Modeling Student Affective State Patterns during Self-Regulated Learning

in Physics Playground

Shiming Kai

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY
2019

ABSTRACT

Modeling of Affective State Patterns during Self-Regulated Learning in Physics Playground

Shiming Kai

This dissertation research focuses on investigating the incidence of student self-regulated learning behavior, and examines patterns in student affective states that accompany such self-regulated behavior. This dissertation leverages prediction models of student affective states in the Physics Playground educational game platform to identify common patterns in student affective states during use of self-regulated learning behavior. In Study 1, prediction models of student affective states are developed in the context of the educational game environment Physics Playground, using affective state observations and computer log data that had already been collected as part of a larger project. The performances of student affective state prediction models generated using a combination of the computer log and observational data are then compared against those of similar prediction models generated using video data collected at the same time. In Study 2, I apply these affective state prediction models to generate predictions of student affective states on a broader set of data collected from students participants playing Physics Playground. In parallel, I define aggregated behavioral features that represent the self-observation and strategic planning components of self-regulated learning. Affective state predictions are then mapped to playground level attempts that contain these self-regulated learning behavioral features, and sequential pattern mining is applied to the affective state predictions to identify the most common patterns in student emotions.

Findings from Study 1 demonstrate that both video data and interaction log data can be used to predict student affective states with significant accuracy. Since the video data is a direct

measure of student emotions, it shows better performance across most affective states. However, the interaction log data can be collected natively by Physics Playground and is able to be generalized more easily to other learning environments. Findings from Study 2 suggest that self-regulatory behavior is closely associated with sustained periods of engaged concentration and .self-regulated learning behaviors are associated with transitions from negative affective states (*confusion, frustration, and boredom*) to the positive *engaged concentration* state.

The results of this dissertation project demonstrate the power of measuring student affective states in real time and examining the temporal relationship to self-regulated learning behavior within an unstructured educational game platform. These results thus provide a building block for future research on the real-time assessment of student emotions and its relationship with self-regulated learning behaviors, particularly within online student-centered and self-directed learning contexts.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Professor Ryan Baker, for his unwavering sponsorship and mentorship throughout my doctoral study. Not only did he guide me through the field of learning analytics and educational data mining, he had encouraged my curiosity as a scholar, and challenged me in different ways to think more critically and creatively about the work I am doing. The quantitative and analytic rigor he brings to every discussion has been one of his greatest gifts, that I can only hope to emulate as a researcher. This dissertation would not have been possible if not for his constant guidance, deep insights and assistance through the whole process.

I would also like to extend my sincere appreciation to my dissertation committee members for their invaluable feedback and questions throughout the different phases of the process. In particular, I would like to thank Professor John Black, for being the first professor who believed in my potential and sponsored my transfer to the PhD program in Cognitive Studies. He has lent many a listening ear and provided much guidance and feedback through my years in the PhD program. To Professor Corter, who has displayed a genuine interest in my dissertation work, thank you for your support in helping me get through the dissertation process, and the insightful feedback that challenged me to think more deeply about my work. To Professor Holbart, for sharing his knowledge and passion in the area of educational games, that has inspired me in new ways. Special thanks also to Professor Luc Paquette, who has been the invaluable advisor throughout my years in the doctoral program, even now when he has moved on to better things. You have been the ultimate mentor and team-mate, with your expertise and willingness to guide me through managing the data I have been working on, tirelessly fielding my questions from the mundane to the most challenging. My dissertation project and related analyses would not have been possible without your assistance and advice throughout the whole process.

I would also like to thank Dr. Nigel Bosch, whose research had played a big part in my work, and who was always quick to lend a hand with his knowledge and expertise. Special thanks also to Professors Valerie Shute and Sidney D'Mello, whose amazing intellect, energy and passion about the Physics Playground project have inspired me to strive to be a better researcher and scholar. To Dr. Jaclyn Ocumpaugh, who has been a constant source of emotional and mental support, especially through the rough patches in my dissertation journey. To current and alumni members of the EDM Lab – Dr. Sweet San Pedro, Dr. Jeanine deFalco, Dr. Elle Yuan Wang, Dr. Mia Almeda, Dr. Aaron Hawn, Miggy Andres, Stefan Slater, Shamya Karumbaiah, Alexis Andres and Chad Coleman – thank you for all the encouragement and friendship, and for always being ready to help. I have sorely missed the camaraderie and bonds we shared in that tiny office here in Teachers College.

A special thank you to the Gates Foundation for the support and funding of the Physics Playground project that this dissertation was based upon. I am honored to have been a part of this remarkable collaborative effort.

Last but not least, I would like to express my heartfelt gratitude to my parents and family. To my mother, Ahmui Chin, and my father, Gungchew Kai, thank you for your unyielding support and unconditional love throughout this journey, particularly during the difficult times. Special thanks to my husband Howard Wu, who has been my counsellor, my personal assistant, my backup nanny, and most of all, my rock throughout this journey. Without him, this project would never have been completed.

DEDICATION

To the past and the future.

In memory of my cranky aunt Kimchin Chen,
who would have loved to brag about my accomplishments,

and to my babies Nami and Garen,
without whom this project would have been completed two years earlier.

CHAPTER I.

INTRODUCTION

Background

Self-regulated learning is an essential aspect of the learning process and has received much attention from researchers since the 1980s. While there are several theoretical models for the concept of self-regulation, self-regulated learning (SRL) generally can be defined as an iterative and internally driven process where learners construct their own learning goals and then monitor, regulate and control their cognition and behavior towards the accomplishment of these goals (Pintrich & Zusho, 2002). The capability to self-regulate is thus a process that plays a mediating role between cognitive and motivational factors as well as learner characteristics (Pintrich, 2003; Pintrich, 2000) to influence the learning process and in turn, learning outcomes. The emerging ability to monitor one's own behavior and self-regulate is an important developmental task starting in early childhood (Kopp, 1982). Studies on individuals' cognitive and emotional development have shown that the ability to regulate cognition and affect are necessary for success in school and academic achievement (Pintrich & de Groot, 1990; Zimmerman, 1998).

While self-regulated learning is important in traditional education settings, it may be even more vital within online and informal learning contexts (Jonassen, 1995) where learning is more student-centered (Artino, 2008; Williams & Hellman, 2004). Multiple studies have demonstrated that the use of self-regulated learning strategies within learner-controlled online learning environments are correlated with achievement outcomes (Joo, Bong & Choi, 2000; Young, 1996; Land & Greene, 2000). Another study found evidence that suggests that higher levels of learner-choice allowed within an online learning environment benefits students with high self-regulated learning strategy use (McManus, 2000). Further studies have shown that students who demonstrate self-regulated behavior have a stronger a sense of connectedness and self-efficacy and a better student experience overall (Cho, Demei, & Laffey, 2010; Turker & Zingel, 2008).

Given the far-reaching implications of SRL behavior and the increasing popularity of online and informal learning platforms, the ability to self-regulate has only become more critical as increasing numbers of students are learning in student-centered settings. As a result, the development of more accurate and reliable measures of student emotions and SRL behaviors will be critical for improving the student learning experience. Specifically, the proliferation of technology-based platforms that implement student-centered learning opportunities has created an urgent need to develop more technology-friendly methods to evaluate and identify student emotions and use of self-regulated learning within these platforms.

Theoretical frameworks of self-regulated learning

Multiple frameworks for self-regulated learning exist, but in general, frameworks for these strategies consist of a few chronological phases. One set of theoretical frameworks maps the various components of student SRL behavior or strategies during learning, based on the diversity of perspectives towards student learning. For example, the SRL framework based on the operant perspective (Mace, Belfiore & Hutchinson, 2001) emphasizes the students' choices in alternative actions during the learning process. In comparison, the socio-cognitive perspective of SRL (Zimmerman, 2000) focuses more on the idea that each component of student self-regulated learning behavior is situationally specific, such that students may engage in some of the components of SRL more in certain contexts than in others. Similarly, Winne & Hadwin's model framework (1998) based on the information processing perspective emphasizes the internal and external cognitive conditions that affect student use of SRL strategies, such as the schemas available to the student, as well as her memory capacity during learning.

While each of these theoretical frameworks emphasizes a different aspect of student SRL, they generally house SRL components within three main phases of learning:

1) Forethought Phase - the planning phase before the start of a task
2) Performance Phase – the phase when the student engages in the task
3) Reflection Phase - the post-performance phase when the student evaluates her performance

Strategic planning and goal setting occurs during the forethought phase (Zimmerman, 2002) before the learning process begins, when the student analyzes the task and plans how she would go about achieving the learning objectives. The performance phase of learning follows the forethought phase, and consists of the self-observation and self-control components. Lastly, self-reflection and evaluation may occur immediately after learning during the self-reflection phase, which involves students' reactions and feelings towards their performance as they review it in comparison to various sets of standards.

Assessment of Self-Regulated Learning (SRL)

Detection and assessment of specific emotions, strategies, and behavior related to self-regulated learning (SRL) have been mostly limited to self-report questionnaires and interviews, or observational reports (Schmitz & Wiese, 2006; Zimmerman, 2008). Common questionnaires that were created to assess self-regulated learning strategies and behavior include the Learning and Study Strategies Inventory (LASSI, Weinstein, Schulte & Palmer, 1987), and the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, Smith, Garcia, & McKeachie, 1993). These examples of extensive questionnaires assess the individual components of self-regulated learning across multiple sections. However, these questionnaires could be further classified as “aptitude” measures of self-regulated learning since they are designed to aggregate self-regulatory behavior over time (Winne & Perry, 2000).

In recent years, there has been a push to assess SRL more as a temporal “event”. One area of research is the development of phase models of SRL that separates students’ SRL process into different learning phases before, during and after the students’ learning attempts (Boekaerts & Corno, 2005; Pintrich, 2000). Researchers have thus shifted towards alternative measures of SRL to capture instances of SRL temporally, such as think-aloud protocols, observations, and using online tools such as structured diaries, computer interaction logs and microanalytic measures in online learning environments (Zimmerman, 2008).

One area of research that has received much attention in recent years has focused on identifying and evaluating student use of self-regulated learning using computer interaction logs from various technology-based learning platforms (Aleven, McLaren, Roll, & Koedinger, 2004; Azevedo, 2005; Roll, Aleven, McLaren, & Koedinger, 2011). In these studies, fine-grained data logs are available that allow researchers to identify specific student affect and behavior patterns within these technology-based learning platforms that constitute self-regulated learning. Aleven and colleagues (2006) identified and developed models of students' help-seeking behavior by capturing computational features that constituted effective versus ineffective help-seeking behavior during their interactions with the Cognitive Tutor in geometry.

Past studies have compared the accuracy of technologically enabled methods against traditional methods for assessing SRL strategies. One study found that student self-reports were generally less effective than computer interaction logs at accurately identifying various components of student self-regulated behavior (Winne & Perry, 2000). While another study showed that questionnaires were less accurate than structured diaries in assessing students’ use of self-regulated learning strategies (Schmitz & Wiese, 2006).

Assessment of student affective states

Several past studies have demonstrated that student affective states directly impact learning experiences, and in turn, their learning outcomes (Bower, 1992; Goetz, Pekrun, Hall, & Haag, 2006; Shutz & Pekrun, 2007). Additionally, emotional control is implicitly present within most self-regulated learning frameworks, often within the components of motivational or metacognitive control (Zimmerman & Schunk, 2001). Several studies have also shown that emotions are an important factor in students' use of SRL strategies during learning (Blair & Diamond, 2008; Mega, Ronconi, & De Beni, 2014; Reschly, Huebner, Appleton, & Antaramian, 2008; Pekrun, Goetz, Titz, & Perry, 2002; Wolters, 1998). In particular, Pekrun and colleagues (2002) were able to identify positive correlations between positive student emotions and students' ability to make use of various self-regulated learning strategies during learning, using the Academic Emotions Questionnaire (AEQ). Similarly, Mega and colleagues (2014) were able to identify a positive relationship between student emotions, self-regulated learning and academic achievements, using a battery of self-report measures. In this study, they found that student emotions influence their use of self-regulated learning strategies, which in turn affects academic achievement. Furthermore, positive emotions were associated with academic achievement only if they were mediated by self-regulated learning and motivation (Mega, Ronconi & de Beni, 2014).

However, while measures of SRL have taken advantage of technological advancements, measures of student emotions have been less common. The assessment of emotions in educational psychology has been mostly limited to the use of traditional aggregated self-report questionnaires, expert coding, and field observations. Some examples of questionnaires include the 27-item Positive and Negative Affect Schedule (PANAS), which assesses how frequently an individual feels certain emotions (Reschly et al., 2008), the Academic Emotions Questionnaire (AEQ; Pekrun et al., 2002), and the Self-regulated Learning, Emotions and Motivation Computerized Battery (LEM-B; Mega et al, 2014), which comprise three separate self-report questionnaires on student self-regulated learning, emotions and motivation. More recent research in intelligent tutoring systems has investigated other methods of identifying student emotions and affective states during learning such as expert coding of video data and field observations (Craig, Graesser, Sullins, & Gholson, 2004; Dragon et al., 2008; Ocumpaugh, Baker, Gaudino, Labrum, & Dezendorf, 2013; Woolf et al., 2009). Despite being labor-intensive and time-consuming, observational measures provide some advantages in identifying high-level student learning behaviors and emotions (Winne & Perry, 2000).

With the recent advancements in technology-based learning systems, native tools are now available to assess students' emotions and affective states during learning. These tools include computer interaction logs, dialogue cues, as well as physiological sensors (Arroyo et al., 2009; Baker, Gowda, & Wixon, 2012; D'Mello et al., 2008). Past studies that have identified student emotions and affective states using such methods have found relationships with student learning outcomes and achievement (Pardos, Baker, San Pedro, Gowda, & Gowda, 2014), as well as future college enrollment (San Pedro, Baker, Bowers & Heffernan, 2013).

Advantages and limitations of fine-grained measures

Self-report measures have been shown to hold certain advantages such as being an efficient and practical method of obtaining information about a study participant, as well as providing informational richness to the researcher (Paulhus & Vazire, 2005), with its potential to provide contextual clues and data. However, the use of self-reports also suffers from various disadvantages, such as being subject to issues of credibility and accuracy due to inaccuracies in memory or self-deception (Paulhus & Vazire, 2005). Self-report items and questions may also be interpreted differently from intended by study participants, which could affect the cognitive validity of the self-report measure (Karabenick et al., 2007). Comparisons made between self-report and fine-grained trace measures have found differences in their correlation with student learning outcomes. One study, compared students' level of bias and accuracy in self-reporting their achievement judgments and self-regulatory strategies to the trace computer logs that were collected throughout the study (Winne & Jamieson-Noel, 2002).found that while the students' judgments of achievement was significantly correlated to their actual scores. However, the self-reported self-regulated learning strategies were incongruent with trace computer logs and were not correlated to the actual strategies used. These results indicate that student self-reports of study tactics and strategies are often not congruent with actual test scores and thus tend to be fallible (Winne & Jamieson-Noel, 2002; Winne & Perry, 2000).

Given the shortcomings of self-report measures, it would thus be more useful for researchers to use these measures in conjunction with more fine-grained trace measures, such as behavior and affective state models built through computer log data, which can provide a more balanced perspective to the researcher and practitioner. With the increasing use of technology-based learning, it is now more possible than before to obtain such fine-grained measures of student SRL behavior, strategy use, and affective states through computer log data. Such measures would provide much-needed support to bolster and balance other self-report measures of SRL to provide researchers with a more accurate and reliable assessment of student self-regulated learning.

Despite the advantages of building behavior and affect models based off fine-grained measures of SRL, there are various limitations. Behavior and affect models are built based on the computer log data specific to a technology-based learning platform. As such, issues arise with generalizing these models across populations (Jaclyn Ocumpaugh, Baker, Gowda, Heffernan, & Heffernan, 2014) and learning systems, despite a few exceptions (Paquette et al., 2013). Because behavior and affect models built based on computer log data are highly dependent on the computation of features that captures the students' interactions with the specific learning platform, the type of features generated is contingent on the learning system itself, making it difficult to apply the same sets of features across different systems.

Trace measures of SRL behavior have been constructed with some success within several online tutoring systems, such as the Cognitive Tutor in Geometry and Algebra (Aleven, Mclaren, Roll, & Koedinger, 2006; Roll et al., 2011), Betty's Brain (Biswas, Jeong, Kinnebrew, Sulcer & Roscoe, 2010), and several others (Greene, Azevedo, 2010; Lee, Lim & Grabowski, 2010). In the study by Aleven and colleagues (2006), features were created from computer log data to identify sequential patterns in student help-seeking behavior and their relationships with learning outcomes.

Problem Statement

Several model frameworks have been proposed to describe the different components of self-regulated learning (SRL) and its influence on the learning process. Some recent theoretical frameworks have emphasized distinct components of SRL that occur during different phases of learning. The growing popularity of frameworks that involve multiple components of self-regulated learning has led to an increasing need for real-time assessments of different types of student SRL behavior within authentic contexts. To address this need, methods have been developed to assess the sequential patterns of student behavior, allowing practitioners and researchers to make better inferences about student SRL behavior over time. (Zimmerman, 2008). While various measures have been successfully implemented within technology-based learning platforms, most still involve post-hoc student self-reports in the form of think-aloud protocols, structured diaries and microanalytic surveys (Winne & Hadwin, 1998; Zimmerman, 2008). As discussed above, student emotions have also been shown to impact SRL behaviors during learning (Blair & Diamond, 2008; Reschly, et al, 2008; Pekrun, et al, 2002; Wolters, 1998). While research on the assessment of student emotions and affective states within different learning contexts have grown steadily in recent years (Ai et al., 2006; Calvo & D'Mello, 2010; Picard, 1997; Rodrigo & Baker, 2009; J. Sabourin, Mott, & Lester, 2011), few studies have looked at how to quantitatively measure the relationship between student emotions and SRL behavior, especially within self-directed learning environments. Established methods that do assess both SRL behaviors and student emotions are still largely confined to post-hoc self-report measures, rather than the application of fine-grained observational measures of student behavior and affect that can be assessed in real-time. These self-report measures have been shown to be less accurate than fine-grain measurements and cannot capture the temporal patterns of student emotions relative to SRL behaviors.

This dissertation project thus aims to present a novel method to quantitatively assess the relationship between student emotions and aspects of self-regulated learning within a learning environment. Additionally, in order to fully realize the potential of the new technology-enabled learning platforms, educators will need to develop automated methods for tailoring educational experiences to students when live human intervention is not possible. In traditional settings, educators can observe cues for when students are struggling and not utilizing SRL behaviors, then intervene to help support the student. In comparison, self-directed learning platforms need to detect and pre-empt the occurrence of student struggles to adapt its programming to ensure continued student learning.

Recent efforts have used a variety of increasingly sophisticated methods to assess various cognitive components of SRL as temporal “events”, as well as student emotions during learning. However, much less work has been done to quantify the mediating relationship between student emotions and self-regulated learning within the same learning context. Evaluating the temporal patterns of emotion associated with SRL will thus improve the understanding of the relationship between student emotions and self-regulated learning behavior in real-time. Furthermore, incorporating both cognitive and emotional features may improve the accuracy of SRL measurement and prediction.

Many studies have examined the relationship between self-regulated learning and achievement, while others have investigated the relationship between student emotions and achievement in an open-ended educational game context. However, far fewer studies have examined the relationship between student emotions and student self-regulated learning behavior. Consequently, part of the analyses in this dissertation attempts to identify components of student self-regulated learning behavior (Study 2), and student affective states as they occur in conjunction with behavioral indicators of student self-regulated learning. From there, I will attempt to isolate specific patterns in affective states that could be indicative of the specific components of SRL, as well as various cognitive components of SRL.

Research Objectives

This dissertation aims to build models that can predict student affective states in the context of an open-ended educational game platform, and using these predictions, identify patterns in student affective states that indicate specific components of SRL, as explained in detail below.

Research Questions

The research questions for this dissertation are:

1. Which machine learning models best predict student affective states in the context of an open-ended learning platform?
2. What are the most common patterns in student affective states (engaged concentration, boredom, confusion, frustration, and delight) that co-occur with self-regulated learning behaviors within an open-ended learning platform?

Significance

The measures of various aspects of SRL in intelligent tutoring systems and online learning platforms have evolved from trait-based formats (in the form of questionnaires and observations) to temporal-based formats (through think-aloud protocols and trace computer logs). While some studies have evaluated various aspects of student SRL behavior and strategies, such research has focused mainly on intelligent tutoring systems. To be able to identify student SRL behavior and strategies in a less structured learning environment would provide a look at how spontaneous use of various SRL strategies and behavior may occur in such environments.

The ability to identify student affective states during the learning process within online learning environments has allowed teachers and facilitators to provide timely interventions to students at risk of becoming disengaged within these learning platforms. Additionally, identifying the temporal patterns of students' affective states and their relationship to SRL could provide facilitators with a better understanding of how to encourage student SRL during learning in less structured learning environments.

This dissertation project provides a new methodological approach to identify self-regulated learning behavior among students by identifying patterns in student emotions. In addition, the project aims to demonstrate a strong relationship between student emotions and self-regulated learning behavior, thus furthering the research in this field on the role of student emotions within self-regulated learning.

CHAPTER II.

LITERATURE REVIEW

This literature review section will discuss and compare the methodologies used to assess self-regulated learning (SRL). It will focus on identifying student engagement in SRL within an exploratory educational game platform. First, this section will discuss the cognitive science research around theoretical frameworks and measurement methods of student engagement in self-regulated learning strategies. Then, this section will explore existing research on automated behavior and affect models which have been successfully built into technology-based software., These systems allow the evaluation and assessment of student engagement during self-regulated learning behaviors and strategies through the analysis of fine-grained measures.

Self-Regulated Learning

While there are several theoretical definitions of self-regulated learning, the common feature among these definitions is that self-regulated learners are active learners who manage their learning through monitoring of their strategy use (Boekaerts, Pintrich, & Zeidner, 2000; Paris & Paris, 2001; Pintrich, 2000; Winne & Hadwin, 1998; Winne & Perry, 2000; Zimmerman & Schunk, 2001). This process is mainly affected by the learners' characteristics as well as the contextual situations that students are learning in (Pintrich, 2000). Within these definitions lies the assumption that SRL involves processes and responses that students must proactively initiate, which imply that students may choose not to self-regulate during learning when they could (Zimmerman & Campillo, 2003). This creates an issue that researchers try to address with various cognitive models for SRL. These models have been posited to provide researchers with theoretical frameworks through which to examine how students evaluate and adapt their learning. In this way, these models seek to explain how and when students choose to self-regulate during learning, as opposed to when they do not. This project will focus on the SRL models and frameworks that have been applied to learning with hypermedia and technology-based software.

