this type of data is not influenced by course design. In this research two course has a different instructional design in which more tools to support metacognitive learning strategies were provided by an instructor in Biology. However, according to studies on motivation or/and metacognition (Wang, Morin, Ryan, & Liu, 2016; Yerdelen-Damar & Peşman, 2013), demographic information is
placed as the antecedent of them. This result is also supported by the same predictive power (i.e., 2% of the variance in both courses).

ii. Motivation

Considering the result that the motivation has the second predictive power in Biology ($R^2 = .06$) and the strongest one in Math ($R^2 = .05$), the resource type can be another powerful predictor with demographic information that is a significant predictor across courses. Many theories demonstrate that motivation significantly predicts academic outcomes including academic achievement (i.e., for self-efficacy, see Zimmerman, 2000a; for achievement goals, see Elliot & Dweck, 1988; for expectancy-value, see Eccles et al., 1983) and the relations have been supported by tons of empirical studies. Therefore, motivation can be considered as the strong predictor of academic outcomes for most of the courses.

However, according to previous literature (Eccles et al., 1983; Meece & Jones, 1996), motivation is theorized to differ by demographic information such as gender. For example, in the certain major field (e.g., STEM), female students are motivated lower than males (Wang, 2013; Wang & Degol, 2013). Therefore, for the stronger predictive power of motivation, the inclusion of interaction terms of demographic information and motivation in the predictive modeling process should be taken into consideration. Another issue is that unlike demographic information, instructors have to put additional efforts to collect the data. Instructors need to design the survey with reliable items and analyze the data so that the result can be used for the class. In this sense, although the data of motivation requires additional processes such as a survey during the course, this type of the data might provide more stable and stronger predictive power across courses. Lastly, some researchers considered motivation personality traits, but this assumption ignores the nature of motivation that learners can be motivated differently depending on time or context (Schunk et al., 2012). In order words, motivation studies are conducted based
on the assumption in which motivational behaviors are not stable, but the behaviors are consistent within the similar contexts (Mischel, 2004). However, in this study, motivation was measured in the first week, and used to predict the final course score. Therefore, if changes in motivation over time is considered, motivation will have the potential of a stronger predictor variable.

iii. Non-theory-aligned behaviors

The only use of Environmental Structure in the first week was found to be significantly related to academic achievement in Biology, showing the lowest explained variance ($R^2 = .01$). Even this resource type is not significant predictor variables of academic achievement in Math. In other words, most of the non-theory-aligned variables were found to be not related to academic achievement. The poor predictive power of non-theory-aligned behaviors demonstrates why it is important to consider which type of data is useful in predicting academic outcome variables (i.e., feature selection). In particular, the log data by LMS or tutoring systems allows researchers to collect the data with ease compared to other methods such as survey or interview (Baker & Inventado, 2014). However, considering the fact that it is time and cost consuming to prepare and analyze the data, the selection of meaningful variables is one of the most important process in prediction modeling. The system-generated variables can be useful for a course-independent prediction model applicable across courses as these features including communication tool and policy are provided in most of the courses. This type of the data makes it possible for the researcher to build a prediction model quickly through a rapid data collection process (Romero et al., 2013).

iv. Theory-aligned behaviors

Seven variables (i.e., monitoring learning, performance, and progress) of theory-aligned behaviors were shown to be significantly related to academic achievement in Biology,
accounting for 6% of the variance, but this type of the data was not significant in Math. The amount of the variance explained by theory-aligned behaviors is greatest compared to other resource types. This finding demonstrates that theory-aligned behaviors capture important learning behaviors related to successful academic achievement. Specifically, this type of the data explained six times as much variance as non-theory-aligned behaviors (i.e., $R^2$: non-theory-aligned behaviors=.01 vs. theory-aligned behaviors=.06).

This result is supported by previous studies. In a study in which student interactions were categorized simply by interaction type, none of the variables were found to be significant (Agudo-Peregrina et al., 2014). In contrast, with interactive behaviors, Xing and his colleagues (2016) developed the prediction model that demonstrated approximately 80% of predictive power by grouping the behaviors based on activity theory.

