ABSTRACT


The goal of this dissertation is to use four scholarly products to demonstrate how multimodal, multichannel process data can augment our conceptual, theoretical, methodological, and practical understanding of cognitive, affective, metacognitive (CAM) self-regulated learning (SRL) processes with intelligent tutoring systems (ITSs). In this dissertation, I argue that detecting and understanding learners’ real-time CAM processes with process measures is necessary to design systems capable of individualizing scaffolding, support, prompting, and ultimately fostering effective SRL with ITSs. I provide specific, empirical evidence that underscores how we detect and understand the influence of CAM processes on learners’ ability to effectively self-regulate their learning with within-subjects experimental designs. I also provide examples of how we can move beyond these types of approaches to design CAM-sensitive ITSs.

The first manuscript (Azevedo, Mudrick, Taub, & Bradbury, in press) is a theoretical handbook chapter, which outlines the issues we face using contemporary theoretical frameworks of cognitive, affect and emotions, metacognition, and SRL to understand behavioral signatures of CAM SRL processes derived from multimodal, multichannel data. The second and third manuscripts are empirical refereed conference proceedings that provide evidence of the complex interplay between affective and cognitive (Mudrick, Rowe, Taub, Lester & Azevedo, 2017) and cognitive and metacognitive (Mudrick, Sawyer, Price, Lester, & Azevedo, 2018) SRL processes during learning with a multimedia-based ITS. The final manuscript (Mudrick, Taub, & Azevedo, in press) is a chapter that describes a new system that we have developed that will allow us to
investigate how to detect, understand, interpret, and respond to real-time CAM SRL processes during learning. Taken together, the empirical results presented in this dissertation and the issues raised in each chapter emphasize how researchers can use multimodal, multichannel process data and specific methodological approaches to identify CAM SRL processes augment underspecified theoretical frameworks and design adaptive systems that can detect and respond to these processes in real-time.

The broader implications and future directions based on this dissertation entail the need to further investigate several methodological and measurement issues regarding the use of these data to detect and understand CAM SRL processes during learning with ITSs. Conceptually, these issues relate to validating identified behavioral signatures by examining generalizability of these findings. Theoretically, the alignment of the levels of explanation between contemporary theories of CAM SRL processes and the grain-size of multimodal multichannel process data needs to be explored. Pragmatically, identifying valid indices of these signatures will overcome cost- and scalability limitations. Lastly, these signatures will contribute to the creation of generalizable domain models of CAM SRL processes that future systems can use to tailor adaptive, real-time scaffolding and support to individualize instruction and ensure successful learning outcomes with ITSs.

by

Nicholas Vincent Mudrick

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology

Raleigh, North Carolina

2018

APPROVED BY:

Dr. Roger Azevedo
Committee Chair

Dr. Jeffrey Greene

Dr. Jing Feng

Dr. James Lester
DEDICATION

To Leanne: You are my number one. You have given me everything, and I cannot express how much your support, patience, and devotion means to me. You make me a better person, and I will forever be grateful.
BIOGRAPHY

Nicholas Vincent Mudrick is a doctoral student and research assistant in the Human Factors and Applied Cognition Program at North Carolina State University. Mudrick received his undergraduate degree in Psychology with Minors in Behavioral Science and Sexual Diversity Studies from McGill University in 2013. During his undergraduate career, Mudrick worked for Dr. Roger Azevedo as an undergraduate research assistant in the Laboratory for the Study of Metacognition and Advanced Learning Technologies (SMART Lab). His experience working at the SMART Lab inspired him to follow Dr. Roger Azevedo to pursue his doctorate at North Carolina State University.

At North Carolina State University, Mudrick continued working at the SMART Lab as a graduate research assistant where he worked on several federally funded and internationally funded research projects dealing with the complexity of learners’ cognitive, affective, metacognitive, and motivational (CAMM) self-regulated learning (SRL) processes during with intelligent tutoring systems (ITSs). In 2015, Mudrick received his master’s degree in Psychology specializing in Human Factors and Applied Cognition under the direction of Azevedo.
ACKNOWLEDGMENTS

First, I need to thank my advisor, mentor, and friend, Dr. Roger Azevedo. Meeting Roger literally changed the course of my life. I am so grateful to have been given the opportunity to follow Roger to North Carolina State University. Roger’s support, encouragement, guidance, and feedback are literally unbelievable. He has moved mountains for me, and under his supervision and with his help, I have achieved things I never would have thought possible. I will never be able to repay Roger for everything he has given me, and I will always be extremely grateful for the time and effort he has put into me. I would also like to thank the members of my advisory committee, Drs. James Lester, Jeffery Greene, and Jing Feng. I very much appreciate their willingness to serve on my committee, their excellent questions, effort, and time.

I would like to thank my fellow lab mates and friends who have assisted and supported me throughout the years. I would like to specifically thank Dr. Michelle Taub who has been like a sister and has always forced me (sometimes with me kicking and screaming) to improve and better myself. For that, I will always be grateful. Thank you to Megan Price and Elizabeth Cloude for being amazing lab mates, friends, and for all of your support.

To my parents, thank you for providing me with everything and giving me the world. I credit my love of learning, determination, and success to the sacrifices they have made for me. My parents are the strongest people that I know and being able to have relied on them throughout my time in graduate school ensured my success. They have given me everything that I could possibly need and more. To my mom, thank you for your patience, support, and all the time you spent with me on the phone whenever I needed it. To my dad, thank you for the excellent and objective advice, you always know how to put it in perspective. I cannot express how lucky I feel to have been born into the most supportive family I know. I will always be so very grateful for
everything my parents have done for me and will always remember their support and love throughout this process.

    Most importantly, thank you to my wife, Leanne Jones. Leanne, you have given your all to make me who I am. I would not be here without your help, support, and love. You have given me everything that I need to succeed and even more. You are the inspiration for my determination. You push me, challenge me, and ultimately, complete me. You are my best friend and the love of my life. You truly are an exceptional person and the best one I know. I am so excited to start the next chapter of our lives together and face the next challenge hand-in-hand with you.
# TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................... vii

LIST OF FIGURES ......................................................................................................... viii

CHAPTER 1: INTRODUCTION .................................................................................... 1
  Conceptual and Theoretical Contributions ................................................................. 3
  Methodological and Experimental Approaches ......................................................... 5
  Goal of this Dissertation ............................................................................................... 7
  Summary ......................................................................................................................... 14

CHAPTER 2: SELF-REGULATION IN COMPUTER ASSISTED LEARNING SYSTEMS ... 15
  History of Research in SRL and CALS ........................................................................ 16
  Computer-Assisted Learning Systems (CALs): An Overview .................................... 17
  CALs and Self-Regulated Learning: Examples .......................................................... 21
  Conceptual and Theoretical Issues Derived from Models on Cognitive, Affective, and Metacognitive Self-Regulated Learning Processes ......................................... 23
  Winne and Hadwin’s Information Processing Theory of Self-Regulated Learning .... 24
  Greene & Aevedo’s Macro and Micro Level Framework of Self-Regulated Learning ................................................................. 26
  D’Mello & Graesser’s (2012) Dynamics of Affective States Model .......................... 28
  Conceptual and Theoretical Dichotomies Related to CAM SRL Processes Related to CALS ................................................................. 30
  Assessing Self-Regulation using Multimodal Multichannel Data ............................ 34
  CALS as a Research and Learning Tool ..................................................................... 35
  Types of Multimodal Multichannel SRL Process Data ............................................. 38
  Using Multimodal Multichannel Data to Measure CAM processes ..................... 42
  Selecting the Appropriate Multimodal Multichannel Data ....................................... 45
  Implications of Using Multimodal Multichannel Data ............................................ 48
  Future Directions and Conclusions .......................................................................... 50
  References ..................................................................................................................... 58

CHAPTER 3: IDENTIFYING HOW METACOGNITIVE JUDGMENTS INFLUENCE STUDENT PERFORMANCE DURING LEARNING WITH METATUTORIVH ................................................................. 66
  Introduction ................................................................................................................... 66
  Current Study ............................................................................................................... 67
  Results ............................................................................................................................ 72
  Discussion ...................................................................................................................... 74
  References ..................................................................................................................... 75

CHAPTER 4: TOWARD AFFECT SENSITIVE VIRTUAL HUMAN TUTORS: THE INFLUENCE OF FACIAL EXPRESSIONS ON LEARNING AND EMOTIONS ......................................................... 76
CHAPTER 5: METAMENTOR: AN INTERACTIVE SYSTEM THAT USES VISUALIZATIONS OF STUDENTS’ REAL-TIME COGNITIVE, AFFECTIVE, METACOGNITIVE, AND MOTIVATIONAL SELF-REGULATORY PROCESSES TO STUDY HUMAN TUTORS’ DECISION MAKING

Abstract ............................................................................................. 82
Introduction ........................................................................................ 82
Using Expert Human Tutors to Inform the Design of ITSs ................. 85
Research with MetaTutor: The Impact of CAMM SRL Processes on Student Learning ........................................................................... 88
Limitations: Fostering CAMM SRL Processes During Learning with MetaTutor .... 90
MetaMentor: An Interactive System Developed to Identify and Foster a Student’s CAMM SRL Processes in Real Time ............................................. 92
MetaMentor: Student Interface ................................................................ 94
MetaMentor: Human Tutor Interface....................................................... 98
Example Learning Session: Towards Developing an Adaptive MetaMentor ........ 104
Possible Challenges and Open Questions for Future Research with MetaMentor ...... 109
Conclusion and Future Directions ............................................................ 113
Acknowledgments ................................................................................ 115
References ........................................................................................... 117

CHAPTER 6: INTEGRATED DISCUSSION ............................................. 126
Broader Impacts and Future Directions ................................................. 131
Conclusion ........................................................................................... 139

REFERENCES .................................................................................... 141
LIST OF TABLES

Chapter 2

Table 1. *Sample Questions and Data Sources Based on Conceptual Dichotomies About Cognitive, Affective, and Metacognitive (CAM) Self-Regulated Learning Processes* .......................................................... 53

Chapter 3

Table 1. *Summary of multiple-choice score by EOL grouping, which summarizes research questions 1 (top row) and 2 (bottom row)* ......................................................... 73

Chapter 4

Table 1. *Human tutor agent facial expressions by study condition* ........................................ 78

Chapter 5

Table 1. *Hypothesized CAMM SRL Processes Depicted from Each Visualization* ........ 106
LIST OF FIGURES

Chapter 2

Figure 1. Screenshot of MetaTutor ............................................................................................................. 56

Figure 2. Examples of specific uses of multimodal multichannel data to investigate CAM processes with different CALSs ............................................................................................................. 57

Chapter 3

Figure 1. Screenshot of MetaTutorIVH’s main interface illustrating the science questions, multimedia content, and intelligent virtual human (IVH) .................................................. 68

Chapter 4

Figure 1. Example screenshot of the question, text, diagram, and human tutor agent with a congruent facial expression of joy ............................................................................................................... 78

Figure 2. Human tutor agent facial expressions of emotion: joy (left), confusion (middle), and neutral (right) .......................................................................................................................... 78

Chapter 5

Figure 1. Example screenshot of the student’s interface during learning with MetaMentor .......................................................... 96

Figure 2. Example screenshot of the tutor’s interface ................................................................................. 100

Figure 3. Example experimental setup with instrumented student (left) and expert human tutor (right) .......................................................................................................................... 105
CHAPTER 1: INTRODUCTION

During self-regulated learning (SRL), learners play an active role in their learning with complex topics and domains. Effective SRL requires learners to concurrently monitor and control their cognitive, affective, and metacognitive (CAM) processes (Schunk & Greene, 2018). Research has shown that learners who engage in effective SRL processes achieve high learning outcomes (Azevedo, Taub, & Mudrick, 2018). However, research indicates that learners do not always engage in effective SRL and are usually unable to self-regulate their learning efficiently. To address learners’ limitations, researchers have developed intelligent tutoring systems (ITSs) to foster these CAM SRL processes during learning (Aleven & Koedinger, 2016; Azevedo, Mudrick, Taub, & Bradbury, in press). ITSs are computer-based learning environments that emulate effective instructional strategies of expert human tutors (e.g., adaptive scaffolding, immediate feedback, fading support, prompting, etc.) (Aleven & Koedinger, 2016). Research shows that ITSs can effectively improve some (i.e., cognitive and metacognitive processes), but not all (i.e., motivational processes), SRL processes during learning (Azevedo, Mudrick, Taub, & Bradbury, in press; Taub, Mudrick, & Azevedo, in press; Graesser, 2016, Woolf, 2009).