Several models of self-regulated learning have been developed by different groups of researchers over the last two decades, based on a diverse set of theories such as operant and phenomenological perspectives, as well as social-cognitive and information processing theories. Most of these theoretical model frameworks involve three or four main phases: the phase involving goal setting or task definition, followed by the phases involving the selection and monitoring of learning strategies and tactics, and lastly the phase of reflection and evaluation of learning (Pintrich & Zusho, 2002; Zimmerman, Bonner & Kovach, 1996; Zimmerman & Schunk, 2003). Within each phase, there are also various aspects of learning that the student may try to control or regulate, including internal conditions such as cognitive, affective (motivational) and behavioral aspects of the individual student, as well as the external context or environment that may be modified to improve the student's learning process (Pintrich & Zusho, 2002).

One example of the theoretical frameworks that have tried to address the various theories definitions around SRL was proposed by Zimmerman (2000), based on the social-cognitive perspective and consists of three main phases of the SRL process: Forethought, Performance, and Self-reflection. Processes in the forethought phase occur before actual learning takes place, and influences the learning context that follows. Performance phase processes occur during learning and influence both student actions and attention, as well as the self-reflection phase afterward. The self-reflection phase occurs after the learning actions have taken place, and influence the student's forethought process relevant to subsequent learning efforts, thus completing an SRL cycle (see Figure 1; Zimmerman & Campillo, 2003).

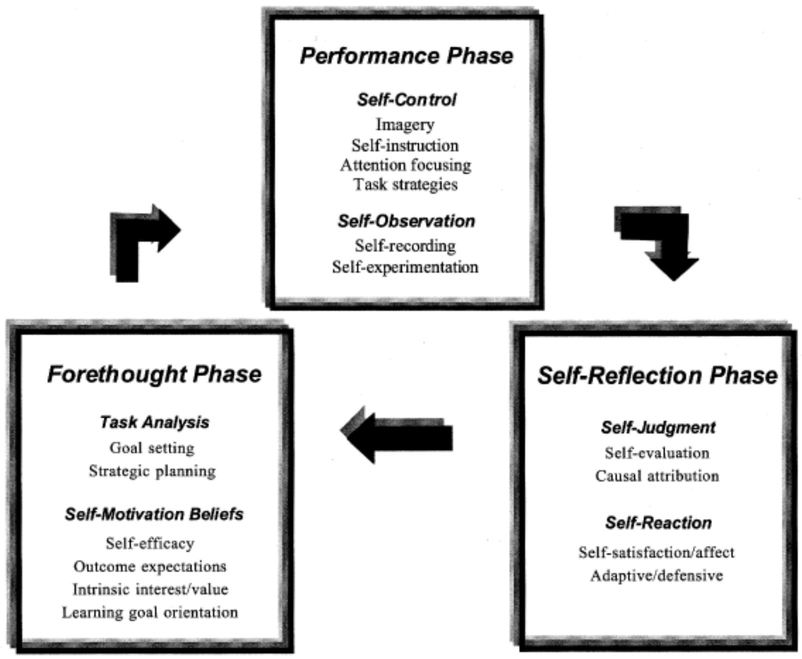

Figure 1. Socio-cognitive framework for self-regulated learning (Zimmerman & Campillo, 2003).

Another theoretical framework was proposed by Winne & Hadwin (1998; see Figure 2), and based on the information processing perspectives. This theoretical framework is made up of 4 phases of a student's learning, involving task definition, goal setting and planning, enactment and lastly, adaptation (Winne & Hadwin, 1998). According to Winne & Hadwin (1998), these phases of learning may be affected by various task and cognitive conditions, which may be transformed through cognitive operations and strategies. This interchange of operations and strategies that affect the students' learning conditions would, in turn, lead to re-evaluations of the different phases of learning, which may result in metacognitive monitoring and updates to the learning products within each phase.

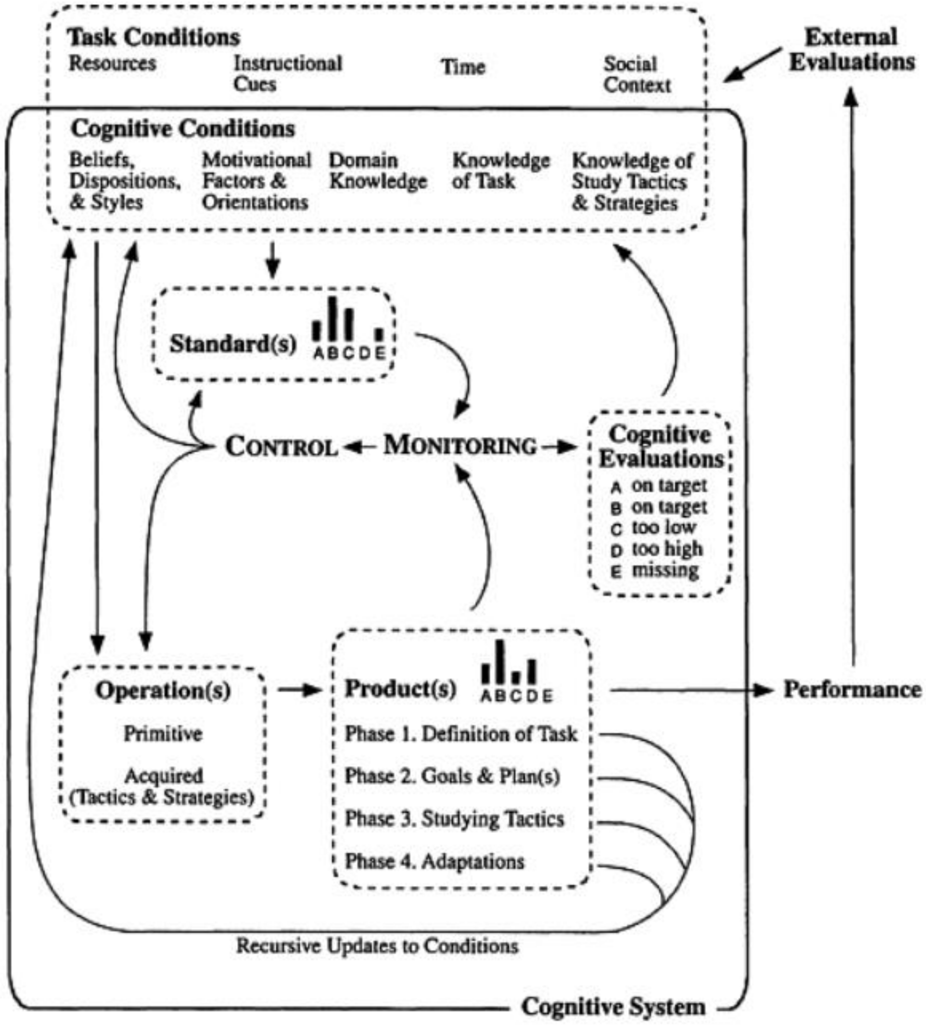

Figure 2. Framework for self-regulated learning based on informational processing theories (Winne & Hadwin, 1998).

Based on the various model frameworks within SRL, interventions targeting the development of self-regulated learning among students have focused largely on the sub-processes within each phase or stage of a model framework. For example, interventions to improve students' use of goal-setting and task definition skills have been implemented to help students better understand their learning tasks and progress (Kitsantas, Robert & Doster, 2004;

Latham, 2004; Latham & Locke, 2007; Locke & Latham, 1990; Manderlink & Harackiewicz, 1984; Zimmerman & Kitsantas, 1999). Other interventions have targeted the metacognitive monitoring sub-process in learning with a selection of strategies such as thinking aloud and help-seeking strategies. Various research studies have also been conducted on the effects of holistic combinations of several self-regulated learning strategies (Cleary & Zimmerman, 2004; Winne, Nesbit, Kumar, Hadwin, Lajoie, Azevedo, & Perry, 2006). In Cleary & Zimmerman (2004), for instance, a school-based holistic program called the Self-Regulation Empowerment Program (SREP) was implemented within middle and high school classroom environments and was found to have positive effects on student achievement and motivation.

Theoretical Frameworks of SRL

Forethought Phase

Most theoretical frameworks for the SRL process include a phase before the actual learning phase. During this phase, students define the task and plan their actions during learning. Research on various interventions that encourage students to engage in setting goals or plan their activities towards the learning objectives has been shown to improve academic outcomes (Fireman & Solomon, 2003; Graham & Harris, 1989). Researchers studying goal-setting in learning contexts found that the impact of goals on behavior depends largely on the specificity, proximity and difficulty of the goals (Bandura, 1988; Locke & Latham, 1990; Latham & Locke, 2007). For example, specific goals and proximal goals have been found to lead to better self-efficacy because progress is more easily tracked (Schunk, 1983; Maderlink & Harackiewicz, 1984; Locke & Latham, 1990). As a result, students show an improvement in the ability to self-regulate during learning. In the subject of reading and writing, for example, interventions that introduce explicit instruction in pre-reading and pre-writing strategies have shown to be beneficial for elementary school students (Pressley, Johnson, Symons, McGoldrick & Kurita, 1989). Research on goal-setting found ample evidence showing that self-set goals improve not only students' achievement outcomes, but also their metacognitive abilities and self-efficacy (Latham & Locke, 2007), thus constituting a key component within self-regulated learning (Latham, 2004).

In addition to goal-setting, explicit instruction in strategic planning activities prior to the learning task has also been found to yield benefits to students learning outcomes. For instance, explicitly teaching 6th-grade students with learning disabilities a planning strategy for writing opinion essays was found to improve students' ability to address the topic in their writing (De La Paz & Graham, 1997; Graham & Harris, 1989).

Performance Phase

The performance phase of learning, as defined by Zimmerman (2000), occurs when the student begins interacting with the learning task or content. There are two categorizations of self-regulated learning behavior during this phase: self-control and self-observations (Zimmerman, 2002). Several components of the SRL process occur during this period, as the student controls her behavior, emotions and attention during the learning process. It is also during this period of learning that students engage in self-observations, as they experiment with the task strategies employed and their effectiveness throughout the learning process, which help to inform their future attitudes and emotions towards the learning within the domain.

Methods that encourage students' self-awareness through self-observational techniques have been found to isolate “the source of error, confusion, or inefficiency” during learning (Zimmerman & Paulsen, 1995, p. 15), and increase their deliberate use of self-regulation

(Ferrari, 1996). Such techniques were thought to empower students by allowing them to make more accurate attributions for poor learning performance (Zimmerman, 1989), and enhance access to cognitive processes as well as other internal states (Gibbons, 1990). For instance, a study involving a form of self-observation on the speech fluency of college students found positive effects with the implementation of self-observational techniques. Specifically, the implementation of self-observational strategies resulted in reductions of self-recorded verbal non-fluencies (e.g. "um," "uh," "er," etc.) (Mace & Kratochwill, 1985). Student groups that made use of this self-observational technique experienced a decrease in the use of verbal non-fluencies as compared to groups that did not employ this self-observation method. Other studies have also found that self-observational techniques can highlight a student's attention to her actions during learning, thus facilitating problem-solving performance (Fosnot, Forman, Edwards & Goldhaber, 1988; Welsch, 1991). An example of such a study is one by Fireman, Kose & Solomon (2003), where elementary school students were shown video recordings of their spontaneous performance during a problem-solving Tower of Hanoi task, and their problem-solving performance was compared with students who watched video-recordings of other students' performance at the task. Results from the study showed that video self-observation significantly enhanced the acquisition and transfer knowledge required to complete a more difficult problem-solving task (Fireman, Kose & Solomon, 2003). On the flip side, however, such self-observation may also have detrimental effects on student learning. Reid & Harris (1993), for instance, found that, although monitoring of off-task behavior reduced the occurrence of student off-task behavior, it also seemed to result in poorer learning, as students became overly focused on their behavior during the learning tasks.

Self-Reflection Phase

After the learning phase, there is often a post-learning phase where students reflect on their performance and strategies and modify their behavior or learning strategies in future learning cycles. There are two main classes of self-reflection: self-evaluation and self-reaction. Self-evaluation refers to the comparisons of self-observed performance to other standards of performance (Zimmerman, 2002), such as previous standards or other external standards of performance. Self-evaluation or judgment may also involve causal attribution, where students attribute the causes behind their performance to internal or external factors. In turn, certain forms of causal attribution may affect students' motivation and beliefs about their learning and are hence very important as well to the learning process (Cleary & Zimmerman, 2004; Zimmerman & Kitsantas, 1999). To date, multiple types of studies have been conducted to assess the effects of different types of causal attribution to students' motivation and beliefs, as well as the effects of different forms of interventions within this phase on student learning outcomes. In Schunk's studies (1996) on elementary school students, for example, implementing self-evaluation strategies in combination with learning goals was found to increase student task orientation and lowered ego orientation, as well as improving self-efficacy and motivation. Among undergraduate students, frequent self-evaluation was shown to produce positive results in student achievement as compared to infrequent self-evaluation (Schunk & Ertmer, 1999). More recent studies such as one conducted on middle school students also showed that the implementation of self-evaluative strategies during learning positively affected student skill acquisition (Kitsantas, Reiser & Doster, 2004).

In addition to self-evaluation, self-reaction also occurs during this phase of learning. Self-reaction refers to the wide variety of students' reactions to their performance after the learning process, which ranges from strategy persistence to change as well as from greater goal commitment and goal adjustment (Zimmerman & Martinez-Pons, 1992). This process is also known as adaptive or defensive inferences, which refer to conclusions drawn by students on whether or not to modify their learning strategies during future learning attempts (Zimmerman, 2000). From the socio-cognitive perspective, self-reactions may involve environmental, personal and behavioral self-reactions, as students make adaptive or defensive inferences based on their learning performance. Recent studies in these areas have also found correlations between students' use of self-regulated learning strategies and their affective states as well as levels of self-satisfaction. For instance, Zimmerman & Bandura (1994) found that students who displayed some level of satisfaction and who attributed poor performance outcomes to their choice of learning strategy were more likely to make adaptive inferences. In comparison, students who were dissatisfied with this performance and attributed poor performance to uncontrollable factors tended to make defensive inferences. In addition, adaptive inferences were found to lead to improvised strategic planning and shifts in goals that benefited future performance (Cleary & Zimmerman, 2001).

Assessment of Self-Regulated Learning

Self-regulated learning is generally measured in two ways: as an aptitude, and an event, with a variety of measurement protocols being commonly used. In this section, I will discuss the use of several protocols that have become increasingly popular in the measurement of self-regulated learning as a temporal "event," focusing in particular on measurement protocols that are more commonly used in online learning environments. These measurement protocols include think-aloud protocols, error detection tasks, trace computer logs and expert observations of performance. In particular, I will focus my discussion on the use of trace measures, which are unique to online learning environments and other learning systems where the use of a computer is integral to learning.

Think-aloud measures

Think-aloud protocols involve students reporting about thoughts and cognitive processes while performing a task, and provide richer information to researchers on how students engage in SRL behavior throughout an entire learning task. It is thus thought to be more appropriately aligned to the dynamic, event-based definition of SRL (Greene, Robertson & Costa, 2011). Learning tasks in which think-aloud protocols have been employed include complex science topics, as well as history and math (Greene et al., 2011). While think-aloud measures have been employed across a wide range subjects and grade levels (Winne & Perry, 2000), few standard procedures exist for this measure. A large section of self-regulated learning research with think-aloud protocols involve the identification of cognitive and metacognitive processes students undergo when learning from text materials (Fox, 2009). Other studies are focused on relationships between student use of SRL strategies and the development of mental models in online learning platforms (Azevedo & Cromley, 2004; Greene & Azevedo, 2007), as well as relationships between the types of SRL strategies used when students fail to learn (Azevedo, Winters & Moos, 2004)

Error detection tasks

To trigger student use of SRL, error detection tasks have been used to allow researchers to observe whether students detect errors in their tasks and what they do upon detection (Perry & Winne, 2000; Pintrich, Wolters & Baxter, 2000). Within online learning systems, eye-tracking has been used as a fine-grained indicator for student monitoring learning materials for errors.

Observations of Performance

Recent research in SRL has expanded to include the contextual relationships in student self-regulated learning. The advantages of observational measures are thus to provide contextual information about learners' behaviors, hence addressing the limitations of self-report measures. Studies that employ the use of observational measures, such as the Child Independent Learning Development (CHILD) (Bryce & Whitebread, 2012; Whitebread et al., 2009), are mostly targeted for younger children in the 3-5 year age range, for whom self-report measures would not be appropriate.

Trace computer logs

Trace methods provide observable fine-grained indicators about students' cognitive processes as they engage with a task. Recent studies that directly measure student engagement in self-regulated learning behavior and affective states employ a variety of analyses methods with the available computer data logs to explore different aspects of student self-regulated learning. Furthermore, the types of online learning systems through which student engagement in self-regulated learning is being evaluated are varied, ranging from online software aiming to improve student studying techniques (Winne & Jamieson-Noel, 2002), to intelligent tutoring systems with teachable agents (Biswas, Roscoe, Jeong & Sulcer, 2009; Bouchet, Azevedo, Kinnebrew & Biswas, 2012).

One example of an online software used in such studies is the gStudy program, which is a shell program that provides students with a learning kit to study about any given topic in any verbal, visual or written format (Winne, Nesbit, Kumar, Hadwin, Lajoie, Azevedo & Perry, 2006). The gStudy program provides cognitive tools that students can use to engage with the different forms of multimedia information, such as analyzing, annotating, classifying, organizing and cross-referencing. These cognitive tools were designed based on cognitive research that encourages the development and engagement of solo/collaborative learning as well as problem-solving skills with student use. Research studies that examined students' engagement in SRL strategies using gStudy as a platform makes use of computer trace logs to identify instances where students make use of certain cognitive tools provided during learning, which may indicate various aspects of engagement in SRL strategies (Winne & Jamieson-Noel, 2006; Hadwin, Oshige, Gress & Winne, 2010). Student interactions with the gStudy tools that indicate engagement in certain types of SRL strategies are coded and recorded. For example, student use of the setting goals strategy in the forethought phase of learning was coded as such when students click on the Objectives button to view objectives right at the beginning of the learning session. Similarly, student use of a planning strategy in the forethought phase of SRL may be inferred if students scrolled through the information first before using the annotation or notes tools. Frequencies of these actions are matched to equivalent self-report items on the Motivational Strategies and Learning Questionnaire (MSLQ) developed by Pintrich et al. (1991) calculated to determine the level of student engagement in the respective SRL strategies.

The investigation of computer trace data in other technology-based learning platforms differs in various ways both in part because of the differing structures of these platforms. For instance, the MetaTutor (Bouchet, Azevedo, Kinnebrew, & Biswas, 2012) and Betty's Brain (Biswas, Roscoe, Jeong, & Sulcer, 2009) are adaptive tutoring systems with multiple agents, while platforms like gStudy and ASSISTments (Feng, Heffernan, & Koedinger, 2009) are adaptive online tools that facilitate students' learning as they practice and learn new content through texts and practice problems respectively. Still others, such as Physics Playground, provide open exploratory environments that help students apply academic content to the learning environment.

MetaTutor is an adaptive tutoring system that teaches biological science content (Bouchet et al., 2012). MetaTutor is a system that is grounded in a theory of SRL and thus contains a combination of features that encourage student engagement in various SRL strategies and process, including four pedagogical agents that function as embodiments of the four main phases of SRL based on Winne & Hadwin's (2009) theoretical model framework. These pedagogical agents guide students through the learning process, prompt them to engage in the various self-regulated learning strategies, and provided adaptive feedback on their actions within the tutoring system. To measure students' overall progress in SRL strategies, Bouchet et al. (2012) employed the use of multiple measures to track and code students' use of cognitive, affective and metacognitive processes at different points during learning. They then made use of clustering and pattern mining analyses techniques to identify distinct patterns of student actions and behavior within the tutoring system that could allow researchers to help facilitate real-time adaptation of the system to cater to different types of student learning. The pattern mining technique and clustering allowed the researchers to identify specific student actions that could differentiate between the various types of student learning.

The Betty's Brain program (Biswas et al., 2009) is an adaptive tutoring program and is similar to the MetaTutor program in that it also involved intervention elements in the form of multiple online agents that help to promote students' use of metacognitive strategies during learning. Instead of the four pedagogical agents present in the MetaTutor, however, Betty's Brain program consisted of two: a mentor agent Mr. Davis, as well as a teachable agent named Betty. This learning platform detects instances during learning when the students' behavioral patterns indicate that metacognitive feedback may be useful. One example of such metacognitive feedback is when the student requests a quiz after teaching a computer agent a given topic using concept maps, to assess the computer agent's learning progress but does not make improvements to her concept maps.

In studies involving other intelligent tutoring systems, Roll, and colleagues (2006, 2007) examined students' help-seeking behaviors within the Cognitive Tutor in Geometry learning platform, through their behavioral patterns before and after asking for hints within the tutoring platform. They then created a model framework to differentiate between groups of students who differ in their help-seeking behaviors when using the Cognitive Tutor platform and built a help-seeking detector based on this framework.

On the other hand, computer logs of student actions within the learning platform were used in a slightly different manner in Winne & Jamieson-Noel's (2002) cross-platform researching system called gStudy. Timestamped log data of students' actions within the platform were logged and the frequencies of the selected student studying actions calculated at the end of the study session. These study actions were created based on items in the traditional MSLQ questionnaire, including student behaviors such as making up questions, outlining goals and summarizing ideas.

Development of Affective State Models

While research in the field of psychology has found relationships between student self-reaction and the future use of self-regulated learning strategies and learning outcomes, there is a limited range of methods that could be used to identify what student behaviors and affect would manifest as a result of specific self-reactions or self-satisfaction. With new affordances in technology, however, improved methods have been created to detect user affect and behavior through the users' interactions with technology-based platforms.

The development of models that can automatically detect student affect now constitutes a considerable body of research (Calvo & D'Mello, 2010; Picard, 1997). Research in this area is particularly focused on computerized learning contexts (Ai et al., 2006; Rodrigo & Baker, 2009; Sabourin, Mott, & Lester, 2011), where researchers have successfully built affect-sensitive learning systems that aim to enhance learning outcomes (Arroyo et al., 2009; Dragon et al., 2008; Graesser & McNamara, 2010). The definitions of affect and affective states in this study are based on affective phenomena covered in the field of affective computing and include emotions, feelings, moods, attitudes, and temperaments. Such affective states and phenomena may also be evaluated via several perspectives that have been derived from traditional theories of emotion. Some of these perspectives include expressions, embodiments, cognitive appraisal, as well as social constructs (Calvo & D'Mello, 2010). Human emotions and affect have traditionally been evaluated through several perspectives, including expressions, embodiments and cognitive appraisals. Emotional expressions refer to various facial expressions of basic emotions that have been universally recognized, such as anger, happiness, and disgust (Darwin, 2002; Russell, 1994). Embodiments of emotion, on the other hand, refer to the physiological changes that an individual's body undergoes when he or she experiences an emotion (James, 1884). Lastly, cognitive appraisals of emotion refer to emotions that are produced as outcomes of an unconscious process of evaluating an event or situation based on some factors such as urgency, novelty, etc. (Arnold, 1960; Dalgleish, Dunn & Mobbs, 2009). Many of the affect models built in the field of affective computing are hence mostly built based on the various traditional theories of emotions posited in research.