Therefore, in building prediction models of academic outcomes, it is necessary to consider educational theories and contexts. Although this process might require additional efforts to design and evaluate student learning, this should bring a lot of benefit to the prediction model. Non-theory-aligned and theory-aligned behaviors are discussed more in section 5.2.

v. **Behavioral sequence patterns**

A sequential pattern consisting of monitoring learning processes significantly predicted academic achievement with 3% of the explained variance in Biology, but did not in Math. Considering the notion that metacognitive learning processes are not static, but procedural knowledge, the application of sequential pattern mining is an appropriate approach for the better understanding of this type of the data (Azevedo, Moos, Johnson, & Chauncey, 2010).

From the result, it should be noted that for the application of sequence mining technique, basically, the behavioral data should be enough. Although the most frequently used behavioral patterns were found, their support lower than ones in the biology course. Further, the patterns
were not significant predictor variables of academic achievement. In contrast, much data of use in metacognition supporting tools were generated, a behavioral pattern found based on them also significantly predicted academic achievement. Behavioral sequence patterns are discussed more in section 5.3.

vi. Models with combined resource types

The finding suggests that models with the combination of multiple resource types demonstrated improved $R^2$ (0.14 in Biology and 0.09 in Math) compared to models with individual resource types (i.e., the best $R^2$ is 0.06 of theory-aligned behaviors in Biology and 0.05 of motivation in Math). Therefore, existing ambient data from university systems can provide value to institutions seeking to identify and support struggling learners. Partnership with learning experts to enrich data can improve the accuracy of models and precision of interventions they afford.

The findings of the hierarchical regression analyses demonstrated relative lower predictive powers (i.e., 14% in Biology and 9% of the explained variance in Math) than previous similar studies (e.g., 33% in Macfadyen & Dawson, 2010). Some of the studies include learning behaviors taking place throughout the semester, whereas the prediction model in this research included only learning behaviors before exam 1 (4 or 5 weeks) to check the potential to develop the early prediction model.

However, in spite of other data resources including demographic information and motivation, the predictive power (i.e., 14% of the variance in Bio) is shown to be not enough to predict the dependent variable, course final scores. Therefore, in order to improve predictive power, it is recommended that additional variables such as regularity of study (Kim, Park, Yoon, & Jo, 2016) or late submission (You, 2016) that can be calculated based on existing data.
B. Differences in the predictive accuracy between theory-aligned vs. non-theory-aligned behavior

In terms of non-theory-aligned behaviors, Environmental Structure in the first week was found to be significant with .01 of the explained variance of academic achievement. However, this variables without the consideration of educational theories sometimes demonstrated poor predictive power. According to Agudo-Peregrina and his colleagues (2014), none of the independent variables, a set of interactions based on system-generated variables, was found to significantly predict academic performance in virtual learning environment (VLE)-supported classes. That is, the non-theory-aligned behaviors might not capture important learning activities, generating meaningless data, which prevent researchers from doing effective prediction modeling processes with acceptable predictive power.

This issue appears more seriously in courses in which contents are not delivered fully online. The reason why this type resource type has a poor predictive power is that the system did not capture all learning behaviors as the course was not fully online, but system-supported F2F courses or F2F courses (Picciano, 2014). Therefore, some learning behaviors that are more critical for successful academic achievement than the use of LMS tools might happen offline. Therefore, if a prediction model for system-supported courses is developed, it will be necessary to take into account an approach that captures offline learning data for the better utilization of the behavioral data.

However, this type of the data, as universal features across courses, is necessary for the scalability of the prediction model in that the specific course-oriented features might cause the likelihood of over-fitting problems (Witten, Frank, Hall, & Pal, 2016).

Unlike non-theory-aligned behaviors, some of the theory-aligned behaviors significantly predict academic achievement in Biology, explaining 6% of the variance, which was the most
powerful predictor variable in the model. In addition to the strongest predictive power, the findings demonstrated that the prediction model developed based on the consideration of context and educational theories provides instructors with interpretable results so that they can use the result to improve the course by identifying and intervening students who are more likely to perform poorly (Gašević, Dawson, & Siemens, 2015). It is also worth noting that such theory-aligned variables allow instructors for a contextual framework that help them understand the importance of findings (Lockyer, Heathcote & Dawson, 2013).

To emphasize the importance of educational theories, Xing and his colleagues (2015) used the term “blunt computational instrument” that means a process to handle the data focused on methodology and algorithms without the consideration of human behaviors based on educational contexts (p.169). That is why instructors who have no background knowledge about sophisticated computation have difficulty utilizing the findings of prediction modeling studies (Romero & Ventura, 2010).