Previous research investigating the relationships between learners’ CAM processes and their SRL has primarily used self-report measures to detect these processes during learning with ITSs (Azevedo et al., 2018, in press). Typically, researchers administer batteries of self-report questionnaires before, during, and after learning sessions to understand how learners’ CAM processes influence their SRL with these systems. Using self-reports to detect CAM processes has substantially increased our understanding of their impact on SRL with ITSs and should continue to be used to study these processes. Furthermore, this research has helped to develop ITSs capable of responding to some of learners’ CAM SRL processes (e.g., cognitive and
metacognitive processes, such as with MetaTutor (Azevedo et al., 2011, 2013, 2018, in press) during learning. However, using self-reports to detect and understand CAM processes during learning with ITSs may be problematic for several reasons. For example, self-reports rely on learners’ self-assessments of their CAM processes, which may not always be accurate and may be subject to possible bias (i.e., social desirability). Additionally, the administration of self-reports during learning may be disruptive and distracting for learners and may deleteriously impact their learning. Furthermore, most self-reports are designed to only measure specific components of CAM processes and do not allow for the investigation of interplay between these processes during SRL. Lastly, though they can be administered multiple times during learning and can be used to examine the influence of specific aspects of the learning session (e.g., the emotional impact of a prompt to engage in a specific cognitive strategy), self-reports are contextually and temporally constrained to the timing of their administration.

However, ITSs now have the potential to foster effective SRL with these systems by detecting, analyzing, and responding to learners’ ability to monitor and regulate their CAM processes during learning (Azevedo et al., 2011, 2018; Lester, Lobene, Mott, & Rowe, 2014). Specifically, these systems can now be instrumented with sensors and devices to collect multimodal, multichannel process data (e.g., log files, eye tracking, videos of facial expressions of emotion, electrodermal activity [EDA] sensors) that measure the complex, real-time, temporal unfolding of learners’ CAM processes and how they relate to their regulation during learning with these systems. For example, ITSs can be equipped with eye trackers to detect and measure learners’ cognitive and metacognitive processes (D’Mello, 2016; van Gog & Jarodzka, 2013). Videos of learners’ facial expressions can be automatically analyzed by commercially available (e.g., iMotions, Noldus FaceReader) and open-source (e.g., Microsoft Azure) software to
automatically detect and analyze learner emotions (e.g., confusion, frustration, boredom) 
(Harley, Lajoie, Frasson, & Hall, 2017; Lester, Mott, Rowe, & Sabourin, 2013). Furthermore, 
EDA can be analyzed to identify learners’ heart rate variability and skin conductance responses 
that may indicate cognitive and affective processes during learning with ITSs (Azevedo et al., 
2018). Lastly, log files capture learner-system interactions (e.g., mouse clicks, time spent on a 
page, accuracy of a specific metacognitive judgment, learners’ summaries and notes, etc.) and 
contextualize other data channels to identify when, and potentially why, these CAM SRL 
processes occur (Taub, Mudrick, & Azevedo, 2018). Therefore, identifying the most appropriate 
set of valid and reliable measures indicative of CAM SRL signatures during learning, reasoning, 
and problem solving with ITSs is necessary to design systems that respond to the psychological 
and educational meaningfulness of these signatures in real-time.

As such, my dissertation outlines the conceptual and theoretical contributions of using 
multimodal, multichannel process data to detect and understand all of a learner’s CAM processes 
during learning with ITSs and presents useful experimental and methodological approaches that 
assist in detecting behavioral signatures of these CAM processes. Ultimately, detecting CAM 
processes using these methodologies will help design systems that automatically detect, respond 
to, and foster effective SRL with ITSs.

**Conceptual and Theoretical Contributions**

Understanding how CAM SRL processes occur and interact during learning with ITSs 
requires the integration of multiple theories and frameworks; no single framework exists to 
address all of these processes. For example, Mayer’s (2014) Cognitive Theory of Multimedia 
Learning can be used to explain the cognitive processes underlying selecting, organizing, and 
integrating text and diagrams embedded in ITSs. However, this model does not specify how
learners’ emotions or metacognitive monitoring processes may facilitate or impede the creation of an adequate mental model of the multimedia content. Though Winne & Hadwin’s (1998, 2008) Information Processing Theory of SRL and Greene & Azevedo’s (2009) macro- and micro-level framework of SRL outline the cognitive and metacognitive processes underlying effective SRL, they do not theorize the impact of learner emotions on these processes. Lastly, D’Mello & Graesser’s (2012) Model of Affective Dynamics discusses how learners’ emotions experienced during learning can positively or negatively influence their cognitive processing, but it does not specify how learners’ metacognitive monitoring processes may cause or be caused by these emotions.

Therefore, using multimodal, multichannel process measures can allow us to concurrently investigate how learners’ CAM SRL processes influence each other during learning with ITSs, thereby increasing our theoretical understanding of how these processes unfold and interact. For example, eye tracking can be used to indicate when learners are cognitively processing learning content, while facial expression recognition and EDA data can identify how they are feeling (e.g., confused or frustrated), and log files can indicate the accuracy of their metacognitive judgments for that content, thereby allowing us to measure all of a learners’ CAM processes concurrently. Furthermore, these data can augment our understanding of the potential sequences of CAM processes during SRL. For example, they can identify if there are specific causal or correlational relationships between learners’ CAM processes, such as if inaccurate metacognitive judgments are preceded by a facial expression of confusion, or if facial expressions of joy are caused by learners’ examining positive feedback provided by the ITS. Ultimately, these multimodal, multichannel process measures can help us understand how specific CAM processes
impact others, and therefore aid in designing systems capable of detecting and responding to these processes to effectively foster learners real-time SRL.

**Methodological and Experimental Approaches**

To augment conceptual and theoretical frameworks of CAM SRL processes, we can use experimental designs that help validate multimodal, multichannel process measures to detect learners’ CAM SRL processes with ITSs. Specifically, within-subjects experimental designs where learners interact with counter-balanced trials and where we also collect data from several process measures (e.g., eye tracking, videos of learners’ facial expressions of emotions, EDA sensors, and log files) are important to identify behavioral signatures of CAM SRL processes. These experimental designs can help constrain the context during which these processes occur and with these designs, we can structure the experimental content to contain specific information (e.g., multimedia science content containing discrepancies within the text, or between the text and graph), prompt learners to make a singular metacognitive judgment, engage in a specific cognitive strategy, etc. Then, we can use these designs to identify the associations between learners’ CAM SRL processes, learning outcomes, and their eye movements, emotions, and EDA. For example, if we want to determine if learners’ skin conductance responses (derived from their EDA) predict their ease of learning judgments (i.e., metacognitive judgments where learners rate their confidence in their ability to learn information needed to answer a question), we can use EDA data and the metacognitive judgments collected during this stage in the experiment. Because we designed the experiment to only prompt ease of learning judgments on certain stages of the trial, we know when they occur, and we know that it is likely that learners are engaging in metacognitive monitoring processes when they make these judgments due to the absence of other content. We can be relatively confident that we are examining the predictive
relationship between learners’ skin conductance responses on these pages and their ease of learning judgments.

Additionally, these within-subjects experiments can be designed to induce the occurrence of CAM SRL processes during learning. For example, if we want to understand how learners’ confusion impacts their ability to accurately evaluate the relevancy of the multimedia materials they are viewing to their current learning goal, we can design content to include discrepancies (e.g., contradictions, incongruities, unexpected information) because we know from previous research that these discrepancies can reliably impact learner confusion (e.g., D’Mello & Graesser, 2012; D’Mello, Lehman, Pekrun, & Graesser, 2014). For example, in a manuscript included in this dissertation, we designed an embedded human tutor to facially express emotions that could be congruent, incongruent, or neutral with the relevancy of the multimedia content (Mudrick, Rowe, Taub, Lester, & Azevedo, 2017). We examined the impact of these facial expressions, and the timing of the human tutor’s facial expressions (i.e., before, during, and after the human tutor’s expression) influenced learners’ levels of confusion (identified from their own facial expressions of confusion). Results indicated that learners’ confusion was significantly impacted by the type and timing of the human tutor’s facial expressions. These types of experimental designs allow researchers to constrain the learning context, induce the occurrence of specific CAM SRL processes, and then turn to the multimodal, multichannel data to understand how these processes can be identified to then be implemented into systems that detect and respond to these processes in real-time.

To address the previously presented issues, more research is needed to answer questions such as: How do learners’ CAM SRL processes unfold, interact, and change during learning with ITSs? Are behavioral signatures that indicate effective CAM SRL processes during learning
measurable with multimodal, multichannel process data? Is there an optimal set of process measures (e.g., eye tracking, videos of learners’ facial expressions, log files, EDA) to indicate CAM SRL processes? Can we use these identified signatures as an evidence base to develop CAM-sensitive ITSs and identify behaviors that predict accurate and effective SRL? Therefore, my dissertation begins to address some of these questions by examining multimodal, multichannel process data to investigate these questions and several others, to ultimately design systems intelligent enough to address all of a learner’s real-time CAM SRL processes.

**Goal of this Dissertation**

Research investigating CAM SRL processes with multimodal, multichannel process data has traditionally converged two data channels (e.g., log files with eye tracking, log files with facial expressions data, etc.) to investigate the influence of specific CAM processes on SRL with ITSs. However, this research has yet to examine the complex interplay of all of these processes with multiple sources of process measures (i.e., converging log files, facial expressions data, and eye tracking). The goal of this dissertation is to demonstrate how multimodal, multichannel process data can augment our conceptual, theoretical, methodological, and practical understanding of CAM SRL processes during learning with ITSs. Specifically, I argue that detecting and understanding learners’ real-time CAM processes with these data is necessary to individualize scaffolding, support, prompting, etc. and ultimately foster effective SRL with ITSs. The manuscripts in this dissertation provide a foundation to address these issues. More specifically, I address how we can answer questions such as: How can multimodal, multichannel process data augment our conceptual and theoretical understanding of CAM SRL processes during learning with ITSs? How well can micro-level metacognitive judgments (e.g., ease of learning judgments) predict performance, and should we design systems capable of adapting to
these judgments? How should we design embedded agents to augment and foster CAM SRL processes? And can expert human tutors augment our understanding of detecting real-time CAM processes to design systems that detect and respond to learners’ CAM SRL processes during learning?

Overall, the goal of this publication-based dissertation is to demonstrate the utility of multimodal, multichannel process measures to detect learners’ CAM SRL processes during learning with ITSs. I provide specific, empirical evidence that underscores how we can detect and understand the influence of CAM processes on learners’ abilities to effectively self-regulate their learning with within-subjects experimental designs. I also provide examples of how we can move beyond these types of approaches to design CAM-sensitive ITSs. This dissertation includes four manuscripts that discuss conceptual, theoretical, methodological, and design components related to assessing CAM SRL processes during learning with ITSs. The next four chapters are copies of submitted manuscripts. The first manuscript is a theoretical handbook chapter (Azevedo, et al., in press), which outlines the issues we face using contemporary theoretical frameworks of cognition, affect and emotions, metacognition, and SRL to understand behavioral signatures of CAM SRL processes derived from multimodal, multichannel data. The next two manuscripts are two empirical refereed conference proceedings that provide evidence of the complex interplay between affective and cognitive (Mudrick, Rowe, Taub, Lester & Azevedo, 2017) and cognitive and metacognitive (Mudrick, Sawyer, Price, Lester, & Azevedo,
SRL processes during learning with a multimedia-based ITS\(^1\). The final manuscript (Mudrick, Taub, & Azevedo, in press) is a chapter that describes a new system that we have developed that will allow us to investigate how to detect, understand, interpret, and respond to real-time CAM SRL processes during learning. In the next section, I briefly outline the contributions of the manuscripts to my dissertation.


In this handbook chapter, Azevedo, Mudrick, Taub, & Bradbury (in press) provide an overview of the challenges of understanding CAM SRL process data using contemporary theories and frameworks from the cognitive, educational, and learning sciences. The authors review ways that researchers can measure and support CAM SRL processes with different types of computer-assisted learning environments (CALSs) such as ITSs and discuss contemporary models related to developing CALSs to scaffold and support these processes. The authors discuss the limitations of these theories to equally address and explain CAM SRL process data during learning with these systems and focus on the challenges related to measuring and supporting CAM SRL processes during learning with CALSs. Specifically, Azevedo et al. (in press) present some conceptual dichotomies (i.e., comparisons between overarching issues) related to detecting, detecting, detecting,

\[^{1}\] Peer-reviewed conference proceedings fulfill the requirements of the publication-based dissertation guidelines.
measuring, understanding, and interpreting CAM SRL processes with these theories. For example, the authors note that several considerations must be made when researchers use multimodal, multichannel process data. These considerations include understanding how these measures can provide evidence for the quality or the quantity of specific CAM SRL processes (e.g., what does a higher frequency of a learners’ facial expression of confusion suggest, and how does it impact their SRL?), if these measures can allow researchers to understand time- vs. event-based thresholds of psychological meaningfulness, and the need to understand if these processes co-occur during learning or if they are discrete events. This chapter outlines these key conceptual and theoretical issues regarding the use of multimodal, multichannel CAM SRL process data to augment contemporary theories and design adaptive ITSs that will be exemplified in the next two empirical manuscripts.