In general, researchers attempting to develop affect models have developed systems falling into two categories: interaction-based models (Baker & Ocumpaugh, 2015) and physical sensor-based models (Calvo & D'Mello, 2010). Many successful efforts to detect student affect in intelligent tutoring systems have used visual, audio or physiological sensors, such as webcams, pressure sensitive seat or back pads, and pressure-sensing keyboards and mice (AlZoubi, Calvo, & Stevens, 2009; Pantic, Pantic, Rothkrantz, & Rothkrantz, 2003; Sebe, Cohen, Gevers, & Huang, 2005; Zeng, Pantic, Roisman, & Huang, 2009).

Interaction-based detection, too, (Baker & Ocumpaugh, 2015) has improved over the last decade. These models infer affective states from students' interactions with computerized learning systems (Baker, Gowda, & Wixon, 2012; Baker, Ocumpaugh, Gowda, Kamarainen, & Metcalf, 2014; Baker & Ocumpaugh, 2015; D'Mello et al., 2008; Paquette et al., 2014; Pardos et al., 2014). The fact that interaction-based affect models rely on student interactions makes it possible for them to run in the background in real time at no extra cost to a school that is using the learning system. Their unobtrusive and cost-efficient nature also makes it feasible to apply interaction-based models at scale, contributing to the growing field of research in the measurement of student academic emotions in the classroom (Baker & Yacef, 2009). For example, interaction-based affect detection has been useful in predicting long-term student outcomes, including standardized exam scores (Pardos et al., 2014) and college attendance (San Pedro et al., 2013). Basing affect detection on student interactions with the system, however, give rise to issues with generalizing such models across populations (Jaclyn Ocumpaugh et al., 2014) and learning systems. Because interaction-based models are highly dependent on the computation of features that captures the student's interactions with the specific learning platform, the type of features generated is contingent on the learning system itself, making it difficult to apply the same sets of features across different systems.

This project proposes the use of interaction-based models of student affect within classroom settings to predict patterns in student affective states that are associated with students' behavioral indicators of specific SRL strategies, in the context of 8^{th} and 9^{th}-grade students using the Physics Playground exploratory game platform

CHAPTER III.

METHODOLOGY

In this dissertation project, I make use of computer interaction data from students participating in the Physics Playground platform (Shute, Ventura, & Kim, 2013). As part of a larger collaborative project, a team of researchers from Florida State University collected data from 137 students from the same school, who played Physics Playground for two hour-long sessions, two days in a row. Various forms of data was collected for this larger project, namely:

1) pre- and post- tests examining students' knowledge of simple machines
2) computer interaction log data recording students' actions within the game environment
3) video camera data recording students' facial expressions as they navigate the game environment
4) observational data of student affective states during gameplay

For this dissertation, I conducted analyses using the computer interaction log data collected for both Studies 1 and 2. In Study 1, I developed prediction models for students' affective states based on the computer log and observational data obtained from the data collection phase. In Study 2, I made use of this computer log data as well to generate feature indicators of specific types of student self-regulated learning behavior.

In addition to computer log data, video of student facial expressions during gameplay was collected to build detectors that could predict student affective states during learning. In the following sections, I describe how the data was collected for the studies. Following this, I discuss how the various affective states were modelled separately using computer log interaction data, such that predictions of these affective states may be made based on interaction data available. I then compare the resulting prediction models with similar detectors that were developed by

Bosch and colleagues (2014) using video data in the same learning environment. Finally, I applied these affective state detectors to predict the affective states accompanying student self-regulated learning behavior within the Physics Playground environment, and study the affective state patterns that co-occur with such behavior.

Physics Playground

Physics Playground, formerly known as Newton's Playground, (Shute & Ventura, 2013) is a 2-dimensional physics game where students apply various Newtonian principles as they create and guide a ball to a red balloon placed on screen (Shute et al., 2013). It offers an exploratory and open-ended game-like interface that allows students to move at their own pace. Thus, Physics Playground encourages conceptual learning of the relevant physics concepts through experimentation and exploration. All objects in the game obey the basic laws of physics, (i.e., gravity and Newton's basic laws of motion). Students can choose to enter one of seven different playgrounds, and then play any of the approximately 10 levels within that playground. Each level consists of various obstacles scattered around the space, as well as a balloon positioned at different locations within the space (see Figure 3). Students can nudge the ball left and right but will need to create simple machines (called "agents of force and motion" in the game) on-screen to solve the problems presented in the playgrounds. There are four possible agents that may be created: ramps, pendulums, levers, and springboards. Students can also create fixed points (called 'pins') along a line drawing to create pivots for the agents they create.

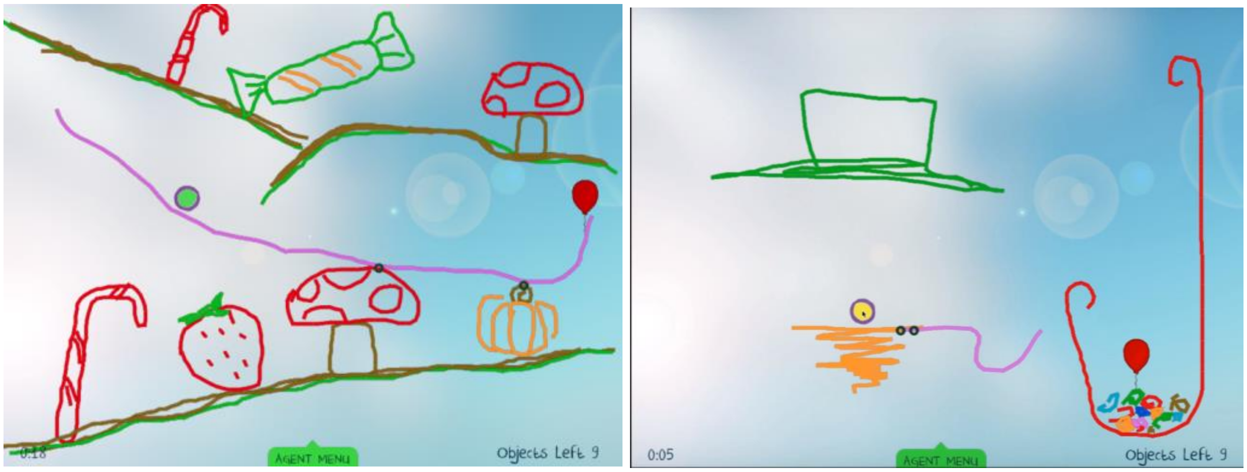

Figure 3. Screenshots of Physics Playground

Students use the mouse to draw agents that 'come to life' after being drawn, and use these agents to propel the ball to the red balloon. In other words, these agents or objects would behave as they would in a real-world context, such as responding to gravity, forces from other objects, etc. Students control the weight and density of objects through their drawings, making an object denser, for example, by filling it with more lines. Each level allows multiple solutions, encouraging students to experiment with various methods to achieve the goal and guide the ball towards the balloon. Silver trophies are awarded for achieving the goal objective while gold trophies are awarded for solutions deemed particularly efficient or creative (that perhaps includes fewer objects created), encouraging students to attempt each playground more than once. This unstructured game-like environment provides a rich setting in which to examine the patterns of students' affect and self-regulated learning as they interact with the game platform.

Data Collection

The dataset used in these analyses is obtained from 137 students as they engaged in the Physics Playground platform for a total of approximately 2 hours each, over 4 days in groups of 20-25. Students in the 8th and 9th grade were selected as participants for this due to the alignment of the curriculum in Physics Playground to the state standards at those grade levels. The study began with pre-tests of student content knowledge, followed by two sessions of actual gameplay with Physics Playground. Upon conclusion of the gameplay, students completed a post-test. Student learning outcomes were measured in the form of online pre- and post-tests on the first and last days of the study, that assessed their content knowledge and skills related to Newtonian physics.

The data collection phase lasted two days (days 2 and 3), during which multiple classes of students worked with Physics Playground in a computer lab. Class periods were 55 minutes in length and class sizes were about 20 students each. Both video data and computer data logs were collected simultaneously during this time over two full sessions of game play. Computer data logs were taken directly from the Physics Playground application, while video camera data recorded students' facial features throughout their interactions with the game platform.

The interaction log files contain all detailed data on students' attempts to complete each playground level and the time taken for each action. Segments of the raw interaction log files are shown below in Tables 1 and 2. Table 1 shows a segment of the raw computer log data that records the summary data of the ball attributes present in the game environment every second, while Table 2 shows a segment of the raw computer log data that records the positional and movement attributes of other objects and machines created in the game environment. A full list of the attributes recorded in the raw computer logs and their respective descriptions can be found in Appendix I.

Table 1

Segment of raw computer log data recording details and position of ball within a playground level in summary data logs.

Event	**Level_path**	**Badge_string**	**Ball_id**	**Game_time**	**Ball_position_x**	**Ball_position_y**	**Ball_velocity_x**	**Ball_velocity_y**	**Ball_touched_count**	**Mouse_distance**	**Mouse_draw_distance**	**fps**
Camera Start	.		.	.	.	.	.	.	.	.	.	.
Level Start	interactive_tutorial.level		1	.	.	.	.	.	.	.	.	.
Summary Data				1000	0	1.2749	0	-8.8556	0	0.9383	0	40
Lost				.	.	.	.	.	.	.	.	.
Summary Data				2000	0	1.0016	0	-0.6649	0	0.3602	0	40
Summary Data				3000	0	1.3176	0	-9.5201	0	0	0	40
Lost				.	.	.	.	.	.	.	.	.
Summary Data				4000	0	1.0064	0	-1.3287	0	0.004	0	40
Summary Data				5000	0	1.3633	0	-10.182	0	0.0027	0	40

Table 2

Segment of raw computer log data recording details and position of freeform objects drawn within a playground level.

userID	time	event	object_id	Start_step	mass	Mass_center_x	Mass_center_y	Pins_count	Position_x	Position_y
203	41487	Draw freeform	3	1812	2937	0.341782	0.000287	0	0	0
203	43858	Erase	-	-	-	-	-	-	-	-
203	44958	Summary data	-	-	-	-	-	-	-	-
203	44958	Draw freeform	4	5360	936	0.504168	0.001624	0	2	0
203	44958	Summary data	-	-	-	-	-	-	-	-

The first facet of the data contains identifiers for each student, the playground and level that was being attempted, as well as a timestamp for each event that occurred. The specific facets of data that were collected are grouped into different types of events. The first type of event is related to initiating or ending a level as either a “Level Start”, “Level Restart” or “Level End”. Level success (whether or not a badge had been achieved) and time is recorded each time a level is started, restarted, or ended,

Within the game space, the ball is subject to the normal laws of physics and gravity, and would thus either move by itself due to gravity, or has to be clicked on to move in either the left or right direction. The summary data events thus record the ball’s position and distance it has moved (if any) every second in the game space.

Single and multiple object events such as drawing and erasing of objects and pin will report the positional, movement and size data of each of the objects and pins created in the game environment. Student actions involved in the creation of the freeform objects are also recorded, such as the distance that the mouse travels in the game space, the distance the mouse travels while drawing a line, even the number of mouse clicks that occurred that did not result in any object being created or ball being moved. Object and object interactions such as 'collisions' are also recorded as an event. Additionally, when the game identifies one of the objects that the student drew as a simple machine, the events recorded will reflect the types of machine created, and other data recorded will include data on the position, rotational velocity, strength and direction at which the machine propels the ball. Data is also recorded on any ball movement and the distance it moves through the machine and within the game space. 67 raw attributes were recorded in the raw computer interaction log data, which were used in this dissertation project to generate aggregated features representing various patterns in student behavior. Overall, 27 action events may be recorded by the Physics Playground platform as the student navigates through the game environment (see Table 3).

Table 3

List of possible events recorded in the raw computer data logs of Physics Playground, during student gameplay.

Event Type	**Description**
Camera Start	Start of the video camera recording
Click	Mouse click
Collision	Collision between ball and object/machine/game environment
Draw Freeform	Freeform object created
Draw Pin	Pin object created

Erase	Erase object created
Game End	Student exited game
Hover Tutorial	Student mouse hovered over Tutorial icon
Level End	End of playground level
Level Pause	Student paused playground level
Level Restart	Student restarted playground level
Level Start	Student started playground level for the first time
Lever	Lever object created
Lost	Ball was lost from game environment
Menu Focus	Student clicked on game menu
Nudge	Student clicked on ball to nudge the ball to move
Object Limit	Student reached the number limit of objects created
Pendulum Object	Pendulum object created
Pendulum Strike	Pendulum object struck ball
Ramp	Ramp object created
Springboard	Springboard object created
Stacking	Student stacked lines/objects on top of one another
Stacking Warning	Warning triggered when student stacked too many lines/objects on top of one another
Summary Data	Summary data of attributes of objects present in game environment, recorded periodically every second
Watch Tutorial	Student watched tutorial of game environment

As part of the larger project, data on student facial expressions were also recorded via video through computer webcams attached to each computer station that the students were using to explore the game environment.

Field observations of student affective states were also recorded during the data collection phase. During these observations, observers coded students' affective states and behavior following the BROMP 2.0 protocol and using the HART (Human Affect Recording Tool) app on an Android. The Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP 2.0) is a momentary time sampling system that has been used to successfully study behavioral and affective indicators of student engagement in a number of learning environments (Baker et al., 2012; Baker, D'Mello, Rodrigo, & Graesser, 2010; Paquette et al., 2014; Rodrigo et al., 2009). BROMP coders observe each student individually, in a predetermined order. They record only the first predominant behavior and affect that the student displays, but they have up to 20 seconds to determine what that might be. To reduce observer effects, observations were conducted using side glances so that students would not be aware that they were being observed. Observations were coded based on the raters' judgment of students' actions, utterances, facial expressions and body language, as well as their interactions with other students or the teacher in the classroom. These are in line with the information used to code student emotions in previously-used methods (e.g. Bartel & Saavedra, 2000), and follow Planalp et al.'s (1996) description that identification of affect is more accurate using multiple cues, rather than based on any individual cue.

The coding process was implemented using the Human Affect Recording Tool (HART) application for Android devices (Ocumpaugh et al., 2015) which enforces the protocol while facilitating data collection. Both the Android devices used in the observations and the Physics

Playground software logging server were synchronized to the same internet time server during observations so that the logged student actions would correlate exactly with the observations. Interactions with the game environment during the twenty seconds before coding entry by the observer were aggregated as a clip and the data features distilled.

In Study 1, observations of student affective states used a coding schema that had previously been used in several other studies of student engagement. This schema included the affective states of *boredom, confusion, concentration, delight* and *frustration* as well as the behavioral states of *off-task*, *on-task* and *on-task conversation*. Given that the *concentration* affective state on its own could apply to both on-task and off-task students (eg. A student could be concentrating on her off-task activity); the *engaged concentration* affective state was derived from a combination of *concentration* affective state and *on-task* or *on-task conversation* behavioral states, in order to better capture the affective states of students who were concentrating on the task at hand. As a result, the *on-task* and *on-task conversation* states were ultimately dropped during the creation of our detectors.

Data Analyses

As mentioned above, data was collected as part of a larger project, from student gameplay in Physics Playground in various ways, including:

1) computer interaction log data,
2) video data of student facial expressions during gameplay, and
3) observational data of student affective states within 20-second windows during gameplay

From these sources of data, I made use of a combination of the computer log and observational data to build a set of affective state prediction models, and compared its performance with a similar set of prediction models built using a combination of the video and observational data collected in the larger project (Study 1). Aggregated features from the raw computer data logs are then generated to create SRL behavioral indicators and mapped to affective state predictions to identify specific affective state patterns co-occurring with SRL behavior in Study 2.

Two sets of analyses were conducted to answer my research questions in this dissertation project. In Study 1, field observations of student affective states were recorded and synchronized to aggregated features generated from raw computer log data. Aggregated features are measurable properties or characteristics of the target construct. In this case, the target construct is a student affective state, and the aggregated features created are measurable characteristics of a student's behavior when she is experiencing a particular affective state. The machine learning process is then applied on the dataset, which involves using supervised learning algorithms (a set of predefined hypotheses) to build a mathematical model of sample data (ie. training data), that can produce a set of classification rules to make correct predictions to a target variable. Examples of supervised learning algorithms include classification algorithms such as JRip and support vector machines (SVM), as well as regression algorithms such as logistic and linear regression. Specifically, in this study I make use of the machine learning process with a selection of five learning algorithms to identify student affective states (whether a student experiences a certain emotion or not) within the Physics Playground dataset. This process have been previously applied to similar affective state models created in a variety of intelligent tutoring systems, including Cognitive Tutor Algebra (Baker, Gowda, & Wixon, 2012) and Reasoning Mind (Miller, Baker, Labrum, Petsche & Wagner, 2014).

Various machine learning algorithms (namely JRip, J48 decision trees, step and logistic regression, and Naïve Bayes) are applied to this dataset to select features that correspond most strongly with each of the affective state observations, thus creating models, or detectors, that can predict a student's affective state based on a combination of aggregated features. A second set of prediction models, or detectors, were also built using a similar machine learning process, but using video data of student facial expressions during gameplay. The accuracy of these two sets of prediction models were then compared against each other using AUC as the performance metric. AUC refers to 'Area under the ROC curve', and provides an aggregate measure of performance across all possible classification thresholds (Fawcett, 2006). An ROC curve (receiver operating characteristic curve) refers to a graph that shows the performance of a classification model at all classification thresholds. One way of interpreting AUC is the probability that the model would rank a random positive example more highly than a random negative example. A model with an AUC of 1.0 means the model is 100% correct in its predictions, whereas an AUC of 0.5 implies that the model performs at chance level in generating correct predictions 50% of the time.

In Study 2, the affective state detectors from Study 1 were used to generate predictions for students' affective states throughout the whole dataset obtained during gameplay. A new set of aggregated features are created from the raw computer log data to identify playground level attempts where aspects of self-regulated learning behavior are shown. The affect predictions are then mapped to these level attempts, and sequential pattern mining is applied to the dataset to track how students' affective states change over time within students' attempts at each playground level. Sequential pattern mining is a popular data mining technique that automatically identifies frequent temporal patterns of actions in data (Agrawal & Srikant, 1995), and can also be used to detect differentially frequent behavioral patterns of different student groups (Kinnebrew, Loretz, & Biswas, 2013).

In the following chapters, I discuss in detail the methods and measures used in each study, and report the corresponding results.

CHAPTER IV.

STUDY 1: COMPARING PREDICTION MODELS FOR STUDENT AFFECTIVE STATES USING COMPUTER LOG DATA WITH MODELS BUILT USING VIDEO DATA

Student affective states during learning have been successfully predicted in prior studies using various methods such as physical sensors, conversational cues and log file interaction data (Baker et al., 2007; Baker, Ocumpaugh, Gowda, Kamarainen & Metcalf, 2014; D'Mello & Graesser, 2012). How these types of detectors compared against one another in terms of accuracy and performance, however, have yet to be investigated. In this dissertation project, I compare the performance of affective state models for the states of boredom, frustration, confusion, delight and engaged concentration based on two forms of data: face-based video data (Bosch et al, 2015), and computer log interaction data (Kai et al, 2015). The affective state detectors were built separately using comparable machine learning algorithms, and their respective performances (computed using the same metric) are then compared against one another.

Prediction models for student affective states using computer log data

We also built predictive models for same student affective states and behavior, this time using data from student interaction logs with the Physics Playground environment. Computer interaction log data of student actions were recorded during every second of gameplay. Segments of the raw interaction logs are shown below (Tables 4 and 5).

Table 4

Segment of raw computer log data recording details and position of ball within a playground level in summary data logs.

Event	Level_path	Badge_string	Ball_id	Game_time	Ball_position_x	Ball_position_y	Ball_velocity_x	Ball_velocity_y	Ball_touched_count	Mouse_dist	Mouse_draw_dist	fps
Camera Start	.		.	.	.	.	.	.	.	.	.	.
Level Start	interactive_tutorial.level		1	.	.	.	.	.	.	.	.	.
Summary Data				1000	0	1.2749	0	-8.8556	0	0.9383	0	40
Lost				.	.	.	.	.	.	.	.	.
Summary Data				2000	0	1.0016	0	-0.6649	0	0.3602	0	40
Summary Data				3000	0	1.3176	0	-9.5201	0	0	0	40
Lost				.	.	.	.	.	.	.	.	.
Summary Data				4000	0	1.0064	0	-1.3287	0	0.004	0	40
Summary Data				5000	0	1.3633	0	-10.182	0	0.0027	0	40

Table 5

Segment of raw computer log data recording details and position of freeform objects drawn within a playground level.

userID	**time**	**event**	**object_id**	**Start_step**	**mass**	**Mass_center_x**	**Mass_center_y**	**Pins_count**	**Position_x**	**Position_y**
203	41487	Draw freeform	3	1812	2937	0.341782	0.000287	0	0	0
203	43858	Erase	-	-	-	-	-	-	-	-
203	44958	Summary data	-	-	-	-	-	-	-	-
203	44958	Draw freeform	4	5360	936	0.504168	0.001624	0	2	0
203	44958	Summary data	-	-	-	-	-	-	-	-

Feature Engineering

From the raw computer log data, interaction features were generated that provided aggregated student actions that may indicate specific affective states and behavior. The feature engineering process for this part of the study was based largely on previous research on student engagement, learning, and persistence. The final set of features comprised 76 gameplay attributes that potentially contain evidence for specific affective states and behavior. Some attributes included:

- The total number of springboard structures created in a level
- The total number of freeform objects drawn in a level

- The amount of time between start to end of a level
- The average number of gold and silver trophies obtained in a level
- The number of stacking events (gaming behavior) in a level

Features created may be grouped into two broad categories. Time-based features focus on the amount of time elapsed between specific student actions, such as starting and pausing a level, as well as the time it takes for a variety of events to occur within each playground level. Other features take into account the number of specific objects drawn or actions and events occurring during gameplay, given various conditions. These features also involve the aggregation of specific attributes per student over varying grain sizes:

1) over a 20-second clip within a given playground level,

2) over a single playground level attempt, as well as

3) across all level attempts within a single playground level.

A complete list of the aggregated interaction features generated to build our affective state models can be found in Appendix II.

Of the 2087 BROMP field observations that were collected, 214 instances were removed as most of these instances corresponded to times when students werenot physically at their workstations. Additional instances were removed where the observer recorded a *?*, the code used when BROMP observers cannot identify a specific affective state or behavior. In total, 171 instances of affect and 63 instances of behavior were coded as ?. As a result, these instances did not contribute to the building of the respective affect and behavior models.