Additionally, through the result in Biology, it is confirmed that metacognitive learning processes are important components together with motivation for successful performance by the SRL process (Winne & Hadwin, 1998). Specifically, the result demonstrated that most of the metacognitive learning process types are important. In addition, for better academic achievement, it is also necessary for students to manage their learning using metacognitive learning processes throughout the course in that the behaviors in each period were found to be significant.

C. What patterns of theory-aligned behaviors emerge?

According to Winne and Perry (2000), SRL can be better explained with events in that events are actual behaviors that students perform rather than a description of the action or mental
states. In this sense, methods to measure events such as behavioral patterns are more appropriate to capture metacognitive learning behaviors than those based on static measurement including trait or aptitude (Winne, 2010). Therefore, in this research, the most frequent behavioral sequences were found respectively in both courses (Taub et al., 2014). The courses were developed by different instructional design, so instructors provided metacognition supporting tools differently, which led to the findings of distinct sequential behavioral patterns in each course.

In Biology, repeated monitoring learning processes through self-assessment quizzes were discovered as the most frequently used metacognitive learning behaviors based on minimum support and confidence. According to Butler (1993), students who desire to master learning contents are more likely to use self-assessment quizzes until the score meets their standard set before. Therefore, the repeated use of self-assessment quizzes indicates the willingness of students to master learning contents, so they are more likely to reach the better academic achievement.

Unlike Biology, the repeated use of Planning was found as the most frequently used metacognitive learning patterns in Math. Considering Planning includes a syllabus, calendar, and schedule, it is likely that students look at the overview of the course in the syllabus and they then check the specific dates through syllabus or schedule. However, the repeated use of Planning with syllabus, course calendar, and exam guides was not significant predictor variables of academic achievement. This finding of regression analysis with individual resource type indicated no significant predictive power of Planning, so it is not surprising the repeated use of planning is not a significant result.
In this research, the purpose of the sequential pattern mining is to discover the most frequently used patterns, but it is not guaranteed that the patterns indicate the most effective behavioral patterns. In other words, there might be other sequential patterns that more positively influence academic outcomes than the repeated use of self-assessment quizzes. Due to the limit of time and cost, however, the effect of all behavioral patterns on academic outcomes was not tested. Therefore, it is necessary to consider the employment of more automatic methods for the more efficient process to deal with all of the possible sequential patterns (Perera, Kay, Koprinska, Yacef, & Zaïane, 2009). In addition, if the most frequently used patterns by performance groups are discovered to find more effective metacognitive patterns, that could be another interesting finding to demonstrate the effectiveness of sequential pattern mining (Taub et al., 2014).

The behavioral pattern was found from the existing data resource, theory-aligned behaviors. In other words, researchers can obtain a predictive power by discovering hidden information through a data mining approach without additional efforts to collect the data. Although the different analytical approach is needed to get this result, systems including an LMS provide sequential and dynamic data (Bernacki, 2018). Therefore, researchers can find further interesting findings by applying another cutting-edge method.

In addition, it should be noted that sequential patterns allow instructors to provide students with a specific intervention. With the significant patterns, instructors figure out how the tool can be utilized for better performance. Specifically, through the finding of repeated monitoring learning, instructors can encourage students to use self-assessment quizzes repeatedly to check the status of their learning status until the learning contents are mastered. According to Vrugt and Oort (2008), students focused on the mastery of learning contents demonstrated the
more active use of metacognitive strategies than those who are interested in outperforming others.

**D. Different predictive power by course**

In this research, two courses, Biology and Math, were not designed differently. Specifically, in Biology, instructors provided students with learning guides designed based on metacognitive learning strategies that allow students to better self-regulate their learning. Also, these materials encourage and further, high-performance students were more likely to have this kind of interest. Therefore, this different instructional course design should be different predictive power of each type of data resource, with particular differences among behavioral data.

Although non-theory-aligned behaviors explained the small amount of the variance in Biology ($R^2 = .01$), a more interesting difference starts from theory-aligned behaviors. In Biology, a variety of tools were provided consistently (67.8 kinds of metacognition supporting features), whereas just two kinds of them were offered consistently across semesters in Math. Therefore, students in Math did not have opportunities to fully utilize metacognition supporting tools to manage their learning process.