In this empirical study, Mudrick et al. (2018) examined the relationship between accurate metacognitive monitoring and learners’ performance with an ITS, MetaTutorIVH, to provide evidence for future designs of learner-model features. These features should be included in the development of new ITSs capable of assessing learners’ metacognitive processes in real-time. Specifically, the authors used a within-subjects design to constrain the occurrence of specific metacognitive and cognitive processes to investigate the accuracy of ease of learning (EOL) judgments (judgments whereby learners gauge their confidence in being able to learn the
information needed to answer questions) and how these judgments predicted learner performance during learning about complex biology content. The authors standardized learners’ EOL judgments to compare across learners and accounted for content difficulty (by accounting for the accuracy of multiple-choice responses across the total sample). Results indicated that learners who judged the content as harder to learn outperformed learners who judged the content as easier to learn. Furthermore, results indicated that learners who judged the content as harder to learn than it actually was significantly outperformed learners who judged the content to be easier to learn than it actually was. Lastly, results indicated that these EOL judgments significantly predicted accurate multiple-choice responses 71.7% of the time.

Though this study only used one data channel (i.e., log files) to assess the relationships between learners’ cognitive and metacognitive SRL processes, the findings emphasize the utility of these within-subjects designs to augment our understanding of these processes. Additionally, this study evidences the utility of this approach to identify features to include in ITSs that are capable of adapting to learners’ CAM SRL processes in real time. For example, the authors note that the cost to develop these EOL-type features in future systems is minimal and would be easy to integrate into systems to provide the ITS with information that they can use to individualize scaffolding, support and feedback to learners based on these judgments.

In this empirical study, Mudrick, Rowe, Taub, Lester, & Azevedo (2017) examined the impact of specific facial expressions provided by an embedded tutor agent on learners’ emotions (i.e., joy, confusion, and frustration) and learning outcomes with the MetaTutor Learning Environment. The authors used a within-subjects design where they manipulated the congruency of the tutor agent’s facial expressions to the relevancy of the multimedia content in the environment. Specifically, learners interacted with three levels of relevant content: high relevancy (where the text and diagram were fully relevant to the science question learners needed to answer), low text relevancy (where the text was less relevant, but the diagram was still fully relevant), and low text relevancy (where the diagram was less relevant, but the text was still fully relevant). In addition, the authors manipulated the congruency of the human tutor agent’s facial expression to be congruent (i.e., joy for fully relevant content, confusion for less relevant content), incongruent (confusion for fully relevant content, joy for less relevant content), and neutral (included as a control). In addition to congruency manipulation, the authors also examined the timing of the human tutor agent’s congruent, incongruent, and neutral facial expressions and the influence time prior to, during, and after the expression on learners’ own facial expressions of joy, confusion, and frustration. Results indicated that learning outcomes were highest when the human tutor agent facially expressed congruent facial expressions. Results also suggested that learners facially expressed more confusion when the human tutor agent provided incongruent expressions. These results emphasize the utility of within-subjects designs to augment our understanding of behavioral signatures of CAM SRL processes and our ability to methodologically constrain the context to investigate these processes. In this case, we found a significant impact of the tutor agent’s facial expressions on learner confusion as the
human tutor made the expression. These findings offer suggestions to design embedded tutor behaviors based on learners’ emotions.


In this chapter, Mudrick, Taub, and Azevedo (in press) discuss a system they developed called MetaMentor, which presents visualizations of multimodal, multichannel data of learners’ cognitive, affective, metacognitive, and motivational SRL processes to experts (both tutors and researchers) in real-time. Specifically, the authors outline the empirical history of using findings from research on expert human tutors to design ITSs to detect and respond to certain CAM processes during learning. The authors argue that presenting these visualizations of learners’ CAM SRL processes (derived from multimodal, multichannel data) can help identify specific CAM SRL signatures that can then be implemented by future iterations of the system to foster these processes for learners in real-time. Ultimately, this chapter explains how to apply findings from the previous experimental studies and augment our experimental approaches to further investigate how to detect and understand learners’ CAM SRL processes during learning with ITSs.

_____________________

This chapter also includes description of the importance of motivational processes during SRL with ITSs, however the focus of this dissertation is to specifically focus on learners’ cognitive, affective, and metacognitive processes.
Summary

Overall, the four proposed manuscripts provide evidence for the use of multimodal, multichannel data coupled with specific experimental approaches to identify accurate and valid CAM SRL behavioral signatures during learning with ITSs. There are many conceptual, theoretical, and analytical issues that must be considered to design systems to address all of a learner’s CAM SRL processes in real-time. These issues are considered with the first publication (i.e., Azevedo et al., 2018), which provides specific examples of how to augment contemporary frameworks related to CAM processes during SRL. Whereas, the two refereed conference proceedings (i.e., Mudrick et al., 2017, 2018) provide specific empirical examples of how to identify and understand the impact of CAM SRL processes with within-subjects experimental designs. Lastly, the final publication outlines a new method to detect and understand CAM SRL data signatures and presents applications of the empirical findings from Mudrick et al. (2017; 2018). As such, this dissertation exemplifies the conceptual, theoretical, methodological, and design issues related to using multimodal, multichannel CAM SRL process data to design adaptive ITSs.
CHAPTER 2: SELF-REGULATION IN COMPUTER ASSISTED LEARNING SYSTEMS

Computer-Assisted Learning Systems (CALSs) such as intelligent tutoring systems (ITSs), game-based learning environments (GBLEs), and virtual reality have the potential to transform educational outcomes by supporting and augmenting learners’ ability to accurately monitor and regulate key cognitive, affective, metacognitive, motivational and social processes. CALSs can now be outfitted with natural language processing, machine vision, haptic devices, and physiological sensors that can potentially improve learners’ self-regulation while significantly challenging current frameworks, models, and theories of self-regulation. These advances can significantly augment existing CALSs both as research tools as well as instructional tools. For example, including natural language processing in CALSs will allow for the collection of concurrent verbalizations that can be coded in real-time to measure the deployment of metacognitive judgments (e.g., determine the relevance of instructional material) as well as be used by the system to determine whether the learner needs scaffolding to address particular metacognitive challenges (e.g., inability to determine the relevance of instructional material). Despite the temptation to design and instrument CALSs with the most recent technological advances researchers need to make theoretically-driven and empirically-based decisions based on current models, frameworks, and theories of self-regulated learning (SRL).

The goal of this chapter is to present research on self-regulation in CALSs by providing examples from contemporary CALSs and also how we use multimodal multichannel data to

---

3 We acknowledge that motivational and social self-regulatory processes are key to understanding self-regulated learning; however, in this chapter we only focus on cognitive, affective, and metacognitive processes due to space limitations.
examine cognitive, affective, and metacognitive (CAM) self-regulatory processes in these systems. Specifically, our chapter focuses on the role, measurement, and support of cognitive, affective, and metacognitive CAM processes during learning and problem solving with CALSs.

The structure of our chapter is as follows. First, we will provide a brief history of research in SRL with CALSs and discuss how different CALSs have been used to study and foster SRL. We will then present and discuss conceptual and theoretical issues derived from several models, frameworks, and theories of SRL that focus on CAM processes. Following this, we discuss several dichotomies related to CAM and the challenges they pose for the measurement and support of SRL with CALSs. Lastly, we present challenges and future directions that need to be addressed by researchers.

**History of Research in SRL and CALS**

The last several decades have seen major shifts in conceptual, theoretical, methodological, and analytical advances that have challenged research on the effectiveness and design of CALSs, as evidenced by certain movements such as cognitivism, constructivism, radical constructivism, constructionism, situated cognition, embodied cognition, etc. (see Mayer & Alexander, 2017; Sawyer, 2014; Dunlosky & Rawson, this volume). Additionally, research methodologies have become more intricate, as researchers have sought to explore the underlying learning processes students use with CALSs. For instance, studies have shifted from experimental designs using pre-test/post-test designs (i.e., product data) to including self-reports (e.g., learners’ self-perception of strategy use, metacognition, motivation, emotions) and more recently, process data (see Azevedo, Taub, & Mudrick, 2018). Process data (e.g., eye tracking, log files, facial expressions of emotions, physiological sensors, screen recordings, gestures, learner-artificial agent interactions, etc.) allow researchers to capture the CAM SRL processes
students deploy when learning with CALSs in real-time (Taub & Azevedo, 2016a). CALSs were originally designed purely as an instructional tool, but with the ability of process data to detect complex CAM processes in real-time, CALSs are now being used to better understand the complex learning processes students engage in during learning, problem solving, conceptual understanding, scientific reasoning, etc. (see Azevedo & Aleven, 2013; Schunk & Greene, 2017). Additionally, research has expanded from focusing solely on cognitive processes during learning to including other areas such as the metacognitive and affective processes involved in learning. Further, a plethora of studies have provided clear evidence that there is no one size fits all CALSs; therefore, the ultimate goal is to develop adaptive CALSs that can accurately detect students’ CAM SRL processes and then act on this information by providing timely, intelligent, and adaptive scaffolding specific to each student’s needs. In the next section, we describe several types of CALSs and how they measure and foster CAM SRL.

**Computer-Assisted Learning Systems (CALSs): An Overview**

There are several different types of CALSs including multimedia, hypermedia, ITSs, GBLEs, teachable agents, computer-supported collaborative learning (CSCL) environments, robots, and more recently intelligent virtual humans (IVHs), virtual reality (VR), augmented reality (AR), and tangible computing. Mayer (2014a) defines multimedia learning as knowledge construction and building mental representations from text (e.g., spoken, written) and pictures (e.g., illustrations, videos, animations, photos, etc.), while hypermedia is defined as knowledge construction from computer-based media (text, images, videos, etc.) navigable via hyperlinks (e.g., the internet; Azevedo & Cromley, 2004). Several studies have demonstrated the educational effectiveness as well as the promotion of self-regulation of both multimedia and hypermedia learning environments (Azevedo, 2014; Azevedo, Feyzi-Behnagh, Duffy, Harley,
Trevors, 2012; Mayer, 2014a). For instance, hypermedia can provide learners with a wealth of non-linear information including text, diagrams, animations, and videos; however, many students lack the CAM SRL skills necessary to effectively navigate these systems (Azevedo et al., 2017; Bannert & Reimann, 2012). Early work with the hypermedia environment Encarta found that training SRL skills prior to using the hypermedia environment led to more sophisticated mental models of the circulatory system compared to a control (Azevedo & Cromley, 2004). Additionally, Bannert and Reimann (2012) found that SRL prompts embedded within a hypermedia environment improved overall learning compared to a control; however, the best effects were observed when students received a short training on SRL in addition to the SRL prompts.

ITSs are computerized learning environments designed to mimic expert human tutors’ pedagogical moves (e.g., feedback, scaffolding) by using computational models from the fields of cognitive psychology and artificial intelligence (Aleven & Koedinger, 2016; Kulik & Fletcher, 2016) to provide learners with individualized, adaptive and intelligent scaffolding in real-time. A recent meta-analysis found that students who used an ITS outperformed students who received traditional classroom instruction 92% of the time (Kulik & Fletcher, 2016). ITSs have a long tradition in cognitive and educational psychology and learning sciences and have been built to address learning and problem solving issues in several well-structured as well as ill-structured domains for learners of all ages (e.g., Cognitive Tutors for algebra problem solving, AutoTutor for physics misconceptions, MetaTutor for teaching biology and self-regulated learning; see Wolff, 2009 for an extensive review). The majority of research on ITSs has focused on learners’ cognitive and metacognitive SRL processes while some recent systems are now addressing
learners’ affective processes such as Graesser and D’Mello’s (2012) extensive work on AutoTutor and other affect-sensitive CASLs.

There is a tremendous emphasis on designing and using GBLEs to promote learning across domains and for learners of all ages (Mayer, 2014b). Some GBLEs are designed to train SRL skills along with educational content embedded within a narrative (e.g., Crystal Island; Rowe et al., 2011) while others can be level-based with new problems being presented at each level (i.e., Physics Playground; Shute et al., 2016). Past research on GBLEs has demonstrated mixed results in terms of their educational effectiveness (Mayer, 2014b); however, a recent meta-analysis found that GBLEs were more educationally effective compared to traditional instructional methods (Wouters, Van Nimwegen, Van Oostendorp, & Van Der Spek, 2013). Other types of CALSs, such as open-ended systems such as Teachable agents (e.g., Betty’s Brain, SimStudent) were explicitly designed using the learning-by-teaching paradigm in which students learn material by teaching a peer agent (Biswas, Segedy, & Bunchongchit, 2016; Matsuda, 2013); several studies have demonstrated their educational effectiveness (Biswas, Segedy, & Bunchongchit, 2016; Matsuda, 2013, Biswas, Jeong, Kinnebrew, Sulcer, & Roscoe, 2010). Additionally, teachable agents, such as Betty in Betty’s Brain, can provide SRL support by having students monitor the learning of the teachable agent via questions and quizzes, therefore monitoring their own understanding. In contrast, CSCL environments are designed to facilitate knowledge and expertise sharing via collaborative activities (e.g., Lin, Preston, Kharrufa, & Kong, 2016). Past research with CSCLs, including Bioworld and Quest Atlantis have demonstrated increased learning gains for Quest Atlantis and improved diagnostic reasoning after using Bioworld (Barab et al., 2007; Lajoie, Poitras, Doleck, & Jarrell, 2015).
CSCL systems are currently being used as testbeds to evaluate assumptions of co-regulated, shared self-regulated learning, and socially shared regulated learning (Järvelä et al., 2016).