Within the field observations (Figure 4), the most common affective state observed was *engaged concentration* with 1293 instances (62.0%), followed by *frustration* with 235 instances (11.3%). *Boredom* and *confusion* were far less frequent despite being observed across both second and third days of observation: 66 instances (3.2%) for *boredom* and 38 instances (1.8%) for *confusion*. *Delight* was only coded on the third day, and was also rare (45 instances), but it still comprised 2.2% of the total observations. The frequency of *off-task* behavior observations was 4.0% (84 instances), which was unusually low compared to prior classroom research in the USA using the same method with other educational technologies (Ocumpaugh, Baker, Gaudino, Labrum, & Dezendorf, 2013; Rodrigo, Baker, & Rossi, 2013). *On-task conversation* was seen 18.6% of the time (388 instances).

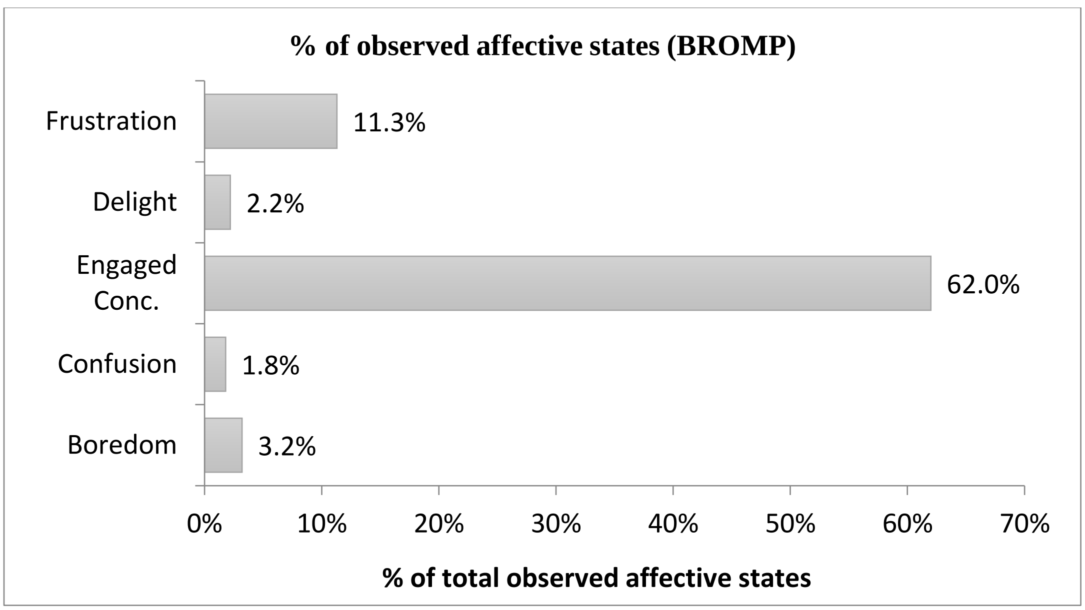

Figure 4. Graph showing the relative frequencies at which each affective state that was observed during data collection.

Machine learning

Data collection was followed by a multi-step process to develop interaction-based models of each affect. A two-class approach was used for each affective state, where that affective state was discriminated from all others. For example, engaged concentration was discriminated from all frustrated, bored, delighted, and confused instances combined (referred to as "all other") (see Figure 5).

Possible affective states	***Model of engaged concentration***
Engaged concentration	Engaged concentration
Boredom Confusion Frustration Delight	Not engaged concentration (All other)

Figure 5. An example of the 2-class approach for each affective state model: An example of the model of engaged concentration.

Because observations of some of the affective states were so infrequent during data collection (with confusion, boredom and delight making less than 5% of the total number of observations), there were large class imbalances in our data distribution. To correct for this, we made use of the *cloning* method to oversample our data. This was done by generating copies of cases associated with each of the smaller classes of affective states within the training data, to make the frequencies of each class more equally distributed for detector development. All test data, however, involved the original distributions.

Correlation-based filtering was used to remove features that had very low correlation with the predicted affect and behavior constructs (correlation coefficient < 0.04) from the initial feature set. This method involved the calculation of the Pearson's Correlation coefficient between generated features and the respective affective states observed. Features with correlation coefficients < 0.04 for all of the five affective states were then removed from the overall feature set. A total of 25 aggregated features were removed from the initial set of features, leaving 76 features that were ultimately used in the development of the affective state prediction models (see Appendix II).

Feature selection for each detector was then conducted using forward selection within the Rapidminer platform, where each feature is evaluated individually. In the forward selection process, the first feature that results in the best performing model is selected, and then all possible combinations of that selected feature and a subsequent feature are evaluated. In this manner, subsequent features are selected and feature selection stops when the required predefined number of features is selected, or when the model does not improve any further with the addition of another feature. Models for each construct are built in the RapidMiner 5.3 data-mining software, using common classification algorithms that have been previously shown to be successful in building affect models: JRip, J48 decision trees, KStar, Naïve-Bayes, step and logistic regression. Models are validated using 10-fold student-level batch cross-validation. In this cross-validation process, students in the training dataset are randomly divided into ten groups of approximately equal size. A detector is built using data from all possible combinations of 9 out of the overall 10 groups, and finally tested on the last group. Cross-validation at this level increases the confidence that the affect and behavior detectors will be more accurate for new students. To ensure comparability between the two sets of video-based and interaction-based detectors, the cross-validation process was carried out with the same randomly selected groups of students.

To handle missing data, several data imputation methods were tested with each machine learning algorithm to optimize model performance. This step was taken for all algorithms particularly since the step regression algorithm could not be conducted in the Rapidminer platform with missing data. We thus tested each algorithm with data that was imputed using zero, the average value, or with no imputation at all. With average imputation, missing values within the dataset would be replaced with the average value of all possible values for the given feature within the whole dataset, while zero imputation meant that missing values would be replaced with a ‘0’.

Finally, theperformance metric of AUC was computed on the original, non-resampled, datasets. In our measures of model performance, we made use of AUC as the primary measure of model goodness, as this metric is recommended to be particularly suitable for skewed data (Jeni, Cohn & de la Torre, 2013). The AUC metric was computed using the A' implementation that incorporates the Wilcoxon statistic (Hanley & McNeil, 1982) (rather than computing the integral of the area under the curve) to avoid having artificially high AUC estimates due to having multiple data points with the same goodness, a bug in the integration-based estimates currently available in most packages (Baker & Ocumpaugh, 2015). A model with AUC of 0.5 performs at chance, and a model with AUC-ROC of 1.0 performs perfectly. It is worth noting that AUC takes model confidence into consideration. From the forward selection process, a combination of features was also selected in each of the affect and behavior models that provide some insight into the type of student interactions that predict the particular affective state. The prediction models developed using computer log data are then compared against similar models built using video data. The section below gives a brief overview of how similar models are developed by another team working on the same project, using video data (Bosch et al., 2015). In this section, I specify several variables that were kept constant during machine learning to ensure a more equitable comparison.

Prediction models for student affective states using video data

Predictive models for the selection of student affective states were built by Bosch and colleagues (2015) using video facial data of student expressions during gameplay, and captured from web cameras affixed to the computers used during data collection.

Feature Engineering

Facial features were extracted using FACET, a commercialized version of the CERT computer vision software (Bosch et al, 2015). The Computer Expression Recognition Toolbox (CERT) (Littlewort et al., 2011) is a computer vision tool used to automatically detect action units as well as head pose and position information. The FACET tool provides likelihood estimates of the presence of 19 action units in total. Action Units (AUs) are labels for specific facial muscle activations (eg. lower brow, downturned lip) (Ekman & Friesen, 1978). These action units provide a small set of features that can be used to train AU detectors to identify various affective states, which can then be applied to new data to generate AU labels.

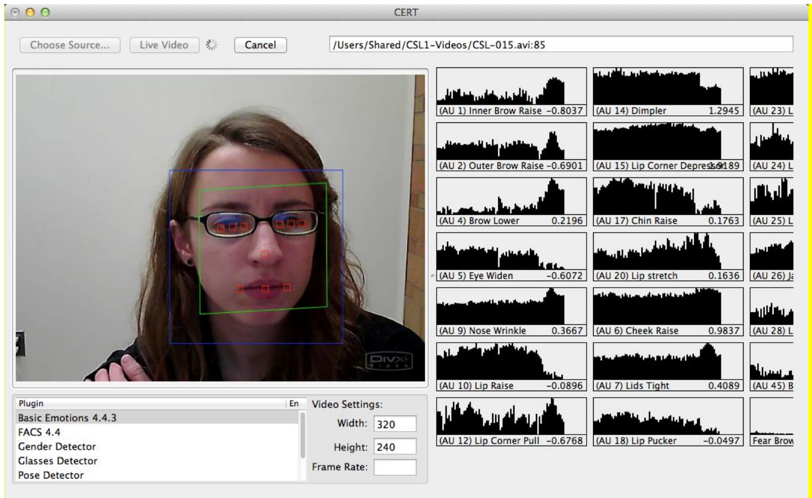

Figure 6. Screenshot of FACET program interface taken from an earlier study.

Of the initial 2087 instances available to train the video-based detectors on, about a quarter (25%) were discarded because FACET was not able to register the face and thus could not estimate the presence of AUs or compute the features. Poor lighting, extreme head pose or position, occlusions from hand-to-face gestures, and rapid movements can all cause face registration errors; these issues were not uncommon due to the game-like nature of the learning environment and the active behaviors of the young students in this study. 9% of the instances were also removed because the window of time leading up to the observation contained less than one second (13 frames) of data in which the face could be detected, culminating in 1224 instances where there was sufficient video data to train the affective state models.

Facial features were thus created by aggregating AUs, orientation, position and body movement estimates in different windows of time (3, 6, 9, 12 and 20-second windows) leading up to each BROMP observation of student affective state. Feature selection was then applied to isolate a smaller set of features for classification, and a set of the highest ranked features were then used in the prediction models for each student affective state, using RELIEF-F (Kononenko,

1994) on the training data. Ten iterations of feature selection were run on the training data with nested cross-validation and using data from a randomly selected percentage of students within the training set for each iteration.

Supervised Learning

Separate detectors were then built for each affective state using a two-class approach, where each given affective state was discriminated from all others (eg. boredom vs. all other) (see Figure 5). A variety of supervised classifiers were experimented with to build the prediction models using the Waikato Environment for Knowledge Analyses (WEKA), a machine learning tool. Due to the high level of class imbalances among the various affective states, downsampling and the synthetic minority oversampling techniques (SMOTE) were used to create more equal class sizes in the training data. Both downsampling and oversampling techniques work to create a balanced dataset using different methods. Downsampling involves the removal of random instances from the majority class, whereas oversampling techniques such as SMOTE creates synthetic training data by interpolating synthetic samples between an instance and randomly chosen nearest neighbors (Chawla, Bowyer, Hall, & Kegelmeyer, 2002).

After a more balanced dataset is obtained, the prediction models were then cross-validated at the student level, using a 10-fold student-level batch cross-validation, as was the case with the interaction-based detectors.

Model performance was measured using AUC values, which refers to ‘Area under the ROC curve’, and provides an aggregate measure of performance across all possible classification thresholds. Baseline results obtained from the supervised learning procedures using video data found that AUC performance was highest overall among models built using various classifiers on data within 12-second windows. However, for the sake of comparison, affective state prediction models were developed using 20-second windows in this study. Affective state predictions based on the 12-second window size was instead used in Study 2 due to better performances of all the affective state prediction models overall..

Results

Performance of affective state models using computer log data

As mentioned in the above section, a selection of machine learning algorithms, or classifiers, were conducted on the combined dataset containing affective state observations and aggregated features aligned in time, namely: JRip, J48 decision trees, Naïve-Bayes, step and logistic regression. The JRip and J48 decision tree algorithms were performed using the Waikato Environment for Knowledge Analyses (WEKA), a machine learning add-on tool in the Rapidminer platform. The results of prediction models built using these algorithms for each of the five affective states, are listed below.

Table 6

Interaction-based prediction models generated using a selection of classifiers, and their respective AUC performance values for the five affective states

Affective State	***Algorithm***	***Imputation Type***	***AUC value***
Boredom	Weka-JRip	Zero	0.509
	Weka-J48	None	0.544
	Step Regression	Zero	0.584
	Logistic Regression	Zero	0.629

	Naïve Bayes	Zero	0.550
Confusion	Weka-JRip	Zero	0.496
	Weka-J48	None	0.492
	Step Regression	Average	0.588
	Logistic Regression	Average	0.585
	Naïve Bayes	Average	0.527
Delight	Weka-JRip	Zero	0.512
	Weka-J48	None	0.569
	Step Regression	Zero	0.663
	Logistic Regression	None	0.679
	Naïve Bayes	Zero	0.606
Engaged concentration	Weka-JRip	Zero	0.505
	Weka-J48	None	0.542
	Step Regression	Zero	0.585
	Logistic Regression	Zero	0.578
	Naïve Bayes	Zero	0.586
Frustration	Weka-JRip	Zero	0.504
	Weka-J48	Zero	0.504
	Step Regression	Zero	0.545
	Logistic Regression	Average	0.559
	Naïve Bayes	None	0.532

Results from the prediction models constructed using the classifier algorithm with the highest performance showed that on average, the interaction-based models yielded an AUC of 0.634, which was higher than chance and comparable to other affective state models created in various intelligent tutoring systems such as ASSISTments, which have had AUC values ranging between from 0.63-0.74 (Pardos et al., 2014). Among the prediction models developed as shown in the table above, models built using different algorithms vary slightly in performance at predicting the various affective states. Specifically, the regression algorithms (step and logistic regression) appeared to perform better in predicting each of the five affective states in this particular learning context. This implies that the regression algorithms may provide a better fit for the computer log data aligned with these affective states, which resulted in slightly better prediction models.

Table 7 (below) lists the model performance based on the best-performing classifier used. Statistical significance may be computed for the AUC values of each affective state, to provide a sense of the performance of these prediction models as compared to chance (Fogarty, Baker, & Hudson, 2005). The probability of each AUC value as compared to chance (AUC = 0.5) may be computed using z-scores based on the formula below:

$$Z = \frac{A'_1 - A'_2.}{\sqrt{\mathrm{SE}(A'_1)^2 + \mathrm{SE}(A'_2)^2}}$$

Where A'_1 refers to the AUC value of the respective prediction mode, and A'_2 refers to the AUC value of chance (0.5), and $\mathrm{SE}(A'_1)$ and $\mathrm{SE}(A'_2)$ refers to the standard errors of the AUC value of each prediction model, and that of chance (0.0) respectively.

Table 7

AUC performance values for affective states using interaction-based detectors

Affective State Construct	Classifier	AUC	Imputation	Z-score	Sig.
Boredom	Logistic regression	0.629	Zero	3.442907	< 0.001
Confusion	Step regression	0.588	Average	1.794956	0.037
Delight	Logistic regression	0.679	None	4.011429	< 0.001
Engaged Concentration	Naïve Bayes	0.586	Zero	5.655054	< 0.001
Frustration	Logistic regression	0.559	Average	2.855977	0.002

Selected Features from interaction-based affective state models

From the forward selection process, a combination of features was selected in each of the affect and behavior detectors that provide some insight into the type of student interactions that predict the particular affective state or behavior. A list of these features are included in the table below (Table 8).

From the selected features for the boredom state, we can infer that a bored student is one who spends more time between actions on average. A bored student would also expend less effort to guide the ball object to move in the right direction, as indicated by fewer nudges made on the ball object to move it, and more ball objects being lost from the screen. On the other hand, the *confusion* state may be characterized by the aggregated features of a student who spends more time before her first nudge to make the ball object move, and drawing fewer objects in a playground level. A student who is confused may also not have known how to draw and move the ball object towards the balloon, thus spending a long time within a given playground level which leads to a lower number of levels attempted in total. From the features selected, *delight* appears to manifest when a student is able to achieve a silver trophy earlier on during gameplay, and completes more levels in total. We can also portray the student who experiences *delight* as someone who was able to achieve the objective without having to make multiple attempts to draw the relevant simple machines (such as springboards and pendulums). The features for *engaged concentration* would describe a student who is able to complete a level in fewer attempts, but erases the ball object more often during each attempt. These repeated draw-erase-draw actions imply that the student was putting in more effort to refine his/her strategies within a single attempt at the level. A student who is experiencing *engaged concentration* would also have had achieved success during gameplay (ie. A trophy or badge) in a shorter than average time.

Table 8

Features selected in the final interaction-based detectors of each affective state.

Affect/ Behavior	***Selected features***	Direction of relationship
Boredom	Time between actions within a level	Positive
	Total number of objects that were "lost" (i.e. Moved off the screen)	Positive
	Total number of nudges made on the ball object to move it	Negative
Confusion	Amount of time spent before the ball object was nudged to move	Positive
	Total number of levels attempted	Negative
	Total number of objects drawn within the level	Negative

Delight	Number of silver trophies achieved	Positive
	Consecutive number of pendulums and springboards created	Positive
	Total number of levels attempted	Negative
	Total number of levels completed successfully	Positive
Engaged Concentration	Total number of silver trophies achieved in under the average time	Positive
	Total number of level re-starts within a playground	Negative
	Total number of times a ball object was erased consecutively	Positive
Frustration	Total number of silver trophies achieved in under the average time	Negative
	Total number of level re-starts within a playground	Positive
	Total number of levels completed successfully	Negative
	Total number of levels attempted	Negative

Lastly, a student who experiences *frustration* is one who has failed to achieve the objective, or achieved fewer silver trophies within the average time taken. A student who is frustrated would also have had to make more attempts at a level due to repeated failure, thus resulting in fewer levels attempted in total.

Comparison with video-based affective state models

Video-based models for the same affective states were constructed (Bosch et al., 2015) as described earlier, for the same Physics Playground data. To facilitate comparison, both types of models were built using the same process of 10-fold student-level batch cross-validation. In this process, students in the training dataset are randomly divided into ten groups of approximately equal size. A detector is built using data from all possible combinations of 9 out of the overall 10 groups, and finally tested on the last group. Cross-validation at this level increases the confidence that the affect and behavior models will be more accurate for new students. To ensure comparability between the two sets of models, the cross-validation process was carried out with the same randomly selected groups of students.

Table 9 shows the performances of both the interaction-based and video-based models. On average, the video-based models had an average AUC of 0.695. This difference can be mainly attributed to the detection of delight, which was much more successful for the video-based models. Accuracy of the two detector suites was much more comparable for the other constructs, though the video-based models showed some advantages for engaged concentration and frustration, and were higher for 5 of the 6 constructs.

To understand how these AUC performances compare to those of the video-based prediction models, I computed similar z-scores of the interaction-based prediction models as compared to the video-based prediction models, using the same formula as above:

$$Z = \frac{A'_1 - A'_2.}{\sqrt{\mathrm{SE}(A'_1)^2 + \mathrm{SE}(A'_2)^2}}$$

In this case, however, A'_2 would refer to the AUC value of the respective video-based prediction model, while $\mathrm{SE}(A'_2)$ is the standard error of the AUC value from the video-based prediction model.

Table 9

*Comparing the AUC performance values for affective states using interaction-based and video-based detectors. * denotes z-score significance at $p \leq 0.05$*

	Interaction-Based Models	***Video-Based Models***	

Affect/ Behavior Construct	**Classifier**	**AUC**	**No. Instance**	**Classifier**	**AUC**	**No. Instance**	***Z-score***
Boredom	Logistic regression	0.629	1732	Classification via Clustering	0.617	1305	0.014
Confusion	Step regression	0.588	1732	Bayes Net	0.622	1293	-0.040
Delight	Logistic regression	0.679	1732	Updateable Naïve Bayes	0.860	1003	-0.165
Engaged Concentration	Naïve Bayes	0.586	1732	Bayes Net	0.658	1228	-0.082
Frustration	Logistic regression	0.559	1732	Bayes Net	0.632	1132	-0.087

The majority of the video-based models performed the best when using the Bayes Net classifier, except for *boredom*, *delight* and *off-task behavior*. In comparison, logistic and step regression classifiers produced the best performance for most of the interaction-based models, with the exception of *engaged concentration*. From the table above, the relative performances of each type of prediction model does not appear to be statistically significant. As such we can conclude that both interaction-based and video-based models perform comparably at predicting the five affective states, and are significantly more accurate than chance at predicting each affective state.

Discussion

Interaction-based vs. video-based models for student affective states

As seen in the results above, the slightly better performance of video models could be influenced by the uncontrolled whole-classroom setting in which video data is collected, where there are higher chances of video data being absent or compromised due to unpredictable student movement. While there were initially 2,087 instances of affect and behavior observed and coded, a moderate proportion of facial data instances were dropped from the final dataset when building the models. For interaction-based models, the exploratory and open-ended user-interface (Shute et al., 2013) constitutes a unique challenge in creating accurate models for student affect and behavior. The open-ended interface included multiple goals and several possible solutions that students could come up with to successfully complete each level. During gameplay, there are also multiple factors that could contribute to a student's failure to complete a level, that is not limited to just a lack of conceptual knowledge. Another issue was that there are fewer indicators of success per unit of time, as compared to other learning software that has been studied previously, such as the Cognitive Tutors (Baker et al., 2012). During gameplay, the system is able to recognize when combinations of objects the student draws forms an eligible agent. However, this indicator of success or failure is not apparent to the student until after he or she creates the ball object and applies a relevant force to trigger a simulation. Since students often spend at least several minutes building agents and ball objects, this results in coarser-grained indicators and evaluations of success and failure. The combination of open-endedness and lack of success indicators per unit of time consequently leads to greater difficulty translating the semantics of student-software interactions into accurate affective state predictions.

When comparing between the two sets of models, models that make direct use of physical traits such as students' facial features and bodily movements as captured by webcams, constitute embodied representations of students' affective states. On the other hand, interaction-based models were built based on student actions within the software, which serves as an indirect proxy of the students' actual affective states. These models rely, therefore on the degree to which student interactions with the software are influenced (or not) by the affective states they experience. Perhaps not surprisingly, video-based models perform somewhat better in predicting some affective states (e.g., delight, engaged concentration, and frustration). Although the video models are limited by missing data, interaction-based models can only detect something that causes students to change their behaviors within the software, which can be challenging given the issues arising from the open-ended game platform. Simply put, face-based affect models appear to provide more accurate affect estimates but in fewer situations, while interaction-based affect models provide less accurate estimates, but are applicable in more situations.

Since the performance of these models using video data was found to be slightly better than that of models using interaction data, this dissertation project will primarily apply prediction models generated from video data in the following analyses of student affective state sequences that indicate self-regulated learning behavior. However, because affective state models generated from interaction data tend to be more generalizable to other learning environments, there is potentially greater future applicability of interaction data models. Thus, affective state patterns will also be generated from these interaction-based prediction models of affective states. Comparisons in the patterns generated from both prediction models will then be discussed.

CHAPTER V.