Also, the role of instructors can be one reason for the great predictive power ($R^2 = .06$) of theory-aligned behaviors in Bio. According to how instructor support student, their use of metacognitive learning strategies can be differentiated (Reingold, Rimor, & Kalay, 2008). To facilitate the use of metacognition supporting tools, instructors provided two additional materials. The first was “Learning to Learn” in which the effects of metacognitive learning strategies was demonstrated with empirical evidences and another one is advice of successful students in previous semesters. They mentioned how they succeeded in the course, focusing on
learning strategies. Therefore, these materials might encourage students who have more interests in their learning to use metacognition supporting tools for the management of learning.

6. Implications

The findings of the research demonstrate that prediction models with combined resource types better explain academic achievement than those with individual resource type. Therefore, existing ambient data from university systems can provide value to institutions seeking to identify and support struggling learners. In practice, partnerships with learning experts to enrich data can improve the accuracy of models and precision of interventions they afford.

Log data generated by a system such as an LMS provide researchers with opportunities to employ new analytical approaches that allow them to examine the data in different ways from traditional methods such as surveys. Therefore, researchers can have a chance to discover the hidden information that can be more powerful and authentic as the finding can align with educational theory or context. Further, through this new approach, researchers can investigate the data from various angles on top of previously existing methods, which does not require additional efforts to collect data.

In addition, educational theory and contexts should be taken into consideration in designing courses and developing the prediction models. The result of prediction models with the consideration of educational theory allows researchers to provide more specific intervention within an educational framework. The inclusion of theory-aligned variables might lead to improved identification of student success or failure in the courses by better explaining learning processes. The improved prediction model can better identify students in need and prevent them from experiencing negative academic outcomes. For the improvement of prediction power through theory-aligned variables, courses design based on educational theories should precede
others. Therefore, in practice, it is necessary to provide theory-based tools enough for students to support and manage their learning. This design approach could be helpful for student academic achievement and further, improve predictive power to identify student success and failure in the course.
Chapter 5: Synthesis, Conclusion and Implications

1. Methodological advancement

The three studies that comprise this dissertation examine relations between student characteristics, motivations, metacognitive learning processes, and academic achievement. Methodologically, the dissertation demonstrated the potential of multiple types of approaches and data resource types. By employing multiple approaches including variable-centered, person-centered, and learning analytics, researchers can understand learning processes from various angles. In addition, through this triangulation by multiple types of methodological approaches, educational theories could be more thoroughly verified and supported by various empirical findings. Multiple types of data resources are related to analytical methods.

Specifically, beyond a traditional variable-centered approach, a person-centered approach allows researchers to identify heterogeneous groups of individuals and figure out how each construct contributes to the groups. Additionally, the aim of learning analytics approach to provide effective intervention to students in need and further improve the learning environment by analyzing students’ interactions and discovering interesting information through cutting-edge methodologies. However, so far, little research has employed multiple types of analytical approaches together to understand learners in a single context.

A person-centered approach has the ability to account for many co-occurring phenomena that describe a learner and derive solutions that can describe common groups of learners who share similar profiles across these dimensions. This kind of solution is advantageous because it can make for more parsimonious analyses via a data reduction phase that can then inform future analyses. For instance, a latent class or profile analysis can handle many motivational variables and identify groups of students with similar patterns of motivations, who can then be compared
in terms of their subsequent behaviors or achievement (i.e., Paper 2). This is particularly valuable when emergent solutions produce profiles where two groups differ in a fashion that affords the test of a theoretical assumption. For instance, a latent profile solution that produces a group with high expectancy and value and another with high expectancy and low value provides an opportunity to examine the assumption that expectancies interact with perceptions of value to influence students’ extent and type of engagement in learning (i.e., Eccles et al. 1983).

In addition to the person-centered approach, the use of behavioral data from an LMS affords opportunities to capture and explain dynamic and complicated metacognitive learning processes that are critical components of SRL models (e.g., Winne & Hadwin, 1998). Compared to traditional education data used for research conducted in experimental settings, machine-generated data are shown to be more appropriate for research in that the SRL process demonstrates dynamic, cyclical, and sequenced patterns (Bernacki, 2018; Biswas et al., 2018). In this sense, by employing sequence pattern mining from educational data mining techniques, this research discovered the frequently used behavioral patterns based on metacognition learning processes and investigated how the patterns influence academic achievement.