As technology advances so do opportunities to design more sophisticated CALs that were not possible decades ago. For example, virtual reality (VR), augmented reality (AR), and tangible computing are the newest CALSs which, as the technology become more readily available, new games, virtual environments, and simulations are being developed. For example, VR has been around since the 1960’s (Heiling, 1962); however, widespread dissemination was limited due to practical and technological constraints (Merchant, Goetz, Cifuentes, Keeney-Kennicutt, & Davis, 2014). VR is an artificial immersive environment in which students experience a virtual world via a head-mounted display. AR layers computer-generated images across real-world stimuli via a head-mounted display or hand-held computer (e.g., table, cellphone, etc.). Where VR diverts attention away from the real world, AR enhances it. A recent meta-analysis gauging the effectiveness of VR-based instruction on learning outcomes in K-12 and higher education found that VR is an effective educational tool with an effect size of .36 (Merchant et al., 2014) while a meta-analysis of AR-based instruction found a moderate positive effect on student performance with an effect size of .56 (Santos et al., 2014). We argue that both VR and AR likely require a high degree of SRL skills (e.g., monitoring effective strategy use, coordinating relevance of multiple information sources) due to the unmatched affordances provided by these environments (e.g., realism that may reduce misconceptions in science, immersion and presence may sustain motivation and emotions while allowing prolonged persistence during complex problem solving); however despite its importance, there is a dearth of research studying the SRL processes students employ or fail to employ during learning with these environments.
**CALSs and Self-Regulated Learning: Examples**

In this section, we describe how SRL has been used to both design and measure SRL processes in several CALS. For example, MetaTutor is an adaptive hypermedia-based ITS designed to detect, model, and foster students’ SRL while learning about the human circulatory system (Azevedo et al., 2009). The design of MetaTutor is based on Winne and Hadwin’s model of SRL (1998, 2008; Winne, 2017) and Mayer’s (2014) Cognitive Theory of Multimedia Learning and has undergone several re-designs based on extensive SRL research with the system for over a decade (Azevedo, Johnson, Chauncey, & Burkett, 2010; Duffy & Azevedo, 2015; Harley, Taub, Azevedo, & Bouchet, in press; Taub & Azevedo, 2016). MetaTutor provides feedback and scaffolding via four pedagogical agents—Gavin the Guide, Pam the Planner, Sam the Strategizer and Mary the Monitor—each embodying theoretical assumptions underlying monitoring and regulation of cognitive and metacognitive SRL processes (see Figure 1 for a screenshot of the MetaTutor interface). In addition, to collecting pretest, posttest and other self-report measures of metacognition, motivation, and emotions, MetaTutor has been used as a research tool that collects multichannel multimodal process SRL data during learning via the use of log-files, eye tracking, videos of facial expressions of emotions, screen recordings, physiological activity, and learner-agent dialogue.

**CRYSTAL ISLAND** is a narrative-centered GBLE designed to foster students’ SRL, scientific reasoning, literacy, and problem-solving skills (Rowe, Shores, Mott, & Lester, 2011). The game is experienced in first person perspective and is set on a tropical island where students learn of a mysterious illness infecting the community. Taking a protagonist role, participants explore the island, collecting clues as they speak with residents and patients, read and answer content on microbiology, and use lab equipment to scan food items, all to discover the source,
identity, and best treatment option for the infectious disease. CRYSTAL ISLAND has been used in several lab experiments involving undergraduates and has also been used in thousands of middle school classrooms to engage students in microbiology content learning and problem solving (e.g., Taub, Mudrick, Azevedo, Millar, Rowe, & Lester, 2017; Taub, Azevedo, Bradbury, Millar, & Lester, 2017). Additionally, the system was first developed over a decade ago and has since served as a research and educational tool for investigating game-based learning issues related to SRL including emotions, motivation, engagement, presence, and agency (e.g., Bradbury, Taub, & Azevedo, 2017; Sawyer, Smith, Rowe, Azevedo, & Lester, 2017).

The design of Betty’s Brain is based on the learning-by-teaching paradigm (Bargh & Schul, 1980) to improve SRL and STEM learning (Biswas et al., 2016) for adolescents. To teach Betty, students first learn science content (ecology) and then construct causal concept maps illustrating ecological processes. Next, they ask Betty questions by getting her to take quizzes on the content, which she answers based on what she was taught (i.e., what information and links students include in their concept map). Her performance on these quizzes provides students feedback on what information needs to be added to the concept map, and what information in the concept map may be incorrect, prompting students to identify strategies for adding relevant information and identifying errors (Biswas et al., 2016). The quizzes are administered by a mentor agent named Mr. Davis who students could seek help from at any time by clicking the “Ask Mr. Davis” button. Mr. Davis can answer questions such as “How do I build a concept map?”, “How do I search for specific information in the resources?”, etc. Additionally, when the student is not performing well, Mr. Davis can suggest a certain learning strategy to get them back on track. Later versions of the system also include a motivational component by having Betty
express happiness when her quiz scores improve and disappointment when her quiz scores fail to improve (Biswas et al., 2016).

In this section, we described how several contemporary CALS have been designed to measure and foster SRL by embodying assumptions of various models. In addition, we have also described how they have been instrumented as research tools to collect learning outcomes, self-reports, as well as rich temporally unfolding SRL process data to detect, measure, and understand the role of CAM processes during learning, problem solving, etc. across domains and learners while addressing contextual constrains related to the collection of process data (e.g., scaling up CALS as educational and research tools from the research laboratory to schools and informal settings limits the ability collect precise eye-tracking data from learners due a variety of economic, privacy, and pragmatic issues).

**Conceptual and Theoretical Issues Derived from Models on Cognitive, Affective, and Metacognitive Self-Regulated Learning Processes**

In this section, we focus on describing and critiquing how three leading models of CAM SRL processes have been used with multimodal multichannel process to explain, measure, and understand these processes in CALSs. Specifically, we present the assumptions and mechanisms, and discuss the levels of explanation (i.e., grain size) and empirical support for Winne & Hadwin’s (1998, 2008) information processing theory (IPT) of SRL, Greene & Azevedo’s (2009) extension of the IPT model of SRL which details macro- and micro-level SRL processes, and D’Mello & Graesser’s (2012) model of Affective Dynamics, which describes the underlying affective processes that may occur during learning with CALSs. We also focus our critique on the limitations of these models and question how we can use them to understand multimodal multichannel CAM SRL process data during learning with CALSs. We end this section by listing
questions arising from the conceptual dichotomies presented in the use of these models to understand CAM SRL processes during learning with CALSs.

**Winne and Hadwin’s Information Processing Theory of Self-Regulated Learning**

Winne & Hadwin’s (1998, 2008) IPT of SRL is one of the leading theories used to detect, measure, and interpret SRL process data to examine the temporal unfolding of SRL during learning with CALSs. The model assumes that SRL is an event and is dependent on the relationship between the learner’s accurate metacognitive monitoring and control (i.e., regulation) of different cognitive operations (i.e., SMART processes) over four loosely sequenced, recursive learning phases (i.e., task definition, goal setting and planning, engaging in cognitive operations, and adaption). Winne (2017) suggests that during each of these four phases, self-regulated learners engage in one or many cognitive operations as they concurrently monitor the quality of the outcomes from these processes. For example, learners engage in searching for specific information within the learning content, assembling that learned information with previously learned information, rehearsing the information by preserving it in working memory, and translating information by changing its representation (e.g., creating a diagram from a from reading a text), while concurrently monitoring the quality of the outcomes from these SMART processes, based on internal standards.

In addition to the enactment of these five cognitive operations, the model specifies that SRL unfolds over four phases. In phase one, self-regulated learners define the learning task based on their perceptions, as well as the constraints of the task. During phase two, the learner generates goals and plans for future learning strategies. In phase three, self-regulated learners engage in the previously discussed cognitive operations based on their plans and goals of phase two. Lastly, in phase four, learners adapt their strategies based on their performance and output
of their enacted learning strategies. As a loosely sequenced and recursive model, the IPT model of SRL specifies that learners do not always progress through these phases in a sequential manner and instead are constantly moving between these phases as they self-regulate their learning.

As Winne & Hadwin’s (1998, 2008) model assumes SRL is an event, it is an optimal model to use when examining how cognitive and metacognitive SRL processes unfold during learning with CALSs (Winne & Azevedo, 2014). There is extensive empirical evidence in support of using Winne & Hadwin’s (1998, 2008) model to examine these processes during learning with CALSs (see Azevedo et al., 2013, 2017; Taub, Azevedo, Bouchet, & Khosravifar, 2014; Taub et al., 2017). However, there are also several issues regarding its ability to explain, describe, and predict SRL processes with multimodal multichannel process data during learning with CALSs. Specifically, the model argues for mechanisms that span across micro- and macro-levels, and as such, does not expand on some processes that are necessary to explain SRL during learning with CALSs. For example, how are learners’ internal conditions (e.g., prior knowledge) enacted during learning? Can they be accessed in real-time during learning with CALSs? What process data are required to assess how a learner evaluates their products based on their internal standards? Can CALSs be designed to make these real-time processes visible to facilitate learners’ metacognitive monitoring and control during learning?

Furthermore, as the model is loosely sequenced and recursive, it does not specify the optimal sequence, temporality, or dynamics of cognitive and metacognitive processes that learners need to engage in during learning, making it difficult for CALSs to respond to the learner effectively by tailoring individualized, adaptive scaffolding and support. For example, is there an optimal period of time during which learners define the task and if so, how do we assess
if a learner does not accurately perceive and define the learning task? Although the model has recently begun to address learners’ affective and motivational processes during learning (e.g., as internal conditions; see Winne, 2017), the model does not account for the specific influence of these processes and instead implies that SRL is dependent on cognitive and metacognitive processes. Lastly, the model is based on the assumption that learners can inherently engage in these learning phases and self-regulate their learning. However, it is possible that some learners may not be able to effectively enact some of these processes. If so, how can CALSs provide support and adapt their scaffolding and feedback to learners who cannot appropriately define their task, or set appropriate learning goals, etc.?

**Greene & Azevedo’s Macro and Micro Level Framework of Self-Regulated Learning**

The seminal work on metacognition (e.g., Nelson & Narens, 1990; Dunlosky & Metcalfe, 2009) and SRL (e.g., Bjork, Dunlosky, & Kornell, 2013; Hacker, Dunlosky, & Graesser, 2009), has been influential in developing Greene and Azevedo’s (2009) macro and micro-level framework of SRL expands on Winne & Hadwin’s (1998, 2008) model of SRL by detailing specific macro and micro-level, valenced, cognitive and metacognitive processes that students enact during SRL. As it is an extension of Winne and Hadwin’s (1998, 2008) model, it is based on the same assumptions that SRL is an event that is dependent on the interplay between accurate metacognitive monitoring and regulation (of cognitive learning strategy use). The Greene & Azevedo (2009) framework argues that SRL is comprised of five macro-level processes of planning, monitoring, strategy use, perceptions of task difficulty and demands, and interest, and within these macro-level processes are 35 specific micro-level processes. In addition to defining these processes, the framework indicates that micro-level cognitive and metacognitive judgments are valenced. In general, valence indicates the accuracy (i.e., positive
or negative) of a learner’s own evaluations, familiarity with, or understanding of their judgments (Azevedo et al., 2013). For example, when a learner performs (e.g., via a mouse click on an interface element or a concurrent utterance during learning with a CALS) a content evaluation (CE; an evaluation of the relevancy of the content to the learners’ current goal, within the macro-level category of monitoring), the evaluation could be positively valenced (i.e., the content is relevant to their goal), or negatively valenced (i.e., the content is irrelevant to their goal). Alternatively, when learners engage in prior knowledge activation (PKA; searching for relevant prior knowledge before beginning the learning task, within the macro-level category of planning), a positively valenced PKA would indicate the learner recalled relevant learning material, while a negatively valenced PKA would indicate that the learner could not recall relevant information. The inclusion of valence within the framework contextualizes these cognitive and metacognitive processes within the specific learning context that may change depending on the learning task, CALS, etc. For example, a learner may make a negatively valenced PKA (i.e., activate irrelevant prior knowledge) when learning about the different vessels of the human heart with a hypermedia system, but may make a positively valenced PKA (i.e., activate relevant prior knowledge) when learning about the circulatory system of invertebrates in a VR system.