STUDY 2: STUDENT AFFECTIVE STATE PATTERNS THAT ACCOMPANY SELF-REGULATED LEARNING BEHAVIOR

Components of Self-Regulated Learning

While there have been a variety of definitions and models for self-regulated learning in the field of education, it is commonly a construct made up of 3 or 4 main components: the task definition and planning or goal setting components before the task begins (such as in the framework proposed by Winne & Hadwin, 1999); the task performance component itself; and lastly, the post-task reflection component. In this dissertation study, I focus on the social-cognitive model of self-regulated learning (Zimmerman, 2000) that includes 3 main areas where self-regulated learning may be manifested in various ways (see Figure 7 below).

The unique game-based nature of Physics Playground provides a much less structured learning environment than many other educational platforms and intelligent tutoring systems. In other words, there are few structured elements in place that explicitly encourage the display of self-regulated learning behavior both before and after the student's attempt of the learning task. During task performance, however, self-regulated learning may be exhibited in the form of self-monitoring behavior; referred to as self-observation in the socio-cognitive framework (Barry J. Zimmerman, 1998), as the student is aware of how she is performing the task and keeps track of how well she is doing in the task. The immediate outcomes of strategic planning may also be observable during the task performance phase, as the student adjusts her actions based on a specific strategy (Zimmerman & Martinez-Pons, 1988). In particular, actions carried out during repeat attempts on the same playground level within the game environment could provide clearer observable information on a student's use of a consistent strategy as opposed to random trial and error.

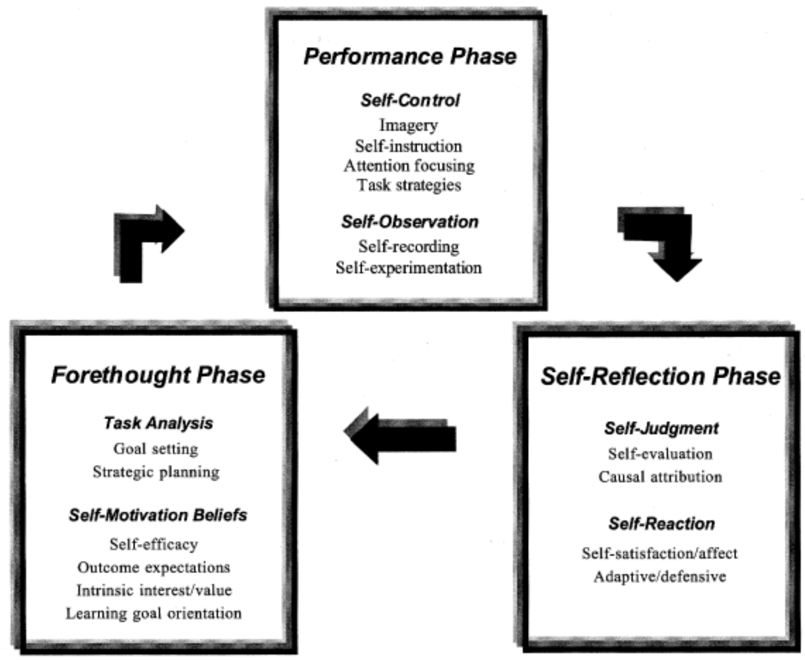

Figure 7. The socio-cognitive framework of self-regulated learning behavior (Zimmerman & Campillo, 2003)

Generating feature indicators of SRL behavior in Physics Playground

Based on the specific design of Physics Playground, I identified several sequences of student actions that could constitute self-regulated learning behavior within the context of Physics Playground. Since this environment is in the format of an open-ended game, users can explore any playground level at any time with no prescribed order, and there are no structured sections where students are explicitly encouraged to engage in pre-task/post-task activities. As such, there are limited ways in which the forethought and self-reflection components of self-regulated learning may be observed and measured in terms of student behavior. The performance component of self-regulated learning could however be measured during the student user's attempt at a selected playground level, when she works to create objects and structures to achieve the level objective. During this time, the student's goal is to create a simple machine to guide a ball towards a red balloon placed elsewhere on the screen. Students achieve a silver trophy if they manage to successfully guide the ball towards the balloon via the creation of a simple machine, whereas a gold trophy is earned only if the student player achieves the same objective using a minimal number of objects overall. As such, gold trophies are rewarded very rarely in the game environment.

For this dissertation study, I propose that we can identify self-monitoring behavior based on how the student user monitors the objects she creates to move the ball towards the balloon. One example of such behavior would be the drawing, erasing, and redrawing of objects or simple machines to improve the shape or position of the objects to achieve the level objective.

An initial analysis of student behavior within each level attempt also shows that only about half of the level attempts – out of nearly 61,000 level attempts in total – include the creation of a simple machine. There could be two possible reasons for this: 1) that the student user tried to draw a simple machine during the level attempt, but was not successful, or 2) that the student did not consciously attempt the strategy of creating a simple machine to achieve the level objective. Therefore, I propose that the creation of a simple machine on consecutive level attempts demonstrates that the student is actively trying to achieve the level objective, and is hence a behavioral indicator of goal setting and strategic planning. Specifically, students who draw simple machines across consecutive level attempts demonstrate the pre-performance planning component of self-regulated learning behavior.

To identify strategic planning self-regulated learning behavior before or in between playground level attempts, I isolated instances where a student attempted the same playground level multiple times back to back (subsequent-level dataset). Based on this subset of data, I created behavior features that identify when the student drew simple machines across consecutive level attempts (irrespective of whether the same machines or a different machines were drawn). The fact that the student is repeating the creation of a simple machine within consecutive level attempts implies that these actions are not due to random trial and error, but a result of a strategic decision made before each of the level attempts. Such behavior may be considered an indicator of strategic planning or goal setting. Another indicator of self-regulated learning may also be the students' repeat of a playground level even after achieving a badge, which indicates some form of strategic planning as they try to achieve the gold trophy.

Following these lines of thought, I analyzed student affective state patterns by splitting them into 4 separate subsets of data: The first dataset (self-observation) contains student affective states that co-occur with self-monitoring behavioral features within a single level attempt of a given playground. The 2nd dataset (no self-observation) contains students affective states co-occurring with student actions that do not contain self-monitoring features. The 3rd and 4th data subsets, on the other hand, make up a subset of student actions and affective states based on actions in consecutive level attempts of the same playground and level. Specifically, the 3rd data subset (strategic planning) consists of students' affective states when self-regulated learning behavioral features are observed in subsequent level attempts, while the last data subset (no strategic planning) consists of the remaining student affective states that accompany student actions when subsequent levels do not contain self-regulated learning features.

The full list of behavioral features generated are listed in the table below (Table 10).

Table 10

List of 8 behavioral features generated that indicate self-observation behavior among students attempting Physics Playground, within a single level (within-level), and strategic planning behavior within a subsequent attempt on a given playground level (subsequent-level).

Behavioral Feature	Level attempt	Description
Draw – erase – draw	Within	Student draws and erases the objects in the platform at least twice, exhibiting self-observation as they draw the freeform object, to make sure object is appropriate to what she had in mind
Draw – erase – draw(object/machine)	Within	Student draws and erases the objects in the platform at least once and creates either an object or a machine, exhibiting self-observation as they explore different strategies to get to the objective
Machine – erase - machine	Within	Student draws and erases a machine in the platform at least twice, exhibiting self-observation as they draw the machine, to make sure the machine is appropriate to what she had in mind.
Machine – erase – draw(object/machine)	Within	Student draws and erases the machine in the platform at least once and creates either an object or a machine, exhibiting self-observation as they explore different strategies to get to the objective
Draw – erase – draw – draw(object/machine)	Within	Student draws and erases the objects in the platform at least twice, exhibiting self-observation as they draw the freeform object, and explores different strategies
Student attempts level after badge	Subsequent-level	Student repeats an attempt on the same playground level after having achieved a badge within that playground level; an indication of strategic planning to achieve the gold trophy.
Any machine → to other machine	Subsequent-level	Student creates a machine within a level attempt, and repeats the same level while drawing a different machine; thus indicating an attempt at self-reflection and change in strategies

Any machine → repeated same machine	Subsequent-level	Student creates any machine within a level attempt, and repeats the same level while drawing the same machine; thus indicating an attempt at strategic planning

In sum, aggregated behavioral features were generated from the raw computer interaction logs, and the occurrence of these features were matched with corresponding affective state predictions within the same time frame. Video-based affective state predictions were made within 12-second time frames, whereas interaction-based affective state predictions were made within 20-second time frames.

Affective State Predictions

Corresponding Affective State Sequences from video-based prediction models

Using video data, prediction models of various student affective states had been generated at 12-second intervals in Bosch et al's, (2015) study (see Study 1). These particular prediction models were selected because of their better performance overall at predicting each of the different affective states in the Physics Playground environment. Because each affective state model was generated independent of the other affective states (affective state vs. all other), there may be cases where multiple affective states were predicted for a particular affective state window. It is also possible to have no affective state prediction at all within a particular 12-second window. In cases where multiple affective states were predicted for a given 12-second window, the affective state with the highest probability of occurrence was selected. In very rare cases where multiple affective states were predicted with the same probability, the more common affective state was selected (e.g. engaged concentration instead of boredom). In cases where no affective state predictions were made, the entire student level attempt was removed from analyses in this study altogether. In other words, student level attempts that contained instances with no affective state predictions were removed entirely from the dataset and not used in the generation of affective state sequences. These affective state predictions were matched to student action events that occur within the same 12-second window, resulting in a dataset with a row for every student affective state prediction. As such, a possible affective state sequence could look something like in the following table (Table 11):

Table 11

An example of an affective state sequence that co-occurs with a sequence of behaviors within Physics Playground.

userID	**Affective State**	**Time of affect (in milliseconds)**	**Action**	**Time of action (in milliseconds)**
203	Engaged concentration	491639	Draw freeform	491639
			Draw pin	494941
			Draw freeform	497059
203	Frustration	503639	Erase	503639
			Draw freeform	505891
			Erase	511282
			Draw ball	513789
203	Engaged concentration	515639	Nudge	515639
			Nudge	516483
			Collision	516989

Given the greater generalizability of affective state prediction models built from computer interaction log data, however, affective state patterns were also generated based on prediction models using interaction data, which will be described in further detail below.

Corresponding affective state sequences from interaction-based prediction models

Using computer interaction log data, affective state predictions have also been generated at 20-second intervals in Kai et al's study (see Study 1). Like the video-based affective state prediction models, each of the affective state prediction models built using interaction data was generated independently of other affective states (ie. Affective state vs. all other). Consequently, it is possible to have multiple affective state predictions for the same time-frame, or no affective state prediction at all. As with the video-based affective state prediction models, the affective state prediction with the highest probability was selected in cases with multiple predictions, and the more common affective state was selected in cases where there were multiple predictions with the same probability of occurring (eg. Engaged concentration selected over confusion). In cases where there were no affective state prediction made in a particular time frame, the entire student level attempt in which this lack of prediction data occurred was removed from the dataset.

Sequential Pattern Mining of Affective States

The research objectives of this dissertation project are to identify interesting transitions between affective states that may be unique to students exhibiting self-regulated learning behavior in the context of an educational game environment. To achieve this objective, sequential pattern mining techniques were to identify patterns in student affective states occurring over time during gameplay within the Physics Playground environment.

Sequential pattern mining has been used in the detection of behavioral patterns that are important for learning (Perera, Kay, Koprinska, Yacef, & Zaiane, 2009), as well as differentially frequent behavioral patterns of different student groups (Bouchet et al., 2012; Kinnebrew et al., 2013; Martínez & Yannakakis, 2011; Sabourin, Shores, Mott, & Lester, 2013), through the differential sequence mining technique. Differential sequence mining combines frequency measures and techniques from sequential pattern mining, which generates the most frequent patterns across a set of sequences, with episode mining, which determines the most frequent patterns within a given sequence (Bouchet et al., 2012). Differential sequence mining techniques have been employed in the investigation of self-regulated learning behavior within computer-based learning environments (Bouchet et al., 2012; Kinnebrew et al., 2013; J. L. Sabourin et al., 2013), and have been used in conjunction with clustering methods to identify different student groups and quantify the differences in their behavior during learning (Martínez & Yannakakis, 2011). Some common measures used to detect differentially frequent behavioral patterns include confidence measures, as well as sequence support (s-support), and instance support (i-support) measures. The s-support metric refers to the percentage of sequences that the pattern occurs in, whereas the i-support metric computes the average number of times the pattern occurs per sequence.

Since the main goal of this dissertation project is to conduct an exploratory analysis on the types of affective state patterns that are unique to students exhibiting self-regulated learning behavior within the game-based Physics Playground context, sequential pattern mining techniques were used to identify these patterns. In particular, the generalized sequential pattern mining (GSP) algorithm is used in this study (Srikant & Agrawal, 1995) using the Rapidminer Studio platform. GSP, or Generalized Sequential Pattern mining, is an A Priori-based algorithm used for sequence mining that makes multiple passes over the dataset to identify sequences of a defined minimum level of support. While the A Priori algorithm outputs patterns that are unordered in time and is mostly used in association rule mining, the GSP algorithm takes into account the order of patterns and identifies these patterns in the form of sequences. With the GSP algorithm, the first pass counts the frequencies of all 1-transaction sequences and identifies the most frequent single items. From this set of items, a set of candidate 2-sequences are identified and their frequencies counted with another pass over the dataset. The most frequent 2-sequences are in turn used to identify candidate 3-sequences, and another pass is made over the dataset to compute the frequencies of these sequences. This process is repeated until no more frequent sequences are found. This cutoff is determined manually, and in the case of this dissertation study, was set at 0.1, or 10% frequency, in order to maximize the number of affective state sequences identified.

To better isolate affective state patterns that are unique to self-regulated learners in this dissertation study, I also conducted a pairwise t-test comparison of each identified affective state sequence by student, which allows us to ascertain how different each of the respective affective state patterns are, in terms of frequency of occurrence. This comparison is similar to the manner in which comparisons were made based on the i-support metric in the differential sequence mining algorithm in other studies (eg. Kinnebrew et al., 2013).

It is important to note that the observations of each affective state made during data collection was heavily skewed, with 62% of all observations made of engaged concentration, as compared to only 11% of observations made of frustration, 3.2% of all observations made of boredom, 2.2% of all observations made of delight and 1.8% of all observations made of confusion. Comparing the frequencies of affective state sequences co-occurring with self-regulated behaviors versus non self-regulated behaviors thus reduces the possibility that various affective state transitions occur more frequently because of the overall prevalence of these affective states occurring within the dataset.

Predictors of self-regulated learning behavior (SRL) versus persistence

The SRL behavior features generated in this dissertation study were based off aggregated computer event logs and student action and followed a simple rational modeling approach. Because the objective of the data collection in this educational game platform was not to evaluate student self-regulated learning behavior, other methods for identifying SRL were not implemented. Consequently, the behavior features we generated may not be specific for identifying SRL, and could also identify underlying processes other than SRL. Specifically, some of the self-observation behavioral features were centered around students repeating a given action, such as Draw->Erase->Draw. While I propose that this set of actions is the result of a student's attempts at self-observation, one could also argue that repeated actions could indicate student persistence without self-observation. Therefore, to further investigate the overlap between persistence and self-regulated learning behavior, I examine persistence as an alternative dependent variable, by creating separate identifiers for student persistence within the game platform.

Specifically, I identify instances where students showed persistence while attempting a specific playground level by identifying levels where students spent the longest continuous duration of time. In the full interaction log dataset, students attempted a total of 6,176 playground levels, spread over 36,121 level attempts. The median amount of time spent on a level was 1.4 minutes with an inter-quartile range of 2.8 minutes. Further analyses of the distributions of the amount of time spent on each playground level showed that students spent longer than 9 continuous minutes on 5% of playground levels, but that these levels accounted for about 20% of total level attempts (see Figure 8). In comparison, SRL behaviors had been identified in about 20% of the level attempts. Based on these considerations, I chose 9 minutes as the time threshold for defining student persistence.

To evaluate the relationship between persistence on levels and the within-level SRL behavior features, I examined the performance of the persistence feature at identifying level attempts with SRL behavior features by computing the precision and recall. Recall, or sensitivity, computes the true positive rate, or the proportion of actual positives that are correctly identified as such. Recall may thus be computed using the following formula:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{P} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} = 1 - \mathrm{FNR}$$

where TPR and FNR refer to the true positive rate and false negative rate respectively, TP refers to the number of true positives, FN refers to the number of false negatives, and P refers to the number of real positive cases in the data. Precision or positive predictive value (PPV) computes the proportion of positive results that are true positive results, and may be calculated using the formula where FDR is the false discovery rate:

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} = 1 - \mathrm{FDR}$$

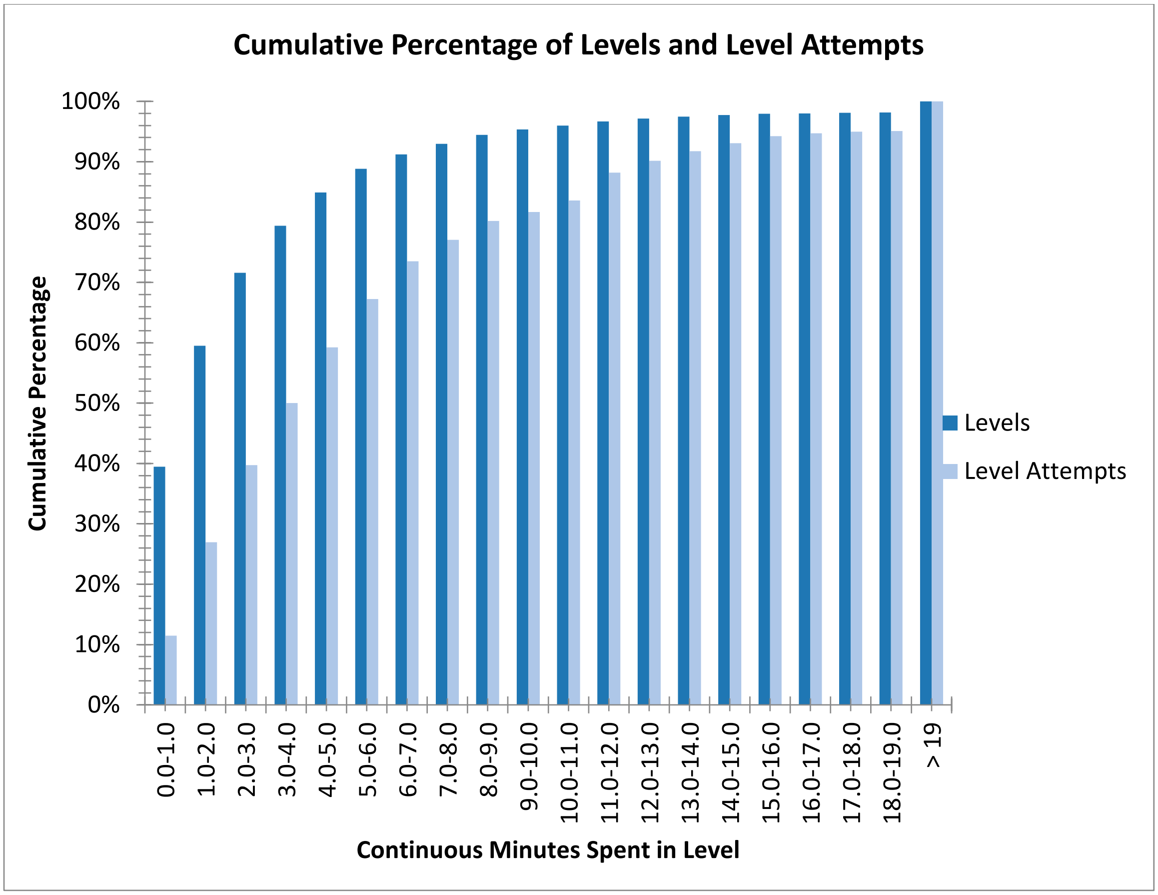

Figure 8. Cumulative distribution function showing the percentage of playground levels and corresponding level attempts made over a cumulative log of total time spent on each playground level, by student.

Results

Student self-regulated learning behavior and corresponding video-based affective states

A total of 8 behavioral features were generated that are indicative of self-regulated learning among students interacting with Physics Playground, listed above. 5 of these features were created from student action sequences within a single level attempt of the playground (within-level), whereas the 3 remaining features were created from student action sequences within a subsequent level attempt of the playground (strategic planning), relative to their first attempt at a given playground level. The table below shows the total number of instances of each behavioral feature that occurred over the entire duration of gameplay, as well as the total number of playground level attempts in which these features were present in.

Because the strategic planning features refer to student actions relative to a prior playground level attempt, the affective state patterns and other results were generated separately from features that contain student actions within any single playground level attempt.

Within the whole dataset of student level attempts, 45,894 affective states were predicted altogether in 12-second intervals, across 19,886 level attempts. Of this number, 7,010 affective state predictions co-occurred with playground level attempts that contained within-level student self-regulated learning behavioral features (self-observation), and 38,884 affective states that co-occurred with level attempts that do not contain self-regulated learning features (within-level-noSRL). In the within-level-SRL dataset, a total of 116 students out of 137 participants were found to exhibit at least one form of self-regulated learning behavior, across 708 playground level attempts. This number of level attempts in the dataset includes both first level attempts as well as subsequent level attempts. The number of instances for each feature created is shown in the table below (see Table 12).

Table 12

Behavioral features created across student action sequences to capture self-observation and strategic planning behavior during gameplay.

Behavioral feature	**SRL Behavior**	**Type**	**Total number of affective state predictions**	**Total number of level attempts**
Draw – erase – draw	Self-observation	Within-level	7933	490
Draw – erase – draw(object/machine)	Self-observation	Within-level	8207	513
Machine – erase - machine	Self-observation	Within-level	524	34
Machine – erase – draw(object/machine)	Self-observation	Within-level	2343	217
Draw – erase – draw – erase – draw(object/machine)	Self-observation	Within-level	938	40
Level repeat after badge	Strategic planning	Subsequent-level	2001	372
Any machine-repeat	Strategic planning	Subsequent-level	12804	2704
Any machine-other	Strategic planning	Subsequent-level	4186	798

From the complete dataset, 30,671 affective states were predicted across playground level attempts that constitute subsequent level attempts. These affective states were in turn split into 11,215 affective states that co-occurred with level attempts containing strategic planning behavioral features (strategic planning), and 19,456 affective states that co-occurred with level attempts that do not contain strategic planning behavioral features (no strategic planning). A total of 118 out of the total 137 student participants were found to exhibit self-regulated behavior in the form of the behavioral features across 3,215 subsequent-level attempts. This number includes only students who had made at least one subsequent-level attempt of the same playground level. On average, the number of subsequent-level attempts made on a given playground level is about 27, and ranges from 1 to 119.