2. Theoretical contributions

A. Examining the Influence of Undergraduates’ Achievement Goals on Metacognitive Behavior Sequences, and Achievement in Science

The purpose of the first paper was to examine relations between achievement goals and metacognitive learning behaviors using a clustering analysis and visualization. A clustering analysis conducted with achievement goals produced three goal profiles; 1) mastery-approach, 2) performance-approach, and 3) performance-avoidance identified three goal profiles. The profiles include High Approach, High Mastery, And High Goal Endorsement groups. The finding demonstrated that students in the High Mastery group, who had greater use of the self-
assessment tool, obtained higher final grades than other groups could be explained from the perspective of SRL. In addition, learners motivated by mastery approach goals engaged in the greater use of self-assessment quizzes. A student in the High Mastery group also used the tools earlier than other two groups for exam 2. As the most frequently used pattern, sequential pattern mining discovered the repeated use of self-assessment quizzes to monitor their learning. More students in the High Mastery group employ this pattern of metacognitive events than students in the High Performance and High-Goal endorsement groups, particularly during sessions in weeks before exams. A subsequent analysis revealed that for all exams, students who conducted a repeated behavior pattern indicative of metacognitive monitoring and control outperformed those who did not. From the research, it is confirmed that the person-centered analysis provided authentic and generalizable groups and afforded observation of the learning behaviors of learners with typical combinations of goals. In addition, sequential patterns provide instructor more interesting information on learning processes than the frequency of accesses.

B. A Latent Profile Analysis of Undergraduates’ Achievement Motivations and Metacognitive Behaviors, and their Relations to Achievement in Science

The purpose of the second research was to identify motivational profiles based on multiple types of motivations including self-efficacy, achievement goals, and expectancy-value from an integrative perspective. For this research, a LPA was conducted with ten types of motivational constructs and three kinds of metacognitive learning processes. The LPA identified four motivational profiles; 1) High Cost, 2) Moderately motivated, 3) High Goals and Values, and 4) Low Performance Goals, and three metacognitive profiles; 1) Infrequent metacognitive processing. 2) Checking performance and planning, and 3) Self-assessment. Student demographic information significantly influenced the membership of motivational profiles. Older students tend to have higher self-efficacy, mastery-approach, and values, but low cost than
younger ones. In addition compared, to Caucasian and Asian students, underrepresented students tend to be more motivated by higher goals and values than high cost or moderate motivation. Lastly, female students are more likely to be members of Moderately motivated and High goals and values than High cost oriented and Low performance goals and cost than males.

In terms of the relations profiles with academic achievement, Low Performance Goals group showed the best performance. Among metacognitive profile groups, students in Checking performance and planning, and Self-assessment demonstrated similar academic performance. The investigation of relations between two profile groups demonstrated that students in the High cost group are more likely to be a member of self-assessment group than checking performance and planning as well as of a member of an infrequent metacognitive process than checking performance and planning. In addition, students in high performance and goals and high goals and values groups relative to the low performance goals group more likely to be a member of the infrequent metacognitive process than checking performance and planning. The findings of this research provide authentic motivation status and metacognition learning process as well as their relations. Addition, this research figured out specific motivational profiles through the multiple types of motivations from the integrative perspective. Therefore, instructors can provide more effective and specific interventions to students who have difficulty utilizing metacognitive learning processes, considering motivational status based on multiple motivations. In addition, instructors can understand motivational profiles by demographics so at the beginning of the semester in which the information on students is not enough to identify students learning processes, they intervene students based on demographic information.

C. Examining the Power of Multiple Data Sources in Predicting Academic Achievement in Undergraduate STEM Courses
The purpose of the third paper was to consider the relative importance of capturing demographic, motivational and metacognitive processes as potential predictors of learning outcomes, and appraises them alongside both traditional prediction modeling approaches in higher education, and emergent methods, sequence pattern mining, arising from the field of educational data mining. The sequence pattern mining discovered the repeated use of self-assessment quizzes in Biology and repeated use of planning contents in Math. A regression model with combined resource types demonstrated the improved predictive power than models with individual resource types. Also, theory-aligned behaviors designed based on metacognitive learning processes better improved the accuracy of the model than non-theory-aligned behaviors automatically provided by the system. Lastly, when applying the same prediction model, the model better explained the variance of academic achievement in Biology in which metacognitive supporting tools designed based on an educational theory than that in Math that has few theory-aligned behavior variables.