Like Winne and Hadwin’s (1998, 2008) model, there exists substantial empirical support for Greene and Azevedo’s (2009) framework of SRL to detect and explain what SRL processes occur and influence learning outcomes during learning with a CALS (see Azevedo et al., 2013, 2017; Taub et al., 2014). However, there are also similar issues for using this framework to identify, explain, and predict cognitive and metacognitive SRL processes with multimodal multichannel process data during learning with CALSs. Despite defining specific cognitive and
metacognitive SRL processes, the model does not specify the optimal sequence, temporality, or dynamics of how these processes should unfold. For example, what is the optimal length of a specific sequence of micro-level SRL processes? Is there a difference between the quality and quantity of these micro-level processes during learning? Does engaging in more processes contribute to higher learning outcomes? And does the quality of these processes improve over time? Additionally, as this framework was derived from analyzing think-aloud data during experiments involving a hypermedia learning system (e.g., Greene & Azevedo, 2009), it may not be representative of how these processes may unfold during learning with different types of CALSs. For example, re-reading (RR; under the macro-level category of strategy use) may not be an appropriate process to examine in a virtual reality-based learning environment with different representations of information. Lastly, similar to Winne & Hadwin’s (1998, 2008) model, Greene & Azevedo’s (2009) framework focuses primarily on cognitive and metacognitive processes at the expense of the affective. As such, it offers no specific information regarding when, why, or how learners’ emotions can influence the accuracy of their metacognitive monitoring or the deployment of effective cognitive learning strategies.

**D’Mello & Graesser’s (2012) Dynamics of Affective States Model**

Unlike the previously discussed models that focused exclusively on cognitive and metacognitive processes, D’Mello & Graesser’s (2012) Dynamics of Affective States Model focuses on the learner-centered emotions that can occur during learning with CALSs. Derived from research on emotions during learning with AutoTutor the model assumes that: (1) learners’ cognitive and affective processes during learning are linked during learning, (2) learners typically experience a specific subset of discrete, learner-centered emotions and, (3) the interplay between these emotions can differentially influence their learning outcomes.
The model predicts that learners in a state of deep learning (i.e., complex learning where learners demonstrate high levels of conceptual understanding) are in a base state of flow/engagement. Once the learner encounters an impasse (i.e., discrepancy in the presented information) they experience a state of cognitive disequilibrium (i.e., state of uncertainty arising from the discrepancy) and subsequently experiences confusion. If the learner does not effectively resolve this confusion by engaging in effortful problem solving to identify the source of the discrepancy (e.g., discrepancy between representations of information), they can transition into a state of frustration. However, if the learner is successful in their problem solving, they can transition back to a state of engagement/flow. Once the learner experiences frustration, the model predicts that persistent failure and blockage to the learner’s goals as they experience their frustration can subsequently lead to boredom and task disengagement.

The Dynamics of Affective States Model is one of the leading models used to describe emotions during learning with CALSs and is based on empirical evidence detailing the presence of these discrete emotions, as well as their dynamics over time (see D’Mello & Graesser, 2012; D’Mello, Graesser, Pekrun, & Leutner, 2014). The model can be easily extended to explain affective processes with multimodal multichannel process data during learning with other CALSs as it assumes that emotions are temporally dependent and unfold over time. However, there are several theoretical and conceptual issues regarding its descriptive and explanatory adequacy for examining emotions during learning with CALSs. First, the model does not provide a concrete operational definition of engagement/flow state learners experience when they engage in deep learning. As such, there remains an open question regarding how researchers can empirically assess engagement/flow with multimodal multichannel process data during learning with CALSs. Additionally, this model is limited in the number and types of the discrete emotions
that it examines and predicts, and cannot offer explanations or predictions regarding the presence of other emotions that can occur in academic settings (e.g., pride, anxiety, etc., Pekrun et al., 2011). Furthermore, while the grain size of this model is primarily micro-level discrete emotions (e.g., confusion, frustration) it does not include all possible learning-centered emotions nor does it emphasize the potential expression of basic emotions (e.g., anger) in cases where CALSs are used in non-traditional educational purposes (e.g., recently diagnosed breast cancer patient learning about basic biomedical knowledge with a MetaTutor-like CALS).

Lastly, the model is also based on the assumption that learners can effectively engage in emotion regulation processes (e.g., Gross, 2015) and can transition between emotions. However, if learners are not effective emotion regulators do they perpetually disengage from the task because of their inability to transition from confusion back to a state of engagement/flow? Additionally, can learners engage in other emotion-regulation strategies (e.g. attentional deployment) other than the effortful problem solving specified in the model? Although the model assumes that cognition and affective processes during learning cannot be separated, it does not account for the presence of metacognitive monitoring processes that may influence the emotions experienced. For example, can learners experience confusion because of their inaccurate metacognitive judgments? Lastly, it also lacks assumptions regarding the temporality of affective states. For instance, how long does a person need to be in a state of confusion before transitioning to frustration?

**Conceptual and Theoretical Dichotomies Related to CAM SRL Processes Related to CALS**

The previously described models and theories have been the leading frameworks researchers have used to examine CAM SRL processes with multimodal multichannel process data during learning with CALSs. However, there are several remaining questions when using
these models to explain CAM SRL processes with multimodal multichannel process data that can be organized by four conceptual dichotomies. In this section, we describe each dichotomy and raise several issues based on the models presented and discussed in the previous section by focusing on detection and measurement of CAM SRL processes during learning with CALSs. We end this section by presenting a table (see Table 1) with process-specific (i.e., cognitive, affective, and metacognitive) research questions based on the subcategories of these dichotomies.

The first conceptual dichotomy that arises from these models deals with the **quality vs. quantity** of the CAM SRL processes. Within this conceptual dichotomy we argue that there are three specific subcategories related to the (1) frequency and (2) duration of, as well as the (3) latency between the CAM SRL processes described in each model. For example, are higher frequencies of CAM SRL processes indicative of more effective SRL? What are the key components that distinguish low vs. high quality SRL (e.g., copying information vs. paraphrasing when given the opportunity to take notes during learning with a CALS) and are these dependent on the specific learning context, individual differences, task demands, etc.? Many of these models are primarily descriptive of the CAM SRL processes they discuss but do not address the qualitative features of these processes (e.g., quality of inferences). Instead, these models imply that a higher frequency and longer duration, as well as a shorter latency between these processes is indicative of effective SRL during learning. For example, both Winne & Hadwin (1998, 2008) and Greene & Azevedo (2009) imply that engaging in more cognitive processes can contribute to higher learning outcomes. Alternatively, the model of Affective Dynamics implies that shorter latencies between emotional state transitions (e.g., confusion to engagement/flow) contribute to increased learning outcomes. As such, these conceptual issues
must be addressed in future work to interpret learners’ CAM SRL processes with multimodal multichannel process data during learning with CALSs.

A second dichotomy presented in these models relates to their descriptions of **macro- vs. micro-level** CAM SRL processes. Within this conceptual dichotomy we argue there are three specific subcategories related to the (1) granularity (i.e., levels of description), (2) valence (i.e., specific features related to each process), and (3) interplay between these levels (i.e., micro-level processes influencing macro-level processes and vice versa). The grain size of CAM SRL processes described in these models range from micro- to macro-level. As such, these differing levels of description make the theoretical interpretation of the interplay between CAM SRL processes difficult and pose questions regarding the interpretation of the multimodal multichannel process data. For example, does a micro-level emotional state (e.g., a 1 sec state of confusion) influence a learner’s macro-level monitoring (e.g., judgment of learning) during a 2-hour learning session? Should grain size be standardized when examining these processes with multimodal multichannel process data? Or does the predictive ability between macro- and micro-SRL process differ based on the specific CAM process, temporal dynamics, evidence of adaptivity during learning, development of SRL competencies during learning with CALSs, etc.? As such, issues related to grain size, valence, and the interplay between micro- and macro-levels are important issues to consider when interpreting CAM SRL processes.

The third dichotomy deals with the **time vs. event-based** thresholds needed to theoretically interpret CAM SRL processes during learning with CALSs. Specifically, these models do not address how long CAM SRL processes last or how many CAM SRL processes should occur during SRL. Furthermore, these theories are limited in descriptions on how to identify the occurrence of CAM SRL processes and as such, researchers are left with several
open questions related to the identification of time- or event-based SRL processes. For example, does time thresholding (i.e., the amount of time a process needs to occur before being considered present) influence our theoretical understanding of how a process occurs during learning with CALSs? How long do CAM SRL processes need to occur for to be psychologically meaningful and detectable by multimodal multichannel process measures (and requiring intelligent intervention by a CALS)? More research is needed to identify the behavioral signatures of CAM SRL processes in order to identify how and when certain processes can be considered psychologically meaningful and influential for understanding their contributions to learning (e.g., longer periods of confusion, an optimal number of summaries, etc.) and where CALSs can potentially intervene to facilitate SRL (e.g., repeated evidence of a lack of ability to control metacognitive discrepancies).

The last conceptual dichotomy arising from these models deals with issues related to interpreting of CAM SRL processes as discrete events vs. co-occurring processes. Within this dichotomy, we argue there are four subcategories related to interpreting CAM SRL as (1) parallel or (2) serial processes, as well as how these processes (3) correlate with and/or (4) cause each other. Although these models interpret CAM SRL processes as events that unfold over time (e.g., Winne, 2017), they do not specify if these processes occur in a serial or parallel fashion. For example, does a learner monitoring themselves by using a judgement of learning (JOL) that leads to the use of a cognitive strategy to re-read stop the cognitive learning strategy of re-reading or does the learner use that cognitive strategy that leads to metacognitive monitoring or does the learner concurrently judge their learning as they continue reading? Alternatively, can a learner experience confusion and frustration simultaneously or does confusion need to serve as the precursor? This issue gets more complicated as researchers and designers consider measuring
CAM processes using multimodal multichannel data—e.g., is confusion expressed before, during, or after the onset of a metacognitive judgment such as a JOL? If there is a transition from confusion to frustration does that affective transition happen serially or in parallel as a result of repeated cognitive strategy use? One can see how using multimodal multichannel SRL process can begin to address these questions (e.g., Azevedo et al., 2017) to both extend models of SRL as well as design SRL-sensitive CALS. Lastly, although some models identify specific causal relationships (e.g., experiencing cognitive disequilibrium after encountering an impasse in the learning content), there are still questions regarding the correlations between CAM SRL processes. For example, are there certain affective states that are more likely to be present when making an incorrect JOL? Can we profile learners based on these correlations to identify optimal scaffolding and feedback during learning with CALS?

We end this section by presenting an initial list of process-specific (i.e., cognitive, affective, and metacognitive) research questions based on the subcategories of the conceptual dichotomies (see Table 1). We also provide examples of the types of multimodal, multichannel process data that can be used to detect, interpret, and analyze these processes during learning with CALSs.

**Assessing Self-Regulation using Multimodal Multichannel Process Data**

There are many ways to assess SRL during learning with CALSs using different types of data, such as self-report questionnaires or multimodal, multichannel data of real-time temporally unfolding processes. Using multimodal, multichannel data to measure SRL can be particularly useful because it allows us to use both obtrusive (e.g., prompted to make a metacognitive judgment) and unobtrusive (e.g., log files) methods to measure learners’ overt (e.g., making a summary after inspecting diagrams and reading text) and covert (e.g., gaze behaviors of reading
behavior and diagram inspection) SRL behaviors, as opposed to relying on self-perceptions of how they engaged, activated, used, reflected, etc. There are various different types of data channels we can use to assess SRL, and selecting the most appropriate data will depend on the research questions being posed, experimental design of the study, design of the CALSs themselves, and other critical issues (e.g., scaling up from lab to outside the lab, securing the privacy and anonymity of learners, pragmatic issues related to conducting data in real-world context such as classrooms, etc.). In the following section, we will provide examples of how different types of data channels can be used to assess CAM SRL processes, as well as the challenges we face when using these data.

**CALSs as a Research and Learning Tool**

One advantage of developing CALSs is that they can be used as both as research tools and learning tools. For learning, they can be used to ensure that students of all ages and ability levels are learning and engaging in effective self-regulation. This is clearly beneficial as we are aiming to improve educational opportunities for students. CALSs can also be used as a research tool, such that not only do we allow students to learn with these systems, but we collect large amounts of data that inform us *how* students are learning, including what types of self-regulation (i.e., metacognitive monitoring and cognitive learning strategies) they are (or are not) engaging in, when they use them, and whether or not they are used effectively.

When using CALSs to conduct research, the researchers themselves have many design considerations they have to make based on the research questions and hypotheses they want to answer, what is the research design, and what are the logistical considerations they have to make, which will then inform the types of data they want to collect (see Hacker et al., 2009; Schunk & Greene, 2017 for examples). First, the specific research questions that researchers want answered
by developing these systems will impact how they design the CALSs and the data they want to collect. For example, if researchers are interested in how a pedagogical agent’s facial expressions impact a student’s facial expressions during learning with a multimedia environment, the system has to include agents that are capable of emoting facial expressions that are commonly found across teachers, tutors, and previous research on students (e.g., confusion and its benefits for learning; D’Mello et al., 2014). Additionally, since the question also aims to investigate students’ emotions, the study protocol must also include collecting emotion data from students during learning. In contrast, if the research is to assess how students self-initiate the use of cognitive and metacognitive processes during learning with MetaTutor, and how this impacts their overall performance on the post-test, the system itself must support engaging in these processes (e.g., the SRL palette), and the study protocol must include a post-test (e.g., Taub et al., 2014). Thus, these questions, and the specificity of these questions, will guide researchers towards using CALSs as a research tool.