Within a single playground level attempt, the number of affective state predictions generated varied according to the length of time a student spent within a single level attempt. Given the fact that an affective state prediction was made at 12-second intervals, the number of affective state predictions made per level attempt ranged from 1 through 178 in the whole dataset. Within the data subset of level attempts exhibiting self-regulated learning behavior, an average level attempt lasting for long enough for about 14 affective state predictions, whereas the average level attempt in the data subset containing no SRL behaviors lasts only for long enough for an average of 2.7 affective state predictions.

Similarly, subsequent level attempts tend to be shorter on average. Among subsequent level attempts that contain strategic planning behavioral features, the number of affective state predictions made per level attempt ranged from 1 through 96, with an average attempt lasting about as long as it takes to make less than 5 affective state predictions. In the data subset containing level attempts that do not contain strategic planning behavioral features, the number of affective state predictions made per level attempt could reach a maximum of 62, with an average attempt only lasting long enough for less than 2 affective state predictions to be made.

Student self-regulated learning behavior and interaction-based affective states

Because of the fact that the interaction-based affective state predictions were based on a different time window of 20 seconds, the number of affective state predictions as well as student level attempts are different from the above dataset based on video-based affective state predictions. The overall dataset contained more predictions of affective states across more level attempts in general, perhaps because there were no instances at all where no affect prediction was made. As a result, more student playground level attempts in total were retained in the dataset. Within the overall dataset, a total of 77,820 affective state predictions were made across 35,301 level attempts. Of this number, 10,818 affective states across 1,263 level attempts co-occurred with playground level attempts that contained self-observation student self-regulated learning behavioral features (within-level-SRL). This is in comparison to the 67,002 affective states across 34,038 level attempts that did not co-occur with any within-level student self-regulated learning behavioral features (not-within-level-SRL).

In terms of strategic planning behavioral features, there were a total of 29,425 subsequent-level attempts made in the game environment, that contained a total of 55,188 affective states predicted. 18,290 affective states across 5,438 level attempts were found to have co-occurred with instances of strategic planning behavioral features (strategic planning). On the other hand, 36,898 affective states across 23,987 level attempts did not co-occur with instances of strategic planning behavioral features (no strategic planning).

In a similar manner to the video-based affective state predictions, the average length a student spends within a single playground level attempt is longer on average among students who exhibit self-observation behavior (self-observation), in contrast to students who do not show any self-observation behavior (no self-observation). Subsequent level attempts among students who exhibit strategic planning behavior on average last longer than students who do not exhibit any strategic planning behavior as well. Since the number of interaction-based affective state predictions present have changed, SRL behavioral data from more level attempts have been included in the dataset (see Table 13). However, the total number of interaction-based affective state predictions made in each sub-category of data was not necessarily higher than the total number of video-based predictions, as the time frame for each interaction-based prediction is longer at 20 seconds, as compared to a video-based prediction at 12 seconds.

Table 13

Instances of behavioral features that capture self-observation and strategic planning behavior during gameplay and co-occur with an interaction-based affective state prediction.

Behavioral feature	**SRL Behavior**	**Type**	**Total number of affective state predictions**	**Total number of level attempts**
Draw – erase – draw	Self-observation	Within-level	8686	892
Draw – erase – draw(object/machine)	Self-observation	Within-level	9027	939
Machine – erase - machine	Self-observation	Within-level	517	56
Machine – erase – draw(object/machine)	Self-observation	Within-level	2594	376
Draw – erase – draw – erase - draw(object/machine)	Self-observation	Within-level	1051	69
Level repeat after badge	Strategic planning	Subsequent-level	1874	602
Any machine-repeat	Strategic planning	Subsequent-level	15805	4636
Any machine-other	Strategic planning	Subsequent-level	4948	1345

Distribution of affective state predictions in video and interaction-based models

As mentioned above, the distributions of student affective states observed during data collection was highly imbalanced. Affective state predictions based on the video-based models were similarly skewed, where *engaged concentration* made up an average of 58% of affect predictions. The *frustration* state prediction occurred the next most frequently, with an average of 27% of affective state predictions generated from the video-based models. This is followed by *delight*, at 8% of affect predictions generated, *confusion* at 4%, and lastly, *boredom* at 2% of all affective state predictions generated in the dataset (see Figure 9). As shown in Figure 9, the distributions of affective state predictions across the four data subsets also did not differ much, with a few exceptions. Engaged concentration predictions appear to occur slightly more frequently among the students exhibiting SRL behavior (within an individual level attempt), than among their counterparts who did not, at 62% and 54% respectively. On the other hand, fewer predictions of *frustration* were made for students with SRL behavior (within an individual level attempt) compared to those who did not show any SRL behavior within a playground level attempt, at 22% and 29% respectively.

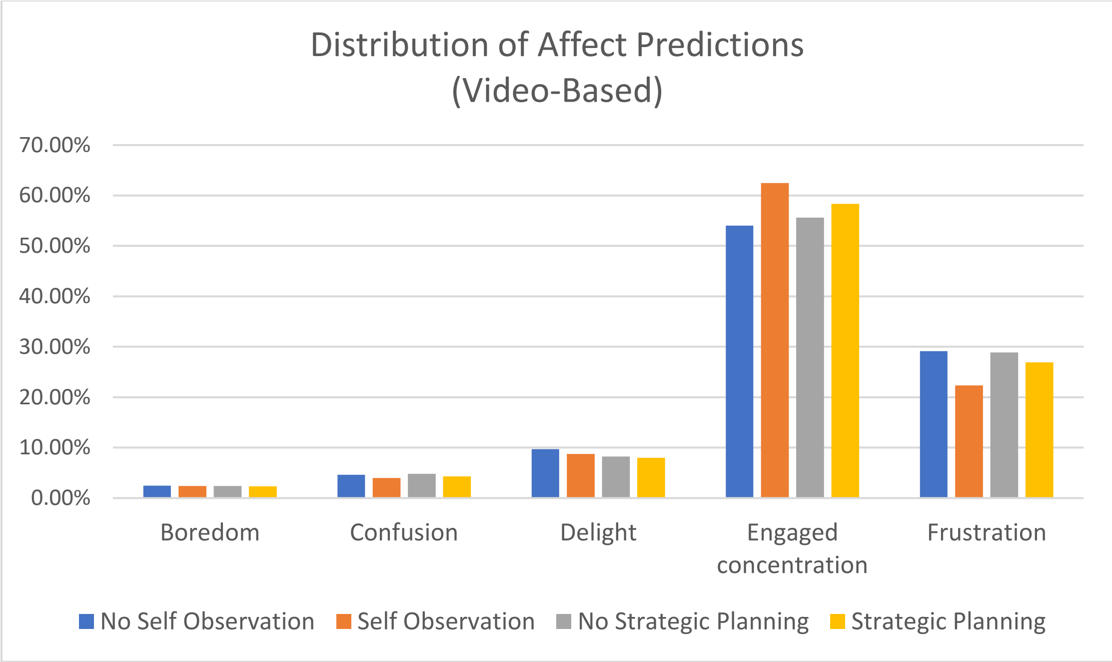

Figure 9. Clustered graph showing the relative distributions of video-based predictions of each affective state across the four data subsets.

In comparison, affective state predictions based on the interaction-based models differed in various ways. More affective state predictions of engaged concentration and boredom were generated, for instance, and fewer instances of frustration were predicted. On average, about 80% of affect predictions made were of engaged concentration (compared to 58% in the video-based models), and 8% of affect predictions made were of boredom (compared to 2% in the video-based models). On the other hand, much fewer predictions of frustration were made in the interaction-based models (5% as compared to 27% in the video-based models). Unlike the video-based model predictions, the distributions of predictions of each affective state appear to be quite similar across the four data subsets, however (see Figure 10).

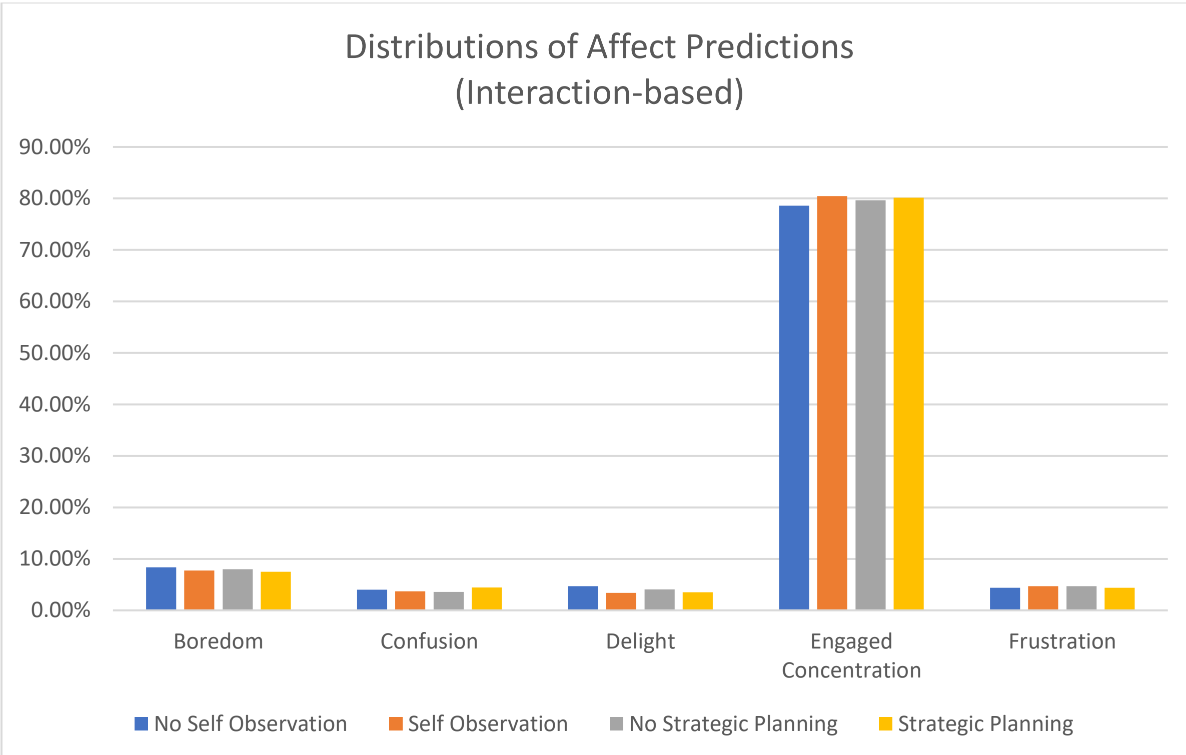

Figure 10. Clustered graph showing the relative distributions of interaction-based predictions of each affective state across the four data subsets.

Sequential Pattern Mining (SPM) using video-based affective state predictions

To achieve the research objectives, I made use of sequential pattern mining methods to identify affective state patterns across time during gameplay of the Physics Playground environments. This was conducted specifically using the Generalized Sequential Pattern algorithm (Srikant & Agrawal, 1996) in the Rapidminer Studio platform. The maximum gap set for this algorithm was 12 seconds, which is the length of an affective state prediction window. Because of the occurrences of any affective state patterns generated from the dataset, I set the support cutoff to 0.1, or 10%, to maximize the number of affective state patterns that were identified to accompany specific self-regulated learning behavior among students during gameplay. Within the sequential pattern mining paradigm, the choice of support cut-off is arbitrary and typically is selected in terms of producing a tractable number of patterns to analyze further – however, comparisons of support values between contexts indicate whether a pattern occurs more frequently in one context than another context. The GSP algorithm has several parameters that adjust its operation: window size, which determines the length of time within which a series of behaviors may be treated as a single behavior, as well as minimum and maximum gap sizes. Minimum and maximum gaps determine the amount of time in between which behaviors may occur and still be considered part of a sequence. Using the behavioral indicators of self-regulated learning that were created using interactive data obtained from Physics Playground, and the affect detector predictions, I identified student affective state patterns that co-occurred within the same level as these behavioral features. To more accurately quantify that the prevalent affective state patterns correlate with the presence of self=regulated learning behavior, I also ran the GSP algorithm with the same algorithm parameters – with the same window sizes and minimum/maximum gap sizes – through the playground level attempts that did not contain any self-regulated learning behavior features (within-level-noSRL). I then conducted paired t-tests on each of these patterns based on the number of students who exhibited a particular affective state pattern.

Self-regulated learning behavior within individual playground level attempts

Using the GSP algorithm, a list of 16 affective state patterns were generated that had supports of higher than 0.1; i.e. these patterns occurred for over 10% of the student level attempts made. From these patterns generated, we can see that the majority involved transitions between frustration and engaged concentration. To a lesser extent, affective transitions from delight and confusion to engaged concentration also occurred with support counts of 0.124 and

0.116 respectively. In particular, a persistence in engaged concentration throughout a level attempt seemed to occur with relatively high support counts, among students exhibiting self-observation behavior.

Table 14

*Affective state patterns observed with support > 0.1 for level attempts with self-observation behavior based on self-observation behavioral indicators, compared with supports for same patterns generated for level attempts with no self-observation behavior. The t-statistic represents paired t-test results between level attempts with and without SRL behavior, with significance based on Benjamini-Hochberg post-hoc corrections. * shows significance at p < 0.05.*

Affective state pattern	**Support – Affective state patterns with self observation behavior**	**Support – Affective state patterns with NO self observation behavior**	**T-statistic**
engaged concentration → engaged concentration	0.565	0.186	0.885
engaged concentration → engaged concentration → engaged concentration	0.434	0.084	2.001
frustration → engaged concentration	0.328	0.081	2.426*
engaged concentration → engaged concentration → engaged concentration → engaged concentration	0.323	0.044	2.830*
Frustration → frustration	0.251	0.083	5.107*
engaged concentration → engaged concentration → engaged concentration → engaged concentration → engaged concentration	0.240	0.025	3.268*

frustration → engaged concentration → engaged concentration	0.213	0.033	3.332*
engaged concentration → frustration	0.184	0.056	0.613
engaged concentration → engaged concentration → engaged concentration → engaged concentration → engaged concentration → engaged concentration	0.181	0.016	3.352*
frustration → engaged concentration → engaged concentration → engaged concentration	0.157	0.017	3.523*
engaged concentration → engaged concentration → engaged concentration → engaged concentration → engaged concentration → engaged concentration → engaged concentration	0.143	0.010	3.517*
frustration → engaged concentration → engaged concentration → engaged concentration → engaged concentration	0.133	0.010	4.148*
delight → engaged concentration	0.124	0.025	0.163
frustration → frustration → engaged concentration	0.121	0.020	2.528*
confusion → engaged concentration	0.116	0.020	2.576*
engaged concentration → engaged concentration → engaged concentration → engaged concentration → engaged concentration → engaged concentration → engaged concentration → engaged concentration	0.106	0.007	2.866*
delight → frustration	0.097	0.025	0.789*

frustration → engaged concentration → engaged concentration → engaged concentration → engaged concentration → engaged concentration	0.097	0.005	4.014*

In contrast, only one affective state pattern occurred with a support count of higher than 0.1 among students who did not exhibit self-observation behavior, which is the engaged concentration → engaged concentration transition (support = 0.186). The occurrence of other affective state transitions between frustration and engaged concentration, and even between delight/confusion and engaged concentration, appeared to occur at much lower frequencies than the frequencies among students who exhibited self-observation behavior.

From the support indices, several types of affective state transitions seem to occur more frequently among the students exhibiting self-observation as compared to students not exhibiting the self-observation self-regulated learning behavior. When we consider only affective state patterns that occur more frequently than 10% of the time, only one of the above 16 patterns emerge among students not exhibiting self-monitoring behavior, and is made up of a re-occurrence of engaged concentration across two affective state transactions.

The affective state transition between two different states with the highest support level among students exhibiting self-regulated learning behavior is the frustration → engaged concentration transition, which has a support of 0.328 as compared to the support of 0.081 among students not exhibiting self-regulated learning behavior.

It is important to note that the prevalence of engaged concentration and frustration affective state transitions could simply be due to the higher incidence of these affective states being observed during data collection. As mentioned in Study 1, the engaged concentration affective state is the most common affective state recorded during data collection, with over 60% of the instances observed, in contrast with 11.3% of the instances observed to be frustration, followed by less than 4% each of the boredom, delight and confusion observations. To examine whether the above affective state sequences were indeed associated with the presence of student self-regulated behavior, however, we conduct paired t-tests (assuming unequal variances) on each of the above affective state patterns, between the self-regulated learners and the non-self-regulated learners.

Paired t-tests & student-level affective state patterns

Paired t-tests were conducted on the affective state transitions that were found to occur at frequencies 10% or higher among students exhibiting self-regulated learning behavior within individual level attempts, versus students who did not exhibit self-regulated learning behavior. The results of these tests were adjusted to control for multiple comparisons using Benjamini-Hochberg (1995) corrections. The table above (Table 14) showed that the majority of these affective state patterns occurred significantly more often among self-regulated learners than among non-self-regulated learners.

The results of the paired t-tests showed that the majority of the affective state patterns generated through the GSP algorithm occurred at significantly higher frequencies among students exhibiting self-regulated learning, as compared to students who did not exhibit self-regulated learning behavior. The only affective state sequences that were found not to significantly differ in frequencies between the self-regulated and non-self-regulated learners were the following transitions:

- Engaged concentration → engaged concentration
- Engaged concentration → engaged concentration → engaged concentration
- Engaged concentration → frustration
- Delight → engaged concentration
- Delight → frustration

In particular, it is interesting to note that the frustration → engaged concentration affective state pattern occurred significantly more often among self-regulated learners ($p < 0.05$), and that this difference in frequency became even more significant when the learner persisted in engaged concentration afterwards (where frustration → engaged concentration → engaged concentration, etc.). In other words, frustration that ultimately led to engaged concentration, particularly sustained engaged concentration, appeared to be much more common among self-regulated learners as compared to their counterparts that did not exhibit self-regulated learning.

Despite the low occurrence of delight and confusion affective states in general, we can also see that affective transitions between delight → engaged concentration (support = 0.124), and confusion → engaged concentration (support = 0.116) still occurred with relatively high frequencies compared to the rest of the affective state transitions. This is despite the fact that these affective state patterns may not have been found to have occurred at significantly higher frequencies among self-regulated learners as compared to non-self-regulated learners. For instance, the delight → engaged concentration pattern was not found to be significantly different between the self-regulated and non-self-regulated learners, whereas the differences in frequencies of the confusion → engaged concentration pattern was significant (p-value < 0.05).

Another interesting finding from the affective state patterns generated was the persistence of the frustration affective state even among students who exhibited self-regulated learning behavior (frustration → frustration; support = 0.251, $p < 0.01$). However, the frustration → frustration → engaged concentration affective state pattern also occurs with relatively high frequency (support = 0.121, p-value < 0.05), which suggests that some extent of frustration may be prevalent among self-regulated learners, but it is not sustained over a long period of time and culminates in a more positive affective state like engaged concentration.

Overall, the relatively high occurrences of affective state patterns that transition from a negative state to engaged concentration, such as frustration → engaged concentration, confusion → engaged concentration, etc., imply that self-regulated learners tend to be able to keep their negative emotions brief and concentrate on completing the task at hand. While transitions from positive to negative affective states have also been found to occur with a relatively high level of support (eg. Delight → frustration, support = 0.097; engaged concentration → frustration, support = 0.184), the differences in frequencies of these patterns were not found to be significant between student level attempts exhibiting self-regulated behavior and level attempts that did not exhibit self-regulated learning behavior.

Self-regulated learning behaviors within subsequent-level attempts

Within the subsequent-level subset of data, considerably fewer affective state sequences were identified that occurred more than 10% of the time, for students who exhibited self-regulated learning behavior. The most prominent affective state patterns discovered were similar to several identified in the self-observation dataset, which include transitions between frustration and engaged concentration, as well as the persistence of engaged concentration across multiple affective state transitions. The affective state transition that occurred between two different affective states with the highest support count was the frustration to engaged concentration transition, with support = 0.161 among students who exhibited self-regulated learning behavior.

Table 15

*Affective state patterns observed with support > 0.1 for level attempts with strategic planning behavior based on strategic planning behavioral indicators, compared with supports for same patterns generated for level attempts with no strategic planning behavior. The t-statistic represents paired t-test results between level attempts with and without strategic planning behavior, with significance based on Benjamini-Hochberg post-hoc corrections. * shows significance at p < 0.05*

Affective state pattern	**Support – Affective state patterns with strategic planning behavior**	**Support – Affective state patterns with NO strategic planning behavior**	**T-statistic**
engaged concentration → engaged concentration	0.369	0.123	5.973*
engaged concentration → engaged concentration → engaged concentration	0.190	0.046	4.759*
frustration → engaged concentration	0.161	0.044	8.124*
frustration → frustration	0.149	0.046	4.187*
engaged concentration → frustration	0.115	0.033	4.508*
engaged concentration → engaged concentration → engaged	0.104	0.022	3.536*

concentration → engaged concentration	

On the flip side, there was only one affective state pattern that was generated with relatively high support among students who did not exhibit self-regulated learning behavior among this subsequent-level subset of data. The only affective state pattern that occurred more than 10% of the time is the engaged concentration → engaged concentration transition, with a support level of 0.123.

Paired t-tests & student-level affective state patterns

As with the self-observation datasets, paired t-tests by student showed the frequencies of the above affective state transitions to be significantly higher among students exhibiting self-regulated learning behaviors than students who did not. Based on results from the paired t-tests conducted for these affective state patterns, I found that all of the patterns occurred at significantly higher frequencies among students exhibiting self-regulated learning behavior, as compared to students not exhibiting self-regulated learning behavior over repeated level attempts.

Sequential Pattern Mining (SPM) using interaction-based affective state predictions

The same sequential pattern mining technique was applied to the affective state dataset based on affective state predictions using detectors from computer interaction log data. Affective state sequences with a support level of 0.1 (or 10%) or higher were generated for all of the subsets of data. Paired t-tests by student were then conducted to compare whether the frequencies of specific affective state sequences co-occurring with the self-observation and strategic planning behavioral features were significantly different from one another. The sections below discuss the affective state sequences generated that co-occurred with the self-observation and strategic planning behavioral features respectively.

Self-regulated learning within individual playground level attempts

Using the GSP algorithm, a total of 5 affective state patterns were generated that had supports of higher than 0.1; i.e. these patterns occurred for over 10% of the student level attempts made. Unlike the patterns above based on video-based affective state predictions, the majority of the affective state patterns here involved only engaged concentration, or a transition between engaged concentration and boredom.

As was the case with the video-based affective state predictions, paired t-tests were conducted on the affective state transitions that were found to occur at frequencies 10% or higher among students exhibiting self-regulated learning behavior within individual level attempts, versus students who did not exhibit self-regulated learning behavior. The results of these tests were adjusted to control for multiple comparisons using Benjamini-Hochberg (1995) corrections. Results of the paired t-tests in Table 15 showed that all of these affective state patterns occurred significantly more often among self-regulated learners than among non-self-regulated learners, even after Benjamini-Hochberg corrections.