Therefore, this study emphasizes the importance of existing ambient data from university systems. Also, log data generated by systems such as LMS allows researchers to examine the same data in different ways with no need for additional data collection. Lastly, educational theory and contexts should be taken into consideration in designing courses and developing the prediction models. Therefore, instructors and researchers, in designing courses, the consideration of educational theories and contexts is the essential process.

3. Synthesis & Conclusion

This dissertation applies multiple data sources and analytical approaches to investigate and explain the influence of motivations on metacognitive processes and examines how these phenomena influence academic achievement. Person-centered approaches provide parsimonious
solutions that accommodate many motivational variables, examine their co-occurrence, and produce insights that cannot be found in a traditional variable-centered approach. In particular, by employing this integrative perspective, the findings could provide more authentic and practical interpretations of motivational theories that propose complex interactions of phenomena, and capture them as they arise from samples collected in authentic learning contexts. Specifically, it was confirmed that multiple motivation constructs operate together according to previous theories in generating motivational profiles (i.e., achievement goal and expectancy value theories). In addition, their relations with metacognitive learning behaviors and academic achievement align to theories of self-regulated learning, and findings largely corroborate and improve upon prior empirical studies.

Similarly, learning analytics using log data provided opportunities to examine metacognitive processes with time stamps and employ sequential pattern mining, which makes it possible to discover hidden information from the same data. In the traditional environment in which a survey such as MSLQ was primarily used to measure metacognition, it was challenging to capture changes in metacognitive learning processes. In addition, and this ambient data allowed researchers to understand the student learning more contextually by investigating relevant variables to the main learning process. Eventually, the multiple types of methodological approaches and data resources produce better findings that in turn allow researchers to provide more specific and effective intervention to students in need.

In addition to demonstrating the importance of these analytical approaches to understanding learning, these studies provide instructors with opportunities to make design choices based on the results. For instance, instructors might consider students’ motivational profile (e.g., their goals, the kinds of value they aim to derive from a course) and adapt their
instruction accordingly by pushing students to access certain content. Results also have implications for university professionals charged with improving students’ academic success. Results from Paper 3 indicates the kinds of data sources and variables university data analysts should prioritize when developing systems to identify students likely to require support. These data can be used to inform instructors, or to initiate connections with campus units like academic success centers, who can provide students with coaching, tutoring, or supplemental instruction. Ultimately, results across papers demonstrate the unique value of rich collection and complex modeling of motivation and metacognitive learning processes for refining theories of self-regulated learning, and for improving these learning processes and outcomes achieved by students.
### Appendix

**Table 27. Factor Loadings**

<table>
<thead>
<tr>
<th>Item</th>
<th>SELF</th>
<th>MAP</th>
<th>PAP</th>
<th>PAV</th>
<th>ATT_V</th>
<th>INT_V</th>
<th>UTI_V</th>
<th>EFF_C</th>
<th>OPP_C</th>
<th>PSY_C</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELF1</td>
<td>.69 (.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SELF2</td>
<td>.77 (.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SELF3</td>
<td>.75 (.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SELF4</td>
<td>.83 (.01)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SELF5</td>
<td>.76 (.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAP1</td>
<td></td>
<td>.68 (.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAP2</td>
<td></td>
<td>.67 (.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAP3</td>
<td></td>
<td>.67 (.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAP1</td>
<td></td>
<td></td>
<td>.64 (.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAP2</td>
<td></td>
<td></td>
<td>.69 (.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAP3</td>
<td></td>
<td></td>
<td>.74 (.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAV1</td>
<td></td>
<td></td>
<td></td>
<td>.70 (.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAV2</td>
<td></td>
<td></td>
<td></td>
<td>.84 (.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAV3</td>
<td></td>
<td></td>
<td></td>
<td>.78 (.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATT_V1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.71 (.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATT_V2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.58 (.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATT_V3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.58 (.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATT_V4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.56 (.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INT_V1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.80 (.01)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INT_V2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.86 (.01)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INT_V3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.86 (.01)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INT_V4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.86 (.01)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTI_V1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.82 (.01)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTI_V2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.63 (.02)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTI_V3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.88 (.01)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTI_V4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.57 (.02)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFF_C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.70 (.02)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFF_C2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.70 (.02)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFF_C3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.71 (.02)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFF_C4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.72 (.02)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPP_C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.89 (.03)</td>
<td></td>
</tr>
<tr>
<td>OPP_C2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.89 (.02)</td>
<td></td>
</tr>
<tr>
<td>OPP_C3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.79 (.02)</td>
<td></td>
</tr>
<tr>
<td>OPP_C4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.83 (.03)</td>
<td></td>
</tr>
<tr>
<td>PSY_C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.60 (.03)</td>
</tr>
<tr>
<td>PSY_C2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.53 (.03)</td>
</tr>
<tr>
<td>PSY_C3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.86 (.03)</td>
</tr>
<tr>
<td>PSY_C4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.82 (.03)</td>
</tr>
</tbody>
</table>