Moreover, the research design itself will also impact how research is conducted assessing these systems. Specifically, using a within-subjects vs. between-subjects design will impact how the researcher can assess the effectiveness of these systems for fostering SRL and learning. When comparing a within- vs. between-subjects design, there can be benefits and drawbacks to each design. A within-subjects design will expose the students to all conditions the system entails, such as progressing through a series of trials that expose students to all combinations of different discrepancy types (e.g., Burkett & Azevedo, 2012). This is a beneficial design because it allows us to examine how students change over time, and specifically across different contexts. If a student was interacting with a pedagogical agent that emoted different facial expressions, we could determine how these different emotions impact performance across different trials.
However, these designs can be time consuming and although trials are randomly ordered, there is still the chance of an order effect, such that a student becomes more familiar with each trial as they complete more of them. In contrast, a between-subjects design is not subject to the effects of trial; however, it can pose an unfair advantage or disadvantage to students in each condition, such that a student who receives prompts and feedback has the advantage of knowing their progress, and getting assistance when they perform poorly, but a student who does not receive this assistance has a disadvantage because they are not taught which processes to engage in, or told how they performed. Therefore, based on these design decisions, assessing how the system fosters the use of CAM processes will differ based on the design choice, such that there will be more time-varying data and data for all conditions for all students from a within-subjects design.

Finally, when designing these experimental procedures, a researcher must consider the logistics of the study, including how much time there is, and how many experimenters there are that have the expertise in setting up and calibrating the participant. Specifically, calibrating the experimental equipment can add up to 30-60 minutes during data collection, and so in addition to the learning session, which can last upwards of 1-3 hours, the entire procedure can last a relatively long time, and potentially induce undue fatigue, stress, etc. on participants. In addition, the setup itself can be complicated, and so the experimenter must have adequate expertise in calibrating the equipment, which includes troubleshooting errors. Thus, to setup a study, the more advanced the technology being used, the more difficult it can be to run the study itself, which is a consideration all researchers must be prepared to make when testing CALSs.

For this chapter, we focus on CALSs as a research tool, such that we discuss how we can use multimodal multichannel data to examine students’ CAM processes during learning with
CALSs. In the next section, we review some of the different types of data we can use to assess these processes.

**Types of Multimodal Multichannel SRL Process Data**

There are many different data channels we can use to assess self-regulation during learning with CALSs. We can use eye-tracking data to examine where students are fixating (i.e., focusing on for at least 250 ms; Rayner, 2009) during learning, which can be indicative of focused attention on particular areas of the screen (Scheiter & Eitel, 2016) thus demonstrating students are monitoring based on where they are paying attention to in the CALSs. Eye-tracking data can also be used to investigate students’ saccades, which are rapid movements between fixations (Rayner, 2009), and can indicate students are reading content, or are assessing whether or not the content is relevant to their overall learning goals. We can also investigate regressions, which involves a student returning back to a previously fixated on area of interest (Rayner, 2009), indicating a student is re-reading because they do not understand the content. Eye-tracking data are increasingly being used to assess self-regulation during learning with CALSs (Conati, Jaques, & Muir, 2013; Scheiter & Eitel, 2016; Taub & Azevedo, 2016), revealing how it can be useful to assess where students are focusing on the system during learning.

Another type of data includes videos of facial expressions of emotions, which reveal which emotions students are expressing during learning with CALSs. Students can experience a range of basic (joy, anger, surprise, fear, disgust, sadness) or learning-centered (confusion, frustration, boredom) emotions during learning, which research has shown can impact how they engage in self-regulation and overall performance (D’Mello et al., 2014; Mudrick, Rowe, Taub, Lester, & Azevedo, 2017). When investigating facial expressions of emotions, software (such as FACET; iMotions, 2017) can measure the evidence of these emotions (e.g., onset, triggering
event, frequency, duration, latency between states, context, alignment with other cognitive and metacognitive processes), defined as the likelihood of human coders coding for that emotion, using a logarithmic (base 10) scale. Therefore, evidence scores of 1, 2, or 3 indicates the likelihood of 10, 100, or 1000 human expert coders coding for that emotion, respectively. Furthermore, a positive value would indicate the likelihood of an emotion being coded as present, while a negative value indicates the likelihood of an emotion being coded as absent. Thus, in addition to identifying which emotions students are experiencing during learning with CALSs, we can also identify which emotions students are not expressing. For example, if the software detects the student is confused during learning, but not frustrated, this can indicate the student is trying to reestablish cognitive equilibrium (D’Mello & Graesser, 2012) by engaging in effective self-regulation strategies. In contrast, if a student is expressing frustration, perhaps they need further assistance to help them regulate this frustration, as to not lead them to a state of boredom (D’Mello & Graesser, 2012). As such, examining students’ facial expressions of emotions can also inform us of how students are learning with CALSs.

A third data type is log-file data, which include time-stamped, behavioral indicators of students’ interactions with a system—i.e., the entire sequence of learner-initiated and system-initiated actions with the CALS at the millisecond level listing what, when, and for how long the learner and system initiated all behavioral actions (e.g., what page they click on, when, and for how long). When extracting log-file data, there are many activities or behaviors that can be extracted, which can usually provide contextual information about what the student is doing. For example, if a student has just completed a sub-goal, the log-file will indicate the time at which they completed the previous sub-goal and began the new sub-goal; and since completing a sub-goal requires students to complete a 10-item quiz, the log file will also indicate the start and end
time of taking the quiz, indicating the duration of that quiz, as well as specific responses to each quiz question, allowing the researcher to know how the student performed on the quiz. As another example, log files can be used to assess gameplay activity, such that we can determine which location the student is in (Bradbury et al., 2017), which activities they engage in for each location (e.g., read 4 books, talked to 1 non-player character), and the outcome of these activities (e.g., number of concept matrix attempts, food item testing positive or negative for pathogenic substances). As such, log files allow us to calculate frequencies, durations, specific responses to, and all other behavioral variables that researchers might want to determine what students were doing during learning (Taub, Azevedo, Bradbury, Millar, & Lester, 2017). These data are typically indicative of cognitive strategy use (e.g., taking notes, summarizing) and can be used to infer metacognitive processes (unless they are embedded self-report measures of metacognitive judgements) during learning with CALSs, as affective data are typically measured using sensing devices.

Finally, another measure of affective data includes electrodermal activity (EDA), which is a measure of students’ physiological arousal, (e.g., heart rate variability, skin conductance, galvanic skin response) measured with wearable biosensors (e.g., Shimmer GSR+, Empatica E4). EDA is an umbrella term used to describe any type of electrical phenomena present in the skin (Boucsein, 2012), and can be processed to measure skin conductance response (SCR) events derived from phasic EDA. Specifically, EDA is comprised of two components: tonic and phasic. Tonic EDA, commonly referred to as skin conductance level is activity that varies over time in the absence of environmental effects or external stimuli (Hardy, Wiebe, Grafsgaard, Boyer, & Lester, 2013). In contrast, phasic activity consists of fast increases in skin conductance forming a peak with a slower decline back to baseline and has been shown to respond to the presence of
specific environmental stimuli (e.g., images, smells, cognitive processes, affective states; Benedek & Kaernbach, 2010, Boucsein, 2012). Research examining students’ emotions using physiological measures has primarily examined the size, duration, and frequency of students’ SCR events as they engage in different learning tasks (Arguel, Lockyer, Lipp, Lordge, & Kennedy, 2016; Brawner, & Gonzalez, 2016; Harley, Bouchet, Hussain, Azevedo, & Calvo, 2015). Furthermore, research has also identified that SCR events can predict students’ perceptions of the relevance of information in retrieval tasks (Barral, Eugster, Toutsalo, Spapé, Kosunen, Ravaja, & Kaski, 2015), or affective responses to specific learning events (e.g., negative feedback; Hardy et al., 2013). Thus, we can use these data to examine the number of SCR events as students read content, take quizzes, make metacognitive judgments, and engage in hypothesis testing during learning. We can also examine how the number of SCR events impact students’ abilities to deem pages as relevant to their sub-goals, or how students’ emotions impact how they respond to reading pages vs. completing quizzes vs. talking to non-player characters vs. testing items during learning with different CALSs. Overall, these different types of data can be used to assess different types of cognitive, affective, and metacognitive processes during learning with CALS. In the next section, we demonstrate specific examples of how we have assessed CAM processes using different data types.

**Using Multimodal Multichannel Data to Measure CAM Processes**

The abovementioned data can be useful for assessing self-regulation during learning with CALSs. Specifically, we can take different approaches to analyzing data for CAM processes. In the following section, we provide specific examples of using different data channels to investigate different CAM processes during learning with different types of CALSs.
Figure 2 illustrates how a student can engage in several cognitive, affective, and metacognitive processes during learning with two different CALSs (MetaTutor, CRYSTAL ISLAND) and how we can use different data channels to measure these different CAM processes. For cognitive processes (under orange header in Figure 2), we can use MetaTutor to examine how students are taking notes. To do so, we can use log-file data to examine the frequency of taking notes (Taub, Azevedo, Bouchet, & Khosravifar, 2014), the duration of how long they spent taking notes (overall, and for each separate instance), and the content of those notes (e.g., Trevors et al., 2014). In addition, we can use eye-tracking data to examine how students spent fixating on the notes (e.g., Taub & Azevedo, 2016). With CRYSTAL ISLAND, we can investigate students’ cognitive learning strategy use based on how they solved the concept matrix after reading a book, how they examined their log-file behavior indicating how long they spent reading a book and completing its concept matrix, and how many attempts they made to complete the matrix correctly. Additionally, we can examine their eye-tracking behavior to reveal the proportion of time they spent fixating on the text and the concept matrices, and how this impacted their performance (Taub et al., 2017). As such, we can use both log files and eye-tracking data to examine how students engage in cognitive self-regulation processes during learning with different CALSs.

To assess affective processes (under blue header in Figure 2), we can use both facial expressions of emotions and EDA data. During learning with MetaTutor, we can measure how students’ facial expressions of emotions impacted their proportional learning gain and we can assess how students’ emotions varied in different locations in CRYSTAL ISLAND (Bradbury et al., 2017). We can also investigate heart rate variability and skin conductance response in terms of how they changed over different content pages about the circulatory system in MetaTutor, and
across different activities (e.g., reading books vs. talking to non-player characters vs. testing food items) during gameplay with CRYSTAL ISLAND. Finally, we can use all these data to create emotion heat maps, which indicate the prevalence of emotions over time, where more prevalent emotions are in red, and the least prevalent emotions are in blue. As such, we can use these emotion heat maps to indicate emotions while interacting with all of these CALSs, where we can indicate if, for example, a student had high levels of confusion at the beginning of the learning session, followed by high levels of frustration as time progressed, indicating the student might have been on a path towards disengagement (D’Mello & Graesser, 2012). Therefore, using both video and EDA data can be used to assess students’ emotions during learning with different types of CALSs.

For assessing metacognitive processes during learning with CALSs (green header in Figure 2), we can use multimodal multichannel data to assess different types of metacognitive processes. First, while a student is reading text content and inspecting diagrams during learning with MetaTutor, we can use eye-tracking data to indicate gaze patterns between the text and diagram, representing metacognitive monitoring to coordinate informational sources. We can also investigate hypothesis-testing behavior, and how students monitored the food items they were testing during gameplay with CRYSTAL ISLAND. The log files can indicate when students used the scanning device, as well as how long students spent testing food items with the scanner and which food items students were testing for (Taub et al., in press). As another example, we can use log-file data from MetaTutor, such that we can examine students’ metacognitive judgments, including which judgment (e.g., content evaluation) they were making and what their responses were (e.g., for concurrent verbalizations). We can also examine students engaging in metacognitive processes by using the SRL palette in MetaTutor, such that students can choose to
engage in a judgment of learning, feeling of knowing, and content evaluation. We can use eye-tracking data to examine the time spent fixating on the SRL palette (Taub & Azevedo, 2016), as well as use the log files to examine when in the learning session students selected to use these processes, which process they selected, how long they spent engaging in the process, and their response (e.g., this page is relevant to my sub-goal). Thus, using log-file and eye-tracking data can provide many details on how students are using metacognitive processes during learning with CALSs.

Given these examples, it is evident that multimodal multichannel data can provide great insight into how learners are using cognitive, affective, and metacognitive processes as they learn with different types of CALSs. However, when selecting which data to use, researchers should always use theoretically-driven and empirically-based justification when selecting data channels. In the next section, we will provide a list of questions that researchers should ask themselves prior to choosing their use of data types.