Table 16

*Interaction-based affective state patterns observed with support > 0.1 for level attempts with self-observation behavior, compared with supports for same patterns generated for level attempts with no self-observation behavior. The t-statistic represents paired t-test results between level attempts with and without self-observation behavior, with significance based on Benjamini-Hochberg post-hoc corrections. * indicates significance at p < 0.05.*

Affective state pattern	**Support – Affective state patterns with self-observation SRL behavior**	**Support – Affective state patterns with NO self-observation SRL behavior**	**T-statistic**
engaged concentration → engaged concentration	0.578	0.220	14.146*
engaged concentration → engaged concentration → engaged concentration	0.264	0.050	11.156*
engaged concentration → boredom	0.132	0.025	7.267*
boredom → engaged concentration	0.131	0.020	7.258*
engaged concentration → engaged concentration → engaged concentration → engaged concentration	0.111	0.010	8.457*

Self-regulated learning within subsequent-level playground level attempts

Within the subsequent-level subset of data, only two affective state sequences were identified that occurred more than 10% of the time, for students who exhibited self-regulated learning behavior. These identified sequences were the same as the ones identified in the self-observation dataset, which involve the persistence of engaged concentration across multiple affective state transitions. There were no affective state sequences that were identified that occurred between two different affective states with a support level of 10% and above.

Table 17

*Interaction-based affective state patterns observed with support > 0.1 for level attempts with strategic planning behavior based on strategic planning behavioral indicators, compared with supports for same patterns generated for level attempts with no strategic planning behavior. The t-statistic represents paired t-test results between level attempts with and without strategic planning behavior, with significance based on Benjamini-Hochberg post-hoc corrections. * shows significance at p < 0.05*

Affective state pattern	**Support – Affective state patterns with strategic planning behavior**	**Support – Affective state patterns with NO strategic planning behavior**	**T-statistic**
engaged concentration → engaged concentration	0.532	0.174	8.376*
engaged concentration → engaged concentration → engaged concentration	0.145	0.033	5.548*
engaged concentration → boredom	0.060	0.013	7.670*

Paired t-tests conducted on the identified affective state sequences also found that these patterns were significantly different between students who exhibited self-regulated learning behavior and those who did not, even after Benjamini-Hochberg corrections.

Predictors of self-regulated learning (SRL) versus persistence

To assess the relationship between persistence on levels and SRL behavior, I defined persistent levels as those in which students spent more than 9 continuous minutes on a given playground level across multiple attempts. This time cutoff derived from initial analyses on the total length of time students spent on each playground level, and includes the top quintile of the level-attempt data, which is similar to the number of level attempts that were flagged with SRL behavior.

To identify whether the predictor for student persistence also predicted student self-regulated learning, I computed the recall and precision values of the persistence variable against the self-regulated learning predictors, with the self-regulated learning (SRL) behavioral features as the ground truth label. Precision and recall values were both found to be low, at 22% for recall and 18% for precision. The recall value implies that the percentage of self-regulated learning (SRL) behavior that is also persistent behavior is only at 22.%, while the precision value implies that the percentage of persistent student level attempts that is also self-regulated learning (SRL) behavior is only at 18%. In other words, these results suggest that student persistence does not necessarily predict self-regulated learning behavior, and vice versa.

When applying the persistence predictor against the SRL predictors, an AUC value of 0.5 was obtained, which indicates that the persistence variable only performs at chance level in identifying student SRL behavior. Altogether, these values suggest that persistence on a given playground level does not predict student self-regulated learning behavior (see Table 18).

Table 18

Results of computations of precision, recall and AUC values of the student persistence identifier based on the aggregated SRL behavioral predictors as ground truth.

Precision	*Recall*	*AUC*
22%	18%	0.496

Discussion

Affective state patterns using video-based predictions

By examining the affective states predictions generated for student level attempts that contain self-regulated learning behavior, I find that engaged concentration is the most common state predicted in both video detectors and interaction detectors. Moreover, the most common sequence of affective states is sustained engaged concentration, two or more consecutive predictions of engaged concentration. The presence of sustained engaged concentration may not be surprising given the prevalence of engaged concentration affective state throughout the dataset. However, the frequency of sustained engaged concentration in level attempts with SRL behavior is significantly higher than in level attempts without SRL behavior across both self-observation and strategic planning behaviors. This suggests that SRL behaviors tend to require more concentration and focus, and that students who are exhibiting SRL behaviors are less likely to enter into other affective states.

While the frustration affective state is also featured quite strongly among the affective state patterns generated, it is worth noting that the more common patterns involve only frustration that persists over periods of no longer than two affective state windows, and tend to transition to the engaged concentration affective state. Furthermore, *confusion* → *engaged concentration* transitions were identified that occurred significantly more frequently among instances of self-regulated learning. The presence of these transitions hence provides evidence to suggest that self-regulated learners may be better at regulating their negative emotions and resolving them during a task, towards a more positive affective state such as engaged concentration.

Affective state patterns using interaction-based predictions

The differences in the distributions of affect predictions made between the video-based and interaction-based models may be attributed to the nature of the type of data used to generate these predictions. As mentioned in Study 1, video-based prediction models constitute a direct measurement of students' affective states during gameplay, whereas interaction-based prediction models are reliant on student actions as a result of these emotions, thus constituting an indirect proxy of student emotions. Various student emotions could hence be better predicted using video data as opposed to interaction data, such as delight and frustration. On the other hand, more subtle emotions that do not manifest in the form of facial expressions, such as boredom may be better predicted by computer interaction data than video data.

Also, fewer affective state sequences were generated from the interaction-based prediction models due to the differences in window size. Among the sequences generated by interaction-based models, the most common sequences involved engaged concentration across multiple predictions This finding is similar to the sequences identified using video-based affective state predictions and is likely due to the prevalence of engaged concentration state predicted throughout gameplay. One affective state sequence that was identified using interaction-based predictors but not present among video-based predictors was the *engaged concentration* → *boredom* and *boredom* → *engaged concentration* sequences. Several reasons could explain the differences in affect sequences identified: 1) the fact that fewer affective state predictions were present as a result of the longer time-windows over which each prediction took place, 2) the relative accuracies of the prediction model of each affective state, or 3) boredom predictions were more common in interaction predictors, boredom was predicted ~8% of the time by interaction predictors, but < 3% of the time by video predictors.

In summary, results of sequential pattern mining showed that there is a significant relationship between student emotional states and specific self-regulated learning behavior. The generalized sequential pattern (GSP) mining algorithm conducted on student affective state predictions found that sustained engaged concentration occurred significantly more frequently among level attempts with self-observation and strategic planning behavior, than level attempts without. This result appears to be consistent across both self-observation and strategic planning level attempt groups, as well as for both sets of video-based and interaction-based affective state predictions. However, deeper analyses are needed to further understand the relationship between self-regulated learning behavior and sustained engaged concentration. Such analyses are especially necessary, as the results of sequential pattern mining also found sustained engaged concentration patterns to occur, but significantly less frequently, among level attempts that do not contain self-regulated learning behavior. As such, it is important to delve into the specific types of behavior that students exhibit when experiencing sustained engaged concentration, to identify exactly what student behaviors or actions are associated with engaged concentration but not self-regulated learning as defined in this study.

While similar affective state patterns have been identified across level attempts that contain self-regulated learning behavior, across both sets of interaction-based and video-based predictions, there are also several key differences in the types of patterns identified between the two. For instance, more patterns involving a transition from *frustration* to *engaged concentration* were identified with video-based affective state predictions, whereas interaction-based predictions tended to turn up more transitions between *boredom* and *engaged concentration*. These identified transitions appear to coincide with the higher percentage of frustration and boredom state predictions generated by the video-based and interaction-based detector respectively, which implies that these transitions may occur more frequently because predictions of these affective states predictions were more common in the Physics Playground dataset. The differing distributions of the affective state predictions made across both interaction- and video-based models also suggest that the interaction-based prediction model may be better at identifying certain student affective states, such as boredom, whereas the video-based prediction model may function better at identifying other affective states such as frustration and delight.

CHAPTER VI.

GENERAL DISCUSSION

This dissertation studies the development of detectors for predicting student affective states within an open-ended educational game environment and examines the relationship between student affective state patterns and self-regulated learning behavior. Specifically, I made use of affective state predictions to identify affective state patterns that co-occurred more frequently with student self-regulated learning behavior. This dissertation combines the development of affective state models with sequential pattern mining techniques across two studies to explore the relationship between student affective states over time and self-regulated learning behavior exhibited within an open-ended educational game platform.

Study 1 involves the development of affective state models within Physics Playground, an open-ended educational game environment, using two different forms of data collected during gameplay. Affective state models were developed for engaged concentration, boredom, confusion, delight and frustration. The performances of these models for each of the affective states were then compared against each other and the advantages and limitations of each method were examined in the context of online learning.

Study 2 builds on the affective state models developed and generates predictions of each affective state for the entire dataset. These predictions are then mapped to level attempts that contain self-regulated learning features so that affective state patterns can be identified through the sequential pattern mining method.

Results from Study 1 show that we can build models that predict student affective states with both video and interaction data. In general, models that make direct use of physical traits such as students' facial features and bodily movements as captured by webcams, constitute embodied representations of students' affective states. On the other hand, interaction-based models built based on student actions within the software are an indirect proxy of the students' actual affective states. These indirect models rely on the degree to which student interactions with the software are influenced (or not) by the affective states they experience. This could potentially explain the differences in performance between the video and interaction-based models, where the direct video-based prediction models appeared to perform slightly better at predicting student affective states of *engaged concentration, confusion, frustration* and *delight*. On the other hand, interaction-based models performed slightly better at predicting student affective states of *boredom* which is more likely to manifest in student actions. However, the differences in model performance between video- and interaction-based detectors were found to be not statistically significant based on the AUC metric. One limitation of the video affect detectors is that fewer instances of video data were available for development, since collection of usable video data is harder to achieve in an uncontrolled classroom setting.

In sum, although the video models are limited by missing data, interaction-based models can only detect something that causes students to change their behaviors within the software, which can be challenging given the issues arising from the open-ended game platform. Simply put, video-based affective state models appear to provide more accurate affect estimates but in fewer situations, whereas interaction-based affect models provide less accurate estimates, but are more generalizable to other learning contexts and can be re-purposed to improve student-centered learning.

Results from Study 2 were based on affect predictions generated using the models developed in Study 1. Aggregated features from raw computer log data were generated to represent self-regulated learning behavior both within an individual level attempt, and across subsequent level attempts. Sequential pattern mining was applied to affective state predictions mapped to level attempts containing these aggregated features. Results of the sequential pattern mining method identified multiple affective state patterns present in data subsets where SRL behavior was present within an individual playground level attempt that involve two patterns 1) staying in *engaged concentration* and 2) a transition from a negative emotion to *engaged concentration*. The first type of pattern is not surprising given the prevalence of engaged concentration affective states observed and predicted throughout the dataset. However, the presence of the transitions *frustration* → *engaged concentration* and *confusion* → *engaged concentration* occurred significantly more often among self-regulated learners than non-self-regulated learners is interesting. These transitions suggest that self-regulated learners do not dwell on negative emotions, and have a higher tendency to transition from a negative emotion to a more positive one (*engaged concentration*).

Among the interaction-based affective state predictions made, fewer patterns were identified in general, but sustained engaged concentration was still the predominant pattern observed. The occurrence of boredom ←→ engaged concentration transitions within interaction-based but not video-based detectors could be due to the higher percentage of boredom predictions within interaction-based models. However, we still find that students staying in engaged concentration is significantly more common in the level attempts with SRL behavior than those without.

In this project, I identified affective state patterns that occurred at the same time as self-regulated learning behavior that was occurring at the same time. From the patterns identified, sustained engaged concentration appears to be strongly associated with the occurrence of self-regulated learning behavior. While this implies that engaged concentration may be associated with specific self-regulated learning strategies, does not suggest that all students showing engaged concentration are utilizing self-regulated learning strategies.

Implications

Theoretical Implications

The results from this dissertation contribute to the existing SRL literature by providing a novel methodological approach to assessing the relationship between student affective states and self-regulated learning behavior, by using fine-grained measurements in the form of raw computer logs and video data. We find that sustained engaged concentration is more associated with SRL behaviors and that students are more likely to transition from confusion or frustration to engaged concentration when employing SRL behaviors.

Through these techniques, we are able to assess student emotions in real-time using either native interaction logs or in conjunction with video capturing software. With this real-time assessment, we are able to better associate specific affective states with SRL behavior and understand how patterns of affective states are associated with SRL behavior. Furthermore, we are able to see that specific changes in affective state could be signals of SRL behavior, specifically transitions from confusion and frustration to engaged concentration. Such temporal analysis is not possible using aggregated self-report measures that lack the time specificity. Since each form of assessment of student emotions and self-regulated learning has a unique set of advantages and limitations, being able to make use of multiple measures provides researchers with opportunities for a more holistic evaluation of student emotions and understand how it correlates with self-regulated learning behavior.

Practical Implications

One of the advantages of using fine-grained measures during learning, particularly in online learning contexts and intelligent tutoring systems, is the automated nature in which these systems could identify students at risk of becoming disengaged during learning. With rapidly developing affordances in technology, it has become more and more common for students to participate in at least some form of online instruction throughout their academic careers. Whereas teachers and facilitators were traditionally the ones to identify disengaged students through behavioral cues, the use of online learning platforms and intelligent tutoring systems make this exercise increasingly difficult. It is therefore important for such learning systems to be able to identify students who may be experiencing sustained negative emotions, or not using the appropriate self-regulated learning strategies, to provide real-time interventions that address these learning issues. Being able to track a student's behavior and her emotions as she engages in the learning content, and identifying her use of self-regulated learning strategies through these inputs, would provide learning systems with greater opportunities to customize interventions to improve the student's learning experience.

In addition to interventions to encourage student self-regulated learning behavior, the temporal tracking of student affective states over time could also provide feedback for the learning system to provide the learning content in a method that engages the student and facilitates learning the most. The ability to monitor students' affective states over time, particularly those already participating in self-regulated learning strategies, allows learning systems to be able to identify the precise moments when students are having trouble with the learning system. As such, the learning system would be better able to adapt to the students' learning needs, to improve the students' learning experience.

Limitations and Future Research

This dissertation project conducts an exploratory analysis of the temporal relationship between student emotions and self-regulated learning behavior. However, because the affective state patterns were identified based on predictions made in parallel, within the same level attempt as self-regulated learning behavior, this relationship is only correlational, and does not allow us to make any conclusions with respect to the causal nature of this relationship. As such, it would be difficult to identify whether the use of self-regulated learning strategies affect student emotions, or if self-regulated learning behavior was facilitated by students' emotions, as posited in previous research (Mega et al., 2014). Further research is therefore needed to conduct analyses on the nature of the relationship between student affective states and the use of self-regulated learning strategies.

The Physics Playground learning environment used in this dissertation project is an unstructured, open-ended game environment that does not explicitly encourage the use of self-regulated learning strategies. Consequently, it may be more difficult to identify quantifiable behaviors that may constitute self-regulated learning behavior, which may reflect only certain components of self-regulated learning behavior but not others. Further research is thus needed to apply this methodological approach to more structured online learning environments, to explore whether similar student affective state patterns can also be identified within these platforms.

Similarly, because of the observational nature of this study, the behavioral features generated to represent self-observation and strategic planning behavior within this platform are a result of a simple rational modeling approach and have not been validated by other measures of self-regulated learning behavior. As such, the action/behavioral sequences identified may be a result of factors other than self-regulated learning strategy use, such as persistence. Replicating Study 2 in a more structured online learning environment in which self-regulated learning behavior has already been assessed with other instruments, or adding additional measurements of student self-regulated learning would help to address this issue and improve the validity of these features of self-regulated learning behavior.

In order to better understand the relationship between self-regulated learning behavior and sustained engaged concentration, further examination is required into the specific behaviors of students within the game platform who are showing engaged concentration. For instance, clustering analyses may be conducted to identify different types of student behavior that occur during sustained engaged concentration patterns, with further labeling of which behaviors could be considered self-regulated learning behavior. This analysis could then corroborated against other measures such as surveys or a code-and-detect approach for identifying different types of student behaviors. In this manner, we would be better able to understand and determine the different student behaviors that occur during engaged concentration and the role that self-regulated learning could play in each of these behaviors.

Moreover, this study has focused on the specific affective state patterns associated with self-regulated learning behaviors. These self-regulated learning behaviors were identified based on a specific set of features identified within student actions using a simple rational modeling approach. Consequently, these behavioral features could also potentially be explained by other underlying processes, such as student persistence, conscientiousness or academic motivation. More investigation is hence needed to differentiate between behavior that is unique to self-regulated learning and behavior that could potentially result from other processes, by validating these features against established measures of self-observation and strategic planning behavior. Identifiers for potential alternative constructs may also be created in further analyses to pinpoint specific actions that could be a result from these alternative processes, and better differentiate between these actions and those that are uniquely a result of self-observation and strategic planning.

One specific student behavior that was used to identify self-regulated learning behavior was repetition and the use of a consistent strategy. Considered in isolation, drawing, erasing, and re-drawing objects could be seen as an act of boredom with the objectives of the game or frustration with the game's user interface. However, sustained engaged concentration during this behavior would suggest that these actions were related to a student strategy. It may thus be interesting to examine sequences where student actions were similar, but students were showing boredom or frustration, to determine whether additional interaction-based features could differentiate these cases. Such analysis could help to isolate additional features that could better identify SRL behavior as this detector would be based on both student actions and affect. These measures could then be compared and validated against other measures of self-regulated learning behavior such as surveys and self-reports.

The current work suggests a strong relationship between student affective states and the occurrence of self-regulated learning behaviors. However, we have not examined to what degree these affective states could modify the effectiveness of these behaviors on learning outcomes, or quantified the impact of how affective states could increase or limit the use of these behaviors. Further work could examine whether specific affective state patterns such as sustained frustration decrease the likelihood of SRL. Also, a future study could look at whether students with frequent periods of sustained engaged concentration are more likely to show SRL, and whether this SRL has a differential impact on learning outcomes than SRL in students without sustained engaged concentration.

REFERENCES

Ai, H., Litman, D. J., Forbes-Riley, K., Rotaru, M., Tetreault, J., & Purandare, A. (2006). Using system and user performance features to improve emotion detection in spoken tutoring dialogs. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (Vol. 2, pp. 797–800).

Aleven, V., Mclaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education*, *16*(2), 101–128. http://doi.org/10.1.1.121.9138

Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2004). Toward Tutoring Help Seeking: Applying Cognitive Modeling to Meta-Cognitive Skills. *Proceedings of the Seventh International Conference on Intelligent Tutoring Systems*, *2004*, 227–239. http://doi.org/10.1007/978-3-540-30139-4_22

AlZoubi, O., Calvo, R. a., & Stevens, R. H. (2009). Classification of EEG for affect recognition: An adaptive approach. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5866 LNAI*, 52–61. http://doi.org/10.1007/978-3-642-10439-8_6

Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009). Emotion sensors go to school. In *Frontiers in Artificial Intelligence and Applications* (Vol. 200, pp. 17–24). http://doi.org/10.3233/978-1-60750-028-5-17

Artino, A. R. (2008). Promoting Academic Motivation and Self-Regulation: Practical Guidelines for Online Instructors. *TechTrends*, *52*(3), 37–45. http://doi.org/10.1007/s11528-008-0153-x

Azevedo, R. (2005). Using Hypermedia as a Metacognitive Tool for Enhancing Student Learning ? The Role of Self- Regulated Learning. *Educational Psychologist*, *40*(4), 199–209. http://doi.org/10.1207/s15326985ep4004

Baker, R.S.; Ocumpaugh, J. (2015). Interaction-Based Affect Detection in Educational Software. In R. A. Calvo, S. K. D'Mello, J. Gratch, & A. Kappas (Eds.), *Handbook of Affective Computing* (pp. 233–245). Oxford, UK: Oxford University Press.

Baker, R. D., Gowda, S., & Wixon, M. (2012). Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra. *Proceedings of the 5th International Conference on Educational Data Mining*, 126–133. Retrieved from http://eric.ed.gov/?id=ED537205

Baker, R. S. J. D., Koedinger, K. R., Corbett, A. T., Wagner, A. Z., Evenson, S. E., Roll, I., … Beck, J. E. (2007). Adapting to When Students Game an Intelligent Tutoring System To cite this version : Adapting to When Students Game an Intelligent Tutoring System. *International Conference on Intelligent Tutoring Systems.*

Baker, R. S. J. D., & Yacef, K. (2009). The State of Educational Data Mining in 2009 : A Review and Future Visions. *Journal of Educational Data Mining*, *1*(1), 3–16. http://doi.org/http://doi.ieeecomputersociety.org/10.1109/ASE.2003.1240314

Baker, R. S. J. d, D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human Computer Studies*. http://doi.org/10.1016/j.ijhcs.2009.12.003

Baker, R. S., Ocumpaugh, J., Gowda, S. M., Kamarainen, A. M., & Metcalf, S. J. (2014). Extending Log-Based Affect Detection to a Multi-User Virtual Environment for Science. In *22nd Conference on User Modeling, Adaptation and Personalization (UMAP 2014)* (pp. 290–300).

Biswas, G., Roscoe, R., Jeong, H., & Sulcer, B. (2009). Promoting self-regulated learning skills in agent-based learning environments. In *Proceedings of the 17th …* (pp. 67–74). Retrieved from http://www.teachableagents.com/papers/2009/Biswas_ICCE.pdf

Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: the promotion of self-regulation as a means of preventing school failure. *Development and Psychopathology*, *20*(3), 899–911. http://doi.org/10.1017/S0954579408000436

Boekaerts, M., & Corno, L. (2005). Self regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology*, *54*(2), 199–231. http://doi.org/10.1111/j.1464-0597.2005.00205.x

Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., … Zhao, W. (2015). Automatic Detection of Learning-Centered Affective States in the Wild. *Proceedings of the 20th International Conference on Intelligent User Interfaces - IUI '15*, 379–388. http://doi.org/10.1145/2678025.2701397

Bosch, N., Mello, S. D., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., … Zhao, W. (n.d.). Automatic Detection of Learning - Centered Affective States in the Wild. In *In Proceedings of the 2015 International Conference on Intelligent User Interfaces (IUI 2015). New York, NY, USA: ACM.*

Bouchet, F., Azevedo, R., Kinnebrew, J. S., & Biswas, G. (2012). Identifying Students' Characteristic Learning Behaviors in an Intelligent Tutoring System Fostering Self-Regulated Learning. *Proceedings of the 5th International Conference on Educational Data Mining (EDM 2012)*, 65–72. Retrieved from http://files.eric.ed.gov/fulltext/ED537188.pdf

Bryce, D., & Whitebread, D. (2012). The development of metacognitive skills: Evidence from observational analysis of young children's behavior during problem-solving. *Metacognition and Learning*, *7*(3), 197–217. http://doi.org/10.1007/s11409-012-9091-2

Calvo, R. A. ., & D'Mello, S. K. (2010). Affect Detection: An Interdisciplinary Review of Models, Methods, and their Application to Learning Environments. *IEEE Transactions on Affective Computing*, *1*(1), 18–37.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. http://doi.org/10.1613/jair.953

Cho, M.-H., Demei, S., & Laffey, J. (2010). Relationships between self-regulation and social experiences in asynchronous online learning environments. *Journal of Interactive Learning*

Research, *21*(3), 297–316. Retrieved from http://www.editlib.org.eres.library.manoa.hawaii.edu/p/29491.