*Note. SELF: self-efficacy; MAP: mastery-approach; PAP: performance-approach; PAV: performance-avoidance; ATT_V: attainment value; INT_V: intrinsic value; UTI_V: utility value; EFF_C: effort cost; OPP_C: opportunity cost; PSY_C: psychological cost.*
Table 28. Factor Correlations

<table>
<thead>
<tr>
<th></th>
<th>SELF</th>
<th>MAP</th>
<th>PAP</th>
<th>PAV</th>
<th>ATT_V</th>
<th>INT_V</th>
<th>UTI_V</th>
<th>EFF_C</th>
<th>OPP_C</th>
<th>PSY_C</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELF</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAP</td>
<td>.67</td>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAP</td>
<td>.16</td>
<td>.41</td>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAV</td>
<td>.04</td>
<td>.24</td>
<td>.95</td>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATT_V</td>
<td>.59</td>
<td>.68</td>
<td>.23</td>
<td>.18</td>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INT_V</td>
<td>.51</td>
<td>.51</td>
<td>.10</td>
<td>.04</td>
<td>.83</td>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTI_V</td>
<td>.42</td>
<td>.47</td>
<td>.15</td>
<td>.12</td>
<td>.85</td>
<td>.74</td>
<td>1.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFF_C</td>
<td>-.46</td>
<td>-.43</td>
<td>-.01</td>
<td>.03</td>
<td>-.70</td>
<td>-.50</td>
<td>-.61</td>
<td>1.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPP_C</td>
<td>-.27</td>
<td>-.2</td>
<td>.02</td>
<td>.02</td>
<td>-.29</td>
<td>-.18</td>
<td>-.19</td>
<td>.54</td>
<td>1.</td>
<td></td>
</tr>
<tr>
<td>PSY_C</td>
<td>-.42</td>
<td>-.2</td>
<td>.10</td>
<td>.15</td>
<td>-.20</td>
<td>-.30</td>
<td>-.12</td>
<td>.41</td>
<td>.51</td>
<td>1.</td>
</tr>
</tbody>
</table>

Note. SELF: self-efficacy; MAP: mastery-approach; PAP: performance-approach; PAV: performance-avoidance; ATT\_V: attainment value; INT\_V: intrinsic value; UTI\_V: utility value; EFF\_C: effort cost; OPP\_C: opportunity cost; PSY\_C: psychological cost.
References


CURRICULUM VITAE

Wonjoon Hong

CONTACT
4505 S Maryland Pkwy, Las Vegas, NV 89154
College of Education, University of Nevada, Las Vegas
Office: CEB 399F
Office Phone: 702-895-1501
Email address: wonjoon.hong0220@gmail.com

EDUCATION

Ph.D. Utah State University
Instructional technology and learning sciences 08/2014-05/2016
(Transfer)
University of Nevada, Las Vegas
Educational Psychology and Higher Education 06/2016-08/2018
Dissertation: Exploring Relations Between Motivation, Metacognition, and Academic Achievement through Variable-Centered, Person-Centered and Learning Analytic Methodologies (Chair: Dr. Matthew L. Bearnacki)

M.S. Seoul National University
Educational Technology 02/2013
Thesis: The effects of segmenting time of the video lecture on learning outcomes
(Chair: Cheolil Lim)