Selecting the Appropriate Multimodal Multichannel Data

A common misconception about using multimodal multichannel data is that the more data types you have, the more information you will obtain; however, this is not the case. There are many considerations a researcher has to make regarding their research questions, hypotheses, and learning context that will inform them of which data channels are the most appropriate, instead of using data channels that cannot address their research goals. In addition, there are theoretical dichotomies (see details in Table 1) that can also impact the selection of different data channels. In this section, we will discuss these considerations in more detail to help researchers select which type of data they should collect to conduct appropriate analyses to investigate CAM processes during learning with CALSs.
**Contextual Considerations.** A researcher’s research questions and hypotheses will impact the data they should use because only certain data channels will be able to address the questions and provide evidence supporting or not supporting the hypotheses. For example, if a research question were to investigate how students’ quiz scores differ based on the condition they were assigned to in MetaTutor, it would be appropriate to use log files to determine their quiz scores, while it would not be appropriate to use eye-tracking data or affect data because those data would not provide information on students’ quiz scores. Alternatively, if the research question were to examine how students’ emotions differ between the two conditions, it would be appropriate to use emotion data, but not eye-tracking or log-file data because they would not allow the researcher to test the hypotheses they formed regarding this research question. Thus, in these cases, we demonstrate how using all data types will not always be useful.

The specific learning context will also play an important role in determining which data channels to collect. Depending on where students are learning (e.g., lab, classroom, home) and what type of CALS they are using, some types of data will be more appropriate than others. For example, if students are learning with a CALS in a classroom and researchers are conducting a classroom study, collecting log-file data and EDA data are feasible to collect because they do not have additional limitations in terms of data collection, such that when collecting video and eye-tracking data, this requires a controlled setting where students must sit still and avoid movement and touching their face. These limitations make it difficult to collect these data channels in the classroom, which is therefore another factor that influences choosing data types. In contrast, a controlled laboratory study does not pose any limitations in terms of moving around the lab; however, when using the eye tracker, it is important to remember that the lighting has to be at an appropriate brightness so as not to interfere with pupil size measurement. As such, based on the
design setting, there are different types of limitations and actions to be taken to avoid data collection issues, and depending on the research setting, this can impact which data to collect, and how.

**Theoretical Dichotomies.** In addition to the contextual considerations, selecting an accurate data type will also depend on factors mentioned in the theoretical dichotomies table from the previous section (see Table 1), where the appropriate data type(s) will depend on the type of variable and the specific CAM process being analyzed. From Table 1, it is evident that for cognitive and metacognitive processes, log-file data can always be used to answer research questions related to the use of these processes, whether it examines quality vs. quantity, macro vs. micro level, time vs. event, or discrete vs. co-occurring variables. However, with more specific and detailed questions examining variables considering different dichotomies, using more data channels will be necessary. For example, a simple question for cognitive processes (based on the frequency level in Table 1), can ask if taking more notes leads to better learning outcomes, while a simple question for metacognitive processes (based on the correlation subcategory in Table 1) can ask if more accurate judgments of learning are associated with more accurate content evaluations. Both of these questions would require the use of log-file data only, thus making them simpler to address. In contrast, a more complex question for cognitive processes (based on the interplay subcategory in Table 1) might ask at what point does individual variation in micro-level (e.g., take notes, summarize) processes lead us to examine macro-level (e.g., learning strategies) processes, while a metacognitive question (based on the causation subcategory in Table 1) might ask which specific cues cause students to engage in judgments of learning. Both of these research questions are far more complex, and as such, to answer them
would require the use of many data channels, such as eye tracking, facial expressions, and EDA, in addition to log files.

For affective processes, a simpler question will also require fewer data channels; however, these are facial expressions and/or EDA data. For example, based on a valence subcategory as described in Table 1, this type of question would ask if experiencing a positive emotion (e.g., joy) influences expressing negative emotions (e.g., confusion), which would require using facial expression and/or EDA data only. In contrast, a more complex question (based on the latency subcategory in Table 1) might ask about the average amount of time occurring between encountering an impasse and expressing confusion, which would require the use of log files and eye tracking with facial expression and/or EDA data. Thus, as with cognitive and metacognitive processes, the appropriate data channels to answer questions related to affect at different complexity levels will differ, where more complex questions require the use of more data channels.

Overall, based on all CAM processes, it appears that to include more contextual information in a research question (e.g., what is a student’s level of frustration while making inaccurate metacognitive judgments, or what contributes to the period of time between opening a new content page and deciding to take notes), the researcher needs to include more than only facial expression or log-file data, respectively, but to include multiple data channels that inform the researcher of the data they need to answer their research questions. Based on these examples, if the researcher wants to investigate a student’s level of frustration during learning they would only need to include facial expression and/or EDA data; or to investigate taking notes alone this would only need log-file data. However, once the researcher wants to know when, where, or how (when making a metacognitive judgment or selecting a new page, on page 6 of the CASL or
when in the infirmary of CRYSTAL ISLAND, or the level of accuracy of this judgment), the
researcher needs to include other data channels as well. Thus, to investigate CAM processes
during learning with CALSs, there are many different types of research questions to pose and
depending on the level of specificity of the research questions, using more than one data channel
will be necessary, and will also depend on the specific details of the research question.

**Implications of Using Multimodal Multichannel Data**

Based on all these examples, it is evident that there are many different ways we can analyze data using the same data channel (e.g., frequency vs. duration), and how we can use multiple different data channels to investigate one research question. In addition to selecting which data channel to use to answer these research questions, one more distinction must be made between using data to assess the *quality* vs. the *quantity* of an SRL process, which may require the use of the same data channel, however in different ways. For example, assessing the use of cognitive strategies, such as taking notes, can be measured by the frequency of use of the note-taking feature, which assesses the quantity of taking notes. However, this does not guarantee that all of these notes contain correct information, and so in order to assess the quality of these notes, further processing of the notes, such as using Latent Semantic Analysis, should be applied. This still involves using the log-file data, it just requires further processing. In contrast, some research questions can involve using more than one data channel to assess the quantity of a CAM process. This can include using both log-file and eye-tracking data to assess if a longer duration engaging in a metacognitive judgment increases its accuracy. However, to assess the quality of the metacognitive judgment, we will not be able to use eye-tracking data and will only assess the accuracy based on the log-files, which indicate the student’s response to the judgment. Thus, we infer from the eye tracking that the student is fixating on the text and diagram to make the
judgment, but to assess the actual quality of that judgment, we would not be able to infer this from the eye-tracking data. As such, not only is it important to differentiate between investigating the quantity and quality of the CAM process, it is also necessary to make this distinction of which data channels will generate these different types of data.

In addition to discussing how these dichotomies impact selecting the appropriate data channel to help us respond to our research questions, selecting the appropriate data channel will also inform us of which specific type of data analysis to conduct. It is beyond the scope of this chapter to address the different types of statistical tests and analysis techniques (e.g., machine learning), however we do note that the selection of a data analysis technique will also depend on the research questions, as well as data channel and variable type (e.g., frequency vs. duration; one time point vs. multiple time points). For example, if a researcher wants to address the total time students were confused during learning, they can use a total duration variable, or proportion of time spent confused compared to the total time spent expressing all emotions (i.e., total session time), and can compare the difference in expressing confusion between conditions using traditional inferential statistics, such as a t-test or ANOVA. However, if the researcher wants to narrow down when specific instances of durations of confusion occurred during learning, and how each instance impacted overall learning, a multi-level modeling approach would be more appropriate. As such, not only is it important to make choices about your research questions for selecting the appropriate data channel, but doing so will also be important toward conducting data analysis, demonstrating the close relationship between all the steps toward conducting a research study.

Overall, using multimodal multichannel data is a beneficial approach to measuring self-regulation because it allows us to pose more sophisticated research questions, allowing us to get
a more in-depth depiction of how students are engaging in different types of cognitive, affective, and metacognitive processes during learning with CALSs, which can ultimately lead the way for designing CALSs that can foster self-regulation and adapt to how students self-regulate during learning.

**Future Directions and Conclusions**

The area of SRL and CALSs has reached an unprecedented moment in history of human learning where we can strategically use our models, frameworks, and theories of SRL to intelligently design CALSs with recent technological advances to measure CAM processes as well as support and foster SRL. Below are a few examples of work that could advance our measurement and understanding of CAM SRL processes with CALSs.

One area of research focuses on the measurement and analysis of learners’ CAM processes in real-time while solving complex STEM and non-STEM problems that require the use of computational thinking (CT) skills (e.g., abstractions and pattern generalization) using virtual reality (VR) both as a research and learning tool. For example, measuring metacognition by converging online trace multimodal multichannel process data (e.g., eye tracking, facial expressions, EDA, log files, concurrent verbal utterances, gestures, body movements), self-reports, and learning outcomes. The results of this type of study can be used to develop a VR metacognition training module (also delivered by VR) that can reify metacognitive monitoring and control processes (e.g., replaying a teacher’s gaze behavior and associated verbal explanations of her metacognitive monitoring and control processes while solving a complex STEM problem). This example naturally leads to dealing with issues of domain specificity, training and transfer of CAM SRL skills, and the development of SRL competencies.
Another area stems from the growing interdisciplinary research documenting adolescents’ inability to accurately monitor and regulate their CAM processes during learning about complex STEM material, which negatively impacts their academic achievement, persistence, engagement, and interest in STEM and in pursuing STEM careers. One promising approach has been to use VR or other immersive systems (e.g., augmented reality, tangible computing) that can intrinsically foster and sustain motivational (e.g., situational interest, persistence, task value, intrinsic motivation, self-efficacy) and affective (e.g., joy, curiosity, pride) processes while allowing learners to focus on cognitive and metacognitive processes necessary to perform accurately when solving complex STEM problems. More specifically, we assume that fostering and sustaining motivational processes and affective states will allow learners to (1) practice a myriad of cognitive learning strategies (e.g., hypothesizing, summarizing, making inferences) and (2) practice making accurate metacognitive monitoring judgments (e.g., judgments of learning, feeling of knowing, monitoring progress towards goals, content evaluation).

A much needed area of research involves the embodiment of theoretical assumptions of models, theories, and frameworks described earlier as affordances of CALSs. While many systems reviewed in this chapter have accomplished this challenge, there are still constructs, mechanisms, assumptions that have yet to be embodied in CALSs. For example, most models of metacognition and SRL make assumptions about internal standards used to monitor and control. However, none of the existing CALSs model learners’ internal standards nor do they make them visible to learners and teachers so they can potentially use them as feedback to monitor the accuracy of their metacognitive processes. Internal standards can be illustrated as part of the
CALSs interface and be augmented with other data visualizations (Azevedo et al., 2017) that provide the learner with information regarding the accuracy of their self-regulatory skills.

Lastly, there are other constructs and mechanisms of SRL (e.g., calibration, adaptivity, self-modification, dysregulation, efficacy) that can be measured through the use of multimodal, multichannel data during learning and problem-solving with CALS. Such endeavours, as illustrated in this chapter along with the key questions presented in Table 1, will significantly advance the science of learning with CALSs. This effort will involve interdisciplinary researchers including psychologists, educational researchers, and instructional designers working alongside statisticians, engineers, computer scientists, and AI researchers to detect, model, track, and foster CAM SRL processes with CALSs. An interdisciplinary approach will advance (1) our models and theories (e.g., test and extend current assumptions by including micro-level and valence in Greene & Azevedo, 2009 processes in Winne’s, 2017 model), and (2) provide intelligent, adaptive scaffolding that addresses learners’ real-time self-regulatory needs during learning with CALSs.

In conclusion, the future of self-regulation and CALS is very bright as recent technological advances allow for unsurpassed approaches to design sophisticated systems that can measure and foster SRL. In this chapter, we presented a brief historical review of research on SRL and CALSs, discussed the role, measurement, and support of CAM processes during learning and problem solving with CALSs, presented ways of measuring and fostering SRL with CALSs, discussed conceptual and theoretical issues derived from several models, frameworks, and theories of SRL, discussed several dichotomies related to CAM and the challenges they pose for the measurement and support of SRL with CALSs, and presented challenges and future directions that that need to be addressed by researchers.
Table 1

Sample Questions and Data Sources Based on Conceptual Dichotomies about Cognitive, Affective, and Metacognitive (CAM) Self-Regulated Learning Processes

<table>
<thead>
<tr>
<th>Dichotomies</th>
<th>Cognitive Questions</th>
<th>Affective Questions</th>
<th>Metacognitive Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quality vs. Quantity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>Does taking more notes contribute to better learning outcomes?</td>
<td>LF</td>
<td>Does a higher frequency of confusion negatively impact learning?</td>
</tr>
<tr>
<td>Duration</td>
<td>Do longer periods of summarizing lead to higher quality summaries?</td>
<td>LF</td>
<td>Do extended periods (i.e., states) of emotion contribute to higher arousal?</td>
</tr>
<tr>
<td>Latency</td>
<td>What contributes to the period of time between a learner selecting a new page of content and deciding to take notes?</td>
<td>LF, FE, ET, EDA</td>
<td>What is the average amount of time between a learner encountering an impasse or discrepancy and experiencing confusion?</td>
</tr>
<tr>
<td><strong>Macro vs. Micro</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Granularity</td>
<td>How does the complexity of the text influence the learner’s ability to make an accurate summary?</td>
<td>LF, ET</td>
<td>Do facial expressions of confusion that last for only one second impact learning outcomes?</td>
</tr>
<tr>
<td>Table 1 (continued)</td>
<td>Valence</td>
<td>What causes a learner to accurately summarize relevant instructional materials?</td>
<td>LF, ET</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------</td>
<td>--------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Interplay</td>
<td>At what point does individual variation in the deployment of valence-level process use lead us to examine micro-level processes necessitate use and subsequently to examine macro-level processes?</td>
<td>LF, ET, FE, EDA</td>
<td>How does a learner’s negatively valenced mood influence their ability to regulate their confusion?</td>
</tr>
<tr>
<td><strong>Time vs. Event</strong></td>
<td>Thresholding</td>
<td>How long should a learner attempt to activate their prior knowledge?</td>
<td>LF</td>
</tr>
<tr>
<td><strong>Discrete vs. Co-occurrence</strong></td>
<td>Parallel</td>
<td>Can cognitive strategy use (e.g., note-taking and summarizing) co-occur or is co—occurrence only possible across cognitive, affective, and metacognitive processes (e.g., note-taking while metacognitive monitoring and</td>
<td>LF</td>
</tr>
</tbody>
</table>
Table 1 (continued)  
facially expressing joy)?