Cleary, T. J., & Zimmerman, B. J. (2004). Self-Regulation Empowerment Program: A school-based program to enhance self-regulated and self-motivated cycles of student learning. *Psychology in the Schools*, *41*(5), 537–550. http://doi.org/10.1002/pits.10177

Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, *29*(3), 241–250. http://doi.org/10.1080/1358165042000283101

D'Mello, S.K.; Graesser, A. (2012). Emotions During Learning with AutoTutor. In *Adaptive Technologies for Training and Education* (pp. 169–187).

D'Mello, S., Jackson, T., Craig, S., Morgan, B., Chip, P., White, H., … Graesser, A. (2008). AutoTutor Detects and Responds to Learners Affective and Cognitive States. In *Proceedings of the Workshop on Emotional and Cognitive issues in ITS in conjunction with the 9th International Conference on Intelligent Tutoring Systems* (Vol. 11, pp. 31–43). http://doi.org/10.1109/TE.2005.856149

De La Paz, S., & Graham, S. (1997). Strategy instruction in planning: Effects on the writing performance and behavior of students with learning difficulties. *Exceptional Children*, *63*(2), 167–181. Retrieved from http://proxy.consortiumlibrary.org/docview/201201924?accountid=14473%5Cnhttp://sz3sa6ce8r.search.serialssolutions.com/?SS_Source=3&amp%5Cngenre=article&amp%5Cnsid=ProQ:&amp%5Cnatitle=Strategy+instruction+in+planning:+Effects+on+the+writing+performance+and+behavio

Dragon, T., Arroyo, I., Woolf, B. P., Burleson, W., El Kaliouby, R., & Eydgahi, H. (2008). Viewing student affect and learning through classroom observation and physical sensors. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5091 LNCS*, 29–39. http://doi.org/10.1007/978-3-540-69132-7-8

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. http://doi.org/10.1016/j.patrec.2005.10.010

Feng, M., Heffernan, N., & Koedinger, K. (2009). Addressing the assessment challenge with an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*, *19*(3), 243–266. http://doi.org/10.1007/s11257-009-9063-7

Fogarty, J., Baker, R. S., & Hudson, S. E. (2005). Case Studies in the use of ROC Curve Analysis for Sensor-Based Estimates in Human Computer Interaction. *ACM International Conference Proceeding Series of Graphics Interface*, 129–136.

Goetz, T., Pekrun, R., Hall, N., & Haag, L. (2006). Academic emotions from a social-cognitive perspective: antecedents and domain specificity of students' affect in the context of Latin instruction. *The British Journal of Educational Psychology*, *76*(Pt 2), 289–308. http://doi.org/10.1348/000709905X42860

Graham, S., & Harris, K. R. (1989). Improving learning disabled students' skills at composing essays: self-instructional strategy training. *Exceptional Children*, *56*(3), 201.

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, *48*(4), 63–85. http://doi.org/10.1007/BF02300500

Karabenick, S. A., Woolley, M. E., Friedel, J. M., Bridget, V., Blazevski, J., Bonney, C. R., … Kelly, K. L. (2007). Cognitive Processing of Self-Report Items in Educational Research : Do They Think What We Mean ?, *42*(3), 139–151.

Kinnebrew, J. S., Loretz, K. M., & Biswas, G. (2013). A Contextualized, Differential Sequence Mining Method to Derive Students' Learning Behavior Patterns. *JEDM - Journal of Educational Data Mining*, *5*(1), 190–219. Retrieved from http://www.educationaldatamining.org/JEDM/index.php/JEDM/article/view/34

Kononenko, I. (1994). Estimating attributes: Analysis and extensions of RELIEF. In F. Bergadano & L. De Raedt (Eds.), *Machine Learning: ECML-94* (Vol. 784, pp. 171–182). Springer, Berlin Heidelberg. http://doi.org/10.1007/3-540-57868-4

Kopp, C. B. (1982). Antecedents of self-regulation: A developmental perspective. *Developmental Psychology*, *18*(2), 199–214. http://doi.org/10.1037/0012-1649.18.2.199

Latham, G. P. (2004). The benefits motivational goal-setting. *The Academy of Management*, *18*(4), 126–129.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. . (2011). The Computer Expression Recognition Toolbox (CERT). In *International Conference on IEEE* (pp. 298–305). Retrieved from http://mplab.ucsd.edu/~marni/Projects/CERT.htm

Martínez, H. P. H., & Yannakakis, G. (2011). Mining multimodal sequential patterns: a case study on affect detection. *… International Conference on Multimodal …*, 3–10. http://doi.org/10.1145/2070481.2070485

McManus, T. F. (2000). Individualizing instruction in a web-based hypermedia learning environment: Nonlinearity, advance organizers, and self-regulated learners. *Journal of Interactive Learning Research*, *11*, 219–251.

Mega, C., Ronconi, L., & De Beni, R. (2014). What makes a good student? How emotions, self-regulated learning, and motivation contribute to academic Achievement. *Journal of Educational Psychology*, *106*(1), 121–131. http://doi.org/10.1037/a0033546

Miller, W., Baker, R., Labrum, M. J., Petsche, K., & Wagner, A. . (2014). Boredom Across Activities, and Across the Year, within Reasoning Mind. In *Workshop on Data Mining for Educational Assessment and Feedback (ASSESS 2014)*. http://doi.org/10.1097/00005110-200012000-00011

Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for Educational Data Mining models: A case study in affect detection. *British Journal of Educational Technology*, *45*(3), 487–501.

Ocumpaugh, J., Baker, R., Rodrigo, M., Salvi, A., van Velsen, M., Aghababyan, A., & Martin, T. (2015). HART: The human affect recording tool. In *Proceedings of the ACM Special Interest Group on the Design of Communication (SIGDOC)* (pp. 4–9). Retrieved from http://dl.acm.org/citation.cfm?id=2775480

Ocumpaugh, J., Baker, R. S. J., Gaudino, S., Labrum, M. J., & Dezendorf, T. (2013). Field Observations of Engagement in Reasoning Mind. In *Proceedings of the 16th International Conference on Artificial Intelligence and Education* (pp. 624–627).

Pantic, M., Pantic, M., Rothkrantz, L. J. M., & Rothkrantz, L. J. M. (2003). Toward an Affect-Sensitive Multimodal Human Computer Interaction. *Proceedings of the IEEE, 91*(9), 1370–1390.

Paquette, L., Baker, R. S. J. d., Sao Pedro, M., Gobert, J. D., Rossi, L., Nakama, A., & Kauffman-Rogoff, Z. (2014). Sensor-Free Affect Detection for a Simulation-Based Science Inquiry Learning Environment. In *Proceedings of the 12th International Conference on Intelligent Tutoring Systems, ITS 2014* (pp. 1–10). Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-07221-0_1

Pardos, Z. a., Baker, R. S. J. d., San Pedro, M. O. C. Z., Gowda, S. M., & Gowda, S. M. (2014). Affective states and state tests: Investigating how affect and engagement during the school year predict end of year learning outcomes. *Journal of Learning Analytics*, *1*(1), 107–128. Retrieved from http://www.columbia.edu/~rsb2162/3536-15645-1-ED-v3.pdf

Paulhus, D. L., & Vazire, S. (2007). The Self-Report Method. In R. W. Robins, R. C. Fraley, & R. F. Kreuger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). New York, NY: The Guilford Press.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. *Educational Psychologist*, *37*(2), 91–105. http://doi.org/10.1207/S15326985EP3702_4

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaiane, O. R. (2009). Clustering and Sequential Pattern Mining of Online Collaborative Learning Data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(6), 759–772. http://doi.org/10.1109/TKDE.2008.138

Picard, R. W. (1997). *Affective Computing*. Cambridge, MA: MIT Press.

Pintrich, P. (2000). An Achievement Goal Theory Perspective on Issues in Motivation Terminology, Theory, and Research. *Contemporary Educational Psychology*, *25*(1), 92–104. http://doi.org/10.1006/ceps.1999.1017

Pintrich, P. R. (2003). A Motivational Science Perspective on the Role of Student Motivation in Learning and Teaching Contexts. *Journal of Educational Psychology, 95*(4), 667–686. http://doi.org/10.1037/0022-0663.95.4.667

Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40. http://doi.org/10.1037/0022-0663.82.1.33

Pintrich, P. R., & Zusho, A. (2002). The Development of Academic Self-Regulation: The Role of Cognitive and Motivational Factors. *The Development of Achievement Motivation*, (2), 249–284.

Pintrich, P., Smith, D., Garcia, T., & McKeachie, W. (1993). Reliabiilty and Predictive Validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*(3), 801–813. http://doi.org/0803973233

Rodrigo, M., Baker, R., & Rossi, L. (2013). Student Off-Task Behavior in Computer-Based Learning in the Philippines: Comparison to Prior Research in the USA. *Teachers College Record*, *115*(10), 1–27. Retrieved from http://www.columbia.edu/~rsb2162/off-task behavior-TCR-v23-fd.pdf

Rodrigo, M. M. T., & Baker, R. S. J. d. (2009). Coarse-grained detection of student frustration in an introductory programming course. In *Proceedings of the fifth international Computing Education Research Workshop - ICER 2009*. http://doi.org/10.1145/1584322.1584332

Rodrigo, M. M. T., Tabanao, E. S., Baker, R. S., Jadud, M. C., Amarra, A. C. M., Dy, T., … Sugay, J. O. (2009). Affective and behavioral predictors of novice programmer achievement. *ACM SIGCSE Bulletin*, *41*(3), 156. http://doi.org/10.1145/1595496.1562929

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction*, *21*(2), 267–280. http://doi.org/10.1016/j.learninstruc.2010.07.004

Sabourin, J. L., Shores, L. R., Mott, B. W., & Lester, J. C. (2013). Understanding and predicting student self-regulated learning strategies in game-based learning environments. *International Journal of Artificial Intelligence in Education*, *23*(1–4), 94–114. http://doi.org/10.1007/s40593-013-0004-6

Sabourin, J., Mott, B., & Lester, J. (2011). Modeling Learner Affect with Theoreticallly Grounded Dynamic Bayesian Networks. In *Proceedings of the Fourth International Conference on Affective Computing and Intelligent Interaction* (pp. 286–295).

San Pedro, M. O. Z., Baker, R. S. J. d., Bowers, A. J., & Heffernan, N. T. (2013). Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. In *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 177–184). Retrieved from http://www.columbia.edu/~rsb2162/EDM2013_SBBH.pdf

Schmitz, B., & Wiese, B. S. (2006). New perspectives for the evaluation of training sessions in self-regulated learning: Time-series analyses of diary data. *Contemporary Educational Psychology*, *31*(1), 64–96. http://doi.org/10.1016/j.cedpsych.2005.02.002

Sebe, N., Cohen, I., Gevers, T., & Huang, T. S. (2005). Multimodal Approaches for Emotion Recognition: A Survey. In *Proceedings of SPIE – The International Society for Optical Engineering* (pp. 56–67). http://doi.org/10.1117/12.600746

Shute, V., & Ventura, M. (2013). *Measuring and Supporting Learning in Games Stealth Assessment*. Cambridge, MA: MIT Press.

Shute, V., Ventura, M., & Kim, Y. J. (2013). Assessment and Learning of Qualitative Physics in Newton ' s Playground Newton ' s Playground. *The Journal of Educational Research*, *29*, 579–582.

Srikant, R; Agrawal, R. (1995). Mining Generalized Association Rules. In *IEEE Computer Society* (p. 3). http://doi.org/10.1080/09500690320000097424

Whitebread, D., Coltman, P., Pasternak, D. P., Sangster, C., Grau, V., Bingham, S., … Demetriou, D. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacognition and Learning*, *4*(1), 63–85. http://doi.org/10.1007/s11409-008-9033-1

Williams, P. E., & Hellman, C. M. (2004). Differences in Self-Regulation for Online Learning Between First- and Secon... *Research in Higher Education*, *45*(1), 71–83. http://doi.org/10.1023/B:RIHE.0000010047.46814.78

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. *Metacognition in Educational Theory and Practice*.

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, *27*(4), 551–572. http://doi.org/10.1016/S0361-476X(02)00006-1

Winne, P. H., & Perry, N. E. (2000). Measuring Self-Regulated Learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (1st ed., pp. 531–566). http://doi.org/10.1016/B978-012109890-2/50045-7

Wolters, C. a. (1998). Self-regulated learning and college students' regulation of motivation. *Journal of Educational Psychology*, *90*(2), 224–235. http://doi.org/10.1037/0022-0663.90.2.224

Woolf, B., Dragon, T., Arroyo, I., Cooper, D., Burleson, W., & Muldner, K. (2009). Recognizing and responding to student affect. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5612 LNCS*(PART 3), 713–722. http://doi.org/10.1007/978-3-642-02580-8_78

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. In *Pattern Analysis and Machine Intelligence, IEEE Transactions* (pp. 39–58). http://doi.org/10.1145/1322192.1322216

Zimmerman, B. J. (1998). Academic studing and the development of personal skill: A self-regulatory perspective. *Educational Psychologist*, *33*(2), 73–86. http://doi.org/10.1207/s15326985ep3302&3_3

Zimmerman, B. J. (2008). Investigating Self-Regulation and Motivation: Historical Background, Methodological Developments, and Future Prospects. *American Educational Research Journal*, *45*(1), 166–183. http://doi.org/10.3102/0002831207312909

Zimmerman, B. J., & Kitsantas, A. (1999). Acquiring writing revision skill: Shifting from process to outcome self-regulatory goals. *Journal of Educational Psychology*, *91*(2), 241–250. http://doi.org/10.1037//0022-0663.91.2.241

Zimmerman, B. J., & Martinez-Pons, M. (1988). Construct Validation of a Strategy Model of Student Self-Regulated Learning. *Journal of Educational Psychology*, *80*(3), 284–290. http://doi.org/10.1037/0022-0663.80.3.284

APPENDIX I.

List of raw computer interaction log data recorded in Physics Playground during gameplay, and their descriptions.

Column #	Category	Column Name	Description
1	Identifiers	RowID	Unique ID for each row of data
2	Observations	UNIQUEID	ID associated with matched coder observation events
3	Observations	OBSTIME	Time associated with matched coder observation events
4	Observations	CODER	ID associated with coder
5	Observations	BEHAVIOR	Coder Observed Behavior
6	Observations	AFFECT	Coder Observed Affect
7	Identifiers	userId	ID associated with each student
8	Identifiers	time	Computer System Time
9	Identifiers	utc_time	Computer System Time converted to UTC
10	Identifiers	timestamp	Computer Timestamp (milliseconds)
11	Identifiers	event	Event occuring in Environment
12	Identifiers	pe_step	???
13	Level Start/Restart	level_path	Type of Level
14	Level Start/Restart	badge_string	???
15	Level Start/Restart	ball_id	ID for ball object
16	Summary	game_time	???
17	Summary	ball_position_x	X position of Ball
18	Summary	ball_position_y	Y position of Ball
19	Summary	ball_velocity_x	X velocity of Ball
20	Summary	ball_velocity_y	Y velocity of Ball
21	Summary	ball_touched_count	Number of times ball touched other objects in game environment
22	Summary	mouse_distance	Distance moved by mouse
23	Summary	mouse_draw_distance	Length of line drawn by mouse
24	Summary	fps	Frames per second
25	Single Object Events	object_id	ID for object
26	Single Object Events	start_step	???
27	Single Object Events	elapsed	Amount of time taken to create object
28	Single Object Events	mass	Mass of object

29	Single Object Events	mass_center_x	X position of object center of mass
30	Single Object Events	mass_center_y	Y position of object center of mass
31	Single Object Events	pins_count	Number of pins in game environment
32	Single Object Events	position_x	X position of object
33	Single Object Events	position_y	Y position of object
34	Single Object Events	width	Width of object
35	Single Object Events	height	Height of object
36	Single Object Events	length	Length of object
37	Single Object Events	draw_data_count	.
38	Single Object Events	draw_data_length	.
39	Single Object Events	type	Type of Object
40	Multiple Object Events	objA	Object A ID
41	Multiple Object Events	positionA_x	X position of Object A
42	Multiple Object Events	positionA_y	Y position of Object A
43	Multiple Object Events	rotationA	Rotation of Object A
44	Multiple Object Events	velocityA_x	X velocity of Object A
45	Multiple Object Events	velocityA_y	Y velocity of Object A
46	Multiple Object Events	rotational_velocityA	Rotational Velocity of Object A
47	Multiple Object Events	objB	Object B ID
48	Multiple Object Events	positionB_x	X position of Object B
49	Multiple Object Events	positionB_y	Y position of Object B
50	Multiple Object Events	rotationB	Rotation of Object B
51	Multiple Object Events	velocityB_x	X velocity of Object B
52	Multiple Object Events	velocityB_y	Y velocity of Object B
53	Multiple Object Events	rotational_velocityB	Rotational Velocity of Object B

54	Machine Events	name	Name of machine drawn
55	Machine Events	strength	Strength at which ball is propelled from machine
56	Machine Events	primary_id	Machine ID
57	Machine Events	badge	Type of badge achieved and machine type
58	Machine Events	item	Playground level
59	Machine Events	touching_movement_x	X position where ball touched machine
60	Machine Events	touching_movement_y	Y position where ball touched machine
61	Machine Events	direction	Direction of ball movement (left or right)
62	Machine Events	wind_up_rotation	Amount of rotation made by machine to propel ball
63	Machine Events	ball_movement_x	X direction of ball movement
64	Machine Events	ball_movement_y	Y direcion of ball movement
65	Machine Events	ball_apex	Maximum height that ball reached
66	Machine Events	apex_rot_velocity	Rotational velocity of ball at apex
67	Machine Events	pin_count	Number of pins planted to build machine
68	Machine Events	freefall_distance	Distance of ball freefall
69	Machine Events	ball_distance	Distance of ball movement

APPENDIX II.

List of aggregated features generated in Study 1 for the development of affective state models, and their descriptions.

Aggregated Features	**Description**
Total Levels per Student	Number of levels attempted
LevelREStartPerStdt	Total number of level restarts per student
TotalREStartsPerLevelPerStdt	Number of restarts attempted per level
TimeBetStartN1stReStartofLevel	Time between Start and 1st restart of level
Levels in Day	Number of levels attempted within a day
TotalLostsinLevel	Number of Ball objects being Lost within a level
TotalPausesinLevel	Number of pauses made in level
TotalNudgesinLevel	Total number of nudges made on ball object within a level
TotalConsecutiveNudges	Number of nudges made on ball object within a level in a row
TotalClicksinLevel	Total number of mouse clicks made on objects in level
TotalNumber of Tutorials Per Level	Total number of tutorials watched within a level
Total Losts+LevelRestarts InLevel	Total number of events where ball objects were lost followed by level restart
LevelStartTime	Time when student first starts a level
TimeBeforeNudge	Time between start and 1st nudge of ball object
TimeBetweenNudgeandLevelEnd	Time between first nudge event made by student and level end time
TotalMachinesInLevel	Total number of machines built in a level
TotalFreeformsDrawn	Total number of freeform objects drawn
TotalDrawObjectsInLevel	Total number of objects (pins + freeforms) drawn
ConsecutiveRamps	Number ramp machines created in a row
ConsecutiveSpringboard	Number springboard machines created in a row

ConsecutivePendulum	Number pendulum machines created in a row
TotalEraseObjectinLevel	Number of objects erased within a level
ConsecutiveClicks	Number of clicks made on objects within a level in a row
ConsecEraseObj	Number of objects erased in a row within a level
Time Betw Lost and Obj Drawn	Time between object lost and new object drawn
TimeBetwLevelStart+Pin	Time between level start and 1st pin placed in level
NoBadgeWon	No badge achieved in a given playground level
TotalNoBadgePerStudt	Total number of playground levels attempted that did not result in a badge being achieved
StudtActionEvents	Total number of student action events, that include draw, erase, mouse click and nudge events
TimeBetwActions	Average amount of time elapsed between student actions within a single playground level
TotalActionsInLevel	Total number of actions taken within a single playground level
NumCollisionsSinceLastAction	Number of collisions made between the ball object and other objects in the game space since the last student action taken
CumSilverBadgePerStdt	Cumulative number of silver badges won by a student
CumGoldBadgePerStdt	Cumulative number of silver gold badges won by a student
BallCollisions	Total number of ball collisions made between the ball object and other objects in the game space in a single playground level
FFCollisions	Total number of ball collisions made between freeform objects drawn and other objects in the game space in a single playground level
CumStackinginLevel	Cumulative number of stacking events in a single playground level
T-LvlStartRestart-NoRestartYet	Amount of time before 1st restart in level
T-LvlStartNudge-NoNudgeYet	Amount of time before 1st nudge in level
T-LostObjDrawn-NoLostYet	Amount of time before 1st lost object in level

T-LostObjDrawn-NoObjDrawnYet	Amount of time before 1st object drawn in level
TotalPinsDrawn	Total number of pins drawn in a level
T-LvlStartPin-NoPinYet	Amount of time before 1st pin drawn in level
T-Actions-NoStdtActionYet	Amount of time before 1st student action made in level
T-ActionsClip-NoStdtActionYet	Amount of time before 1st student action made in 20-sec clip
AvgGoldSoFar	Average number of gold badges achieved so far
AvgSilverSoFar	Average number of silver badges achieved so far
PostSilver	Number of levels attempts made after achieving a silver badge
PostSilverPlay	Amount of time spent in a playground level after achieving the silver badge so far in the current level
PostSilverPlayTotal	Total amount of time spent in the game environment after achieving the silver badge
PostSilverPlayTotalDivByLevels	Total amount of time spent in a single playground level after achieving the silver badge
TimeFirstSilverThisLevel	Time taken to achieve first silver badge within level
AvgTimeToFirstSilver	Average time taken to achieve first silver badge
PlayerFirstSilverRelative	Amount of time taken for a student to achieve her first silver badge relative to the total amount of time spent in a single playground level
SumPlayerFirstSilverRelativeSoFar	Total amount of time taken for a student to achieve her first silver badge in a single playground level
AvgPlayerFirstSilverRelativeSoFar	Average amount of time taken for a student to achieve her first silver badge in a single playground level
SilverInUnderAverageTimeSoFar	Number of silver badges achieved in under the average time
PctSilverInUnderAverageTime	Percent of silver badges achieved in under the average time