B.S. Kongju National University
Computer Education 02/2008

RESEARCH INTERESTS
- Data mining & learner analytics (development of infrastructure to log & analyze learning events)
- Self-regulated learning (cognitive, metacognitive & motivational processes)
- Learning technologies (intelligent tutoring systems, learning management systems, hypermedia)
- K-20 STEM learning (K-12 math and science; undergraduate math, science, & engineering)
- Technology Integration in educational settings

RESEARCH EXPERIENCE
06/2016 - present  Research on Metacognition and Motivation in Learning Technology
06/2016 - present  Visualization of Learning activities in University of Nevada, Las Vegas
08/2014 - 05/2016  Data Mining and Learning Analytics from the CANVAS at USU
01/2015 - 04/2015  Data visualization using Tableau
05/2013 - 04/2014  Development of Design Principles of Emotional Interface for SLR
03/2011 - 06/2011  The Developing the Instructional Program for the New Engineer spring
07/2012 - 08/2012  The Evaluation Methodology for developing the smart learning solution
09/2011 - 02/2013  Designing the Learning Environment Model for Creative Problem Solving
03/2011 - 06/2012  Usability Testing for the Speech Recognition Word Processing Software

**Publications**


**Conference Presentations**


**Graduate Teaching Assistant**

2018 Fall  
Self-Regulated Learning, Metacognition and Motivation (EPY 752)  
- Teaching, Evaluation, Data Preparation for AERA 2019

**Work Experiences**

<table>
<thead>
<tr>
<th>Date</th>
<th>Location/Role</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>07/2005 - 03/2008</td>
<td>XMD Inc. (ERP-Enterprise Resource Plan company), Korea</td>
<td></td>
</tr>
</tbody>
</table>
- DBA (MS-SQL, My-Sql)  
- Web-Programming (ASP, PHP)  
- Developing ERP systems for logistics  
| 09/2008 - 10/2010 | Kongju Medical Center, Korea                       |  
- Military Public Service Personnel  
- Caring for patient, Managing the stocks  
| 01/2012 - 08/2012 | Graduate Student Instructor in Seoul National University, Korea |  
| 08/2012 - 12/2012 | System Assistant in Seoul National University Dormitory, Korea |  

**Workshops**

<table>
<thead>
<tr>
<th>Date</th>
<th>Title</th>
</tr>
</thead>
</table>
| 04/2017    | Community Based Educational Data Repositories and Analysis Tools in International Educational Datamining Conference 2017 (China)  
- Presentation title: A Prediction and Early Alert Model Using Learning Management System Data  

**Certificates**

<table>
<thead>
<tr>
<th>Date</th>
<th>Certificate</th>
</tr>
</thead>
</table>
| 07/2009    | MCITP(Microsoft Certified IT Professional)  
- Database Developer 2008  
| 08/2008    | MCTS(Microsoft Certified Technology Specialist)  
- SQL Server 2008: Implementation and Maintenance  
- SQL Server 2008: Database Development  
| 08/2012    | Completion of the Course for Developing the Android Application  
| 08/2012    | Completion of the Course for the Android Application Design  
| 02/2017    | Splunk for Analytics and Data Science  

GRANTS

05/2013 – 04/2014 National Research Foundation of Korea. ($20,000)
  • Cheol-II Lim, Tae-Jung Park, Won-Joon Hong, Jeong-Eun Park
  • Developing Design Principles of Emotional Interface for Self-Regulated Learning in an E-Learning Environment

AWARDS & HONORS

12/2009 The certificate of commendation from Chief of Kongju Medical Center
03/2002 Enter University at the top of my major

SCHOLARSHIPS

Doctoral Program
07/2016 - 08/2018 Graduate Research Assistantship at University of Nevada, Las Vegas
08/2014 - 05/2016 Graduate Research Assistantship at Utah State University

Master Course
02/2011 Superior Academic Performance ($900)
01/2012 Work-Study Scholarship ($300)
01/2012 Lecture & Research Support Scholarship ($4356)
02/2012 Work-Study Scholarship ($2800)

Bachelor
04/2002 Outstanding Academic Achievements ($1442)
07/2002 Outstanding Academic Achievement ($300)
02/2005 Language Achievement ($1237)

VOLUNTEER

08/2004 International Volunteer (Mongolia – 24days)

COMPUTER SKILLS

• DBMS (database management system) – MS-SQL & My-SQL
• Maintaining Web Programming – ASP & PHP
• Statistical software package - SPSS, R, Mplus, and SAS
• Splunk