<table>
<thead>
<tr>
<th>Serial</th>
<th>Should learners engage in note taking on every page?</th>
<th>LF</th>
<th>At what point does confusion transition into frustration?</th>
<th>FE, EDA</th>
<th>Should previewing learning content always be followed by a content evaluation?</th>
<th>LF, ET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causation</td>
<td>When reading text, what causes the learner to inspect the corresponding diagram?</td>
<td>LF, FE, ET, EDA</td>
<td>To experience frustration, do learners need to first experience confusion?</td>
<td>FE, EDA</td>
<td>Which specific cues cause learners to engage in judgments of learning?</td>
<td>LT, FE, ET, EDA</td>
</tr>
<tr>
<td>Correlation</td>
<td>Are individual differences correlated with increased note taking?</td>
<td>LF, FE, ET, EDA</td>
<td>Do higher levels of confusion lead to higher levels of frustration?</td>
<td>FE, EDA</td>
<td>Are more accurate judgments of learning associated with more accurate content evaluations?</td>
<td>LF</td>
</tr>
</tbody>
</table>

**Note.** LF = log-file, FE = facial expressions of emotion, ET = eye-tracking, EDA = electrodermal activity. We acknowledge that this table does not include all possible multimodal multichannel SRL process data (e.g., concurrent think aloud data). Please note that this table does not include a comprehensive list of all possible questions, issues, and multimodal multichannel SRL processes data.
Figure 1. Screenshot of MetaTutor
Figure 2. Examples of specific uses of multimodal multichannel data to investigate CAM processes with different CALSs.
References


CHAPTER 3: IDENTIFYING HOW METACOGNITIVE JUDGMENTS INFLUENCE STUDENT PERFORMANCE DURING LEARNING WITH METATUTORIVH

Identifying How Metacognitive Judgments Influence Student Performance During Learning with MetaTutorIVH

Nicholas V. Mudrick\(^1\), Robert Sawyer\(^1\), Megan J. Price\(^1\), James Lester\(^2\), Candice Roberts\(^1\), and Roger Azevedo\(^1\)

\(^1\) Department of Psychology, North Carolina State University, Raleigh, NC, USA (nvmudric, mjprice3, razevedo@ncsu.edu)
\(^2\) Department of Computer Science, North Carolina State University, Raleigh, NC, USA (rssawyer, lester@ncsu.edu)
\(^3\) Natural Sciences Department, Wake Technical Community College, Raleigh, NC, USA

Abstract. Students need to accurately monitor and judge the difficulty of learning materials to effectively self-regulate their learning with advanced learning technologies such as intelligent tutoring systems (ITSs), including MetaTutorIVH. However, there is a paucity of research examining how metacognitive monitoring processes such as ease of learning (EOL) judgments can be used to provide adaptive scaffolding and predict student performance during learning ITSs. In this paper, we report on a study investigating how students’ EOL judgments can influence their performance and significantly predict their learning outcomes during learning with MetaTutorIVH, an ITS for human physiology. The results have important design implications for incorporating different types of metacognitive judgements in student models to support metacognition and foster learning of complex ITSs.

Keywords: Metacognitive monitoring · Ease of learning judgments · Performance · Predictive modeling · Intelligent tutoring systems

1 Introduction

The use of advanced learning technologies, such as intelligent tutoring systems (ITS), for learning is becoming ubiquitous and students learning with these environments are expected to act autonomously and self-regulate their learning [1]. Furthermore, several ITSs, such as MetaTutor, Betty’s Brain, AutoTutor, and nSTUDY have been developed to detect, support, and foster students’ metacognition and self-regulated learning (SRL) [2]. As such, it is imperative for students to accurately monitor the difficulty of the material they learn with these environments. These judgments regarding how difficult content will be to learn, or ease of learning (EOL) judgments, are important contributors to academic achievement and learning with ITSs as they can influence attention, time, strategy use, and effort allocation to the learning content [1–3]. Past research investigating these judgments has assessed their accuracy and confidence (i.e., difference between their judged and demonstrated levels of performance) and has found in most
cases, students are largely inaccurate and overconfident [1–4]. However, this research has primarily investigated EOLs in laboratory-based contexts (i.e., paired-associates learning) that may not reflect how students make these judgments in educational contexts. Thus, determining the utility of EOL judgments in predicting learning outcomes provides a valuable research contribution to using metacognitive judgment features in student modeling during learning with ITSs.

1.1 Related Work

EOL judgments can contribute to successful learning outcomes because they are made early on in the learning process and can influence study behavior and allocation of effort during self-regulated learning [1–3]. Research examining these judgments in laboratory settings suggests that EOLs are poor to moderate predictors of learning outcomes because they are prospective judgments that are made without seeing the instructional materials. However, much of this literature examines how EOL judgments can influence learning with simple tasks (i.e., paired-associates learning) and as such, most of the factors that influence the accuracy of these judgments are relatively micro-level (e.g., semantic relatedness between word pairs, fluency of perceptual processing, [2]). Although the results from laboratory studies have provided evidence regarding the inability of EOLs to predict performance, there is a potential for using more multimedia instructional materials (e.g., text and diagrams), to identify how EOLs can be embedded within ITSs to predict student performance during complex learning.

To explore this potential, we investigate how micro-level metacognitive judgments (EOL judgments) that are made after examining a science question can influence student performance. EOLs may force a learner to activate and successfully retrieve relevant prior knowledge (if any), plan and generate sub-goals for learning based on successful retrieval of relevant prior knowledge, to prepare to actively and accurately monitoring and regulate their cognition (e.g., select learning strategies), motivation (e.g., expect to persist given the complexity of the materials and lack of relevant prior knowledge), and emotions (e.g., engage in cognitive reappraisal when experiencing frequent and prolonged bouts of confusion and frustration). These are possible cognitive, affective, motivation, and metacognitive SRL processes that can be activated prior to an EOL that may fluctuate once the multimedia material is made available by an ITS and can therefore substantially influence and predict students’ performance, especially in those with low prior knowledge such as the ones who participated in our study (see Sect. 2.1).

2 Current Study

To assess the relationship between students’ EOL judgments and student performance during learning with MetaTutorIVH, we investigated the following research questions:

1. Are there differences in performance when students judge the content as easier to learn vs. when students judge the content as more difficult to learn?
2. Are there differences in performance for when students judge the content as easier to learn than it actually is vs. when students judge the content as more difficult to learn than it actually is?
3. Can students’ ease of learning (EOL) judgments predict their performance?

2.1 Participants

48 undergraduate students (77% female) enrolled at a large mid-Atlantic North American University participated in this study. Their ages ranged from 18 to 30 ($M = 20.30$, $SD = 2.35$). Scores from the 18-item pre-test assessing their prior knowledge of the science domains covered in the study revealed that students had low to moderate prior knowledge of the science content ($M = 11.20$ [62.22%], $SD = 1.48$ [8.22%]). Students were monetarily compensated up to $30 dollars for their participation.

2.2 MetaTutorIVH

The study was conducted with MetaTutorIVH, where students made several metacognitive judgments, inspected multimedia materials, and answered a series of multiple-choice questions regarding 9 different human physiological systems (e.g., circulatory, endocrine, nervous, etc.). MetaTutorIVH was designed to examine the influence of an intelligent virtual human’s (IVH) behavior on students’ cognitive learning strategies, metacognitive judgments, and emotions during learning about complex biology topics (Fig. 1). The environment consists of an IVH, text passages and diagrams about human body systems, and metacognitive judgment prompts. For this study, the IVH’s behavior

![Fig. 1. Screenshot of MetaTutorIVH’s main interface illustrating the science questions, multimedia content, and intelligent virtual human (IVH).](image-url)
consisted of specific facial expressions that were dependent on the relevancy of the content (see Research Design in Sect. 2.3).

Students interacted with MetaTutorIVH over 18 counter-balanced, randomized, self-paced trials that consisted of science questions, metacognitive judgment prompts, multimedia science content, and multiple-choice questions. The 18 trials were identical in format. In each trial, students were first presented with a science question regarding a particular body system on a separate slide before being presented with the multimedia science content. An example science question was, "Please explain the process by which we inhale more oxygen molecules than we exhale." After viewing the science question, students were then asked to submit an EOL judgment by answering, "How easy do you think it will be to learn the information needed to answer this question?" Students submitted their responses on a 0–100% scale, increasing in increments of 1%.

Following the submission of their EOL judgment, students were presented with a content page containing the text passage, diagram depicting the concept described in the text, the IVH, and the science question that was presented previously. After 30 s, students were prompted to judge the relevancy of the text and diagram to the science question they needed to answer by responding on a 3-point Likert-style scale. After students made their text and diagram content evaluations, the IVH facially expressed a congruent, incongruent, or neutral facial expression depending on the content relevance. Students returned to reading the text and inspecting the diagram. After they were finished viewing the content, students were required to answer the science question they were presented previously by choosing a correct response from 4 options. After they submitted their answer, students were prompted to make a judgment assessing their confidence in their chosen answer. After they submitted their judgment, students were prompted to justify their answer by typing a response into a text box (to ensure they had not skimmed the material and guessed). They were then asked to make another confidence judgment based on their justification. This procedure was repeated for the remainder of the 18 trials following the experimental session.

2.3 Research Design

This study used a 3 × 3 × 2 within-subjects design resulting in 18 trials. The first factor was content relevancy, which referred to the relationship between the level of description of the concept presented in the text/diagram to the science question asked. Students interacted with 3 levels of relevancy: high relevancy (where both the text and diagram were fully relevant to the science question asked), low text relevancy (where the diagram was fully relevant to the science question, but the text depicted the science topic in more general terms), and low diagram relevancy (where the text was fully relevant to the science question, but the diagram depicted the science topic more generally). Despite the presence of less relevant text or diagrams, the content still contained the information needed to correctly answer the question. The second factor was the congruency of the IVH's facial expressions such that the IVH facially expressed a congruent (i.e. the facial expression matched the relevancy of the content, joy for fully relevant content, confusion for less relevant), incongruent (i.e. the facial expression did not match the relevancy of the content, confusion for fully relevant content, and joy for less relevant content), or a
neutral (included as a comparison) based on the relevancy of the content. For example, if the text was only somewhat relevant to the content, the IVH facially expressed confusion to be congruent with the content, or joy to be incongruent with the content. The third factor was whether the science question asked about a standard function or malfunction of a particular body system. For example, a function question about the human respiratory system was, “Please explain the process by which we inhale more oxygen molecules than we exhale,” while a malfunction question was, “Please explain how, in cystic fibrosis patients, a missing chloride channel alters diffusion of oxygen in the respiratory system.”

2.4 Materials and Procedure

The study materials and equipment included the following: demographics questionnaire, pretest made of 18 4-option multiple-choice questions used to assess prior knowledge of the body systems described within the environment. Students’ EOL judgments and answers to the 4-option multiple-choice questions were automatically collected by MetaTutorIVH.

Students completed an informed consent form and then asked to complete a computerized demographic questionnaire and an 18-item science content pretest assessing their basic biology content knowledge. After students completed the pretest, they completed the 18 previously described trials with MetaTutorIVH. The average interaction with MetaTutorIVH lasted approximately 1 h ($M = 58.5 \text{ min, } SD = 20.40 \text{ min}$).

2.5 Data Sources and Preprocessing

Traditionally, metacognitive confidence judgments like EOL judgments have been examined by calculating their absolute and relative accuracies. Absolute accuracy is defined as the difference between the judgment and performance, while relative accuracy is the relationship between a set of judgments and students’ performance scores. While absolute accuracy identifies how precise a student is in their metacognitive judgments, relative accuracy assesses the correspondence between judgments and overall performance [1,4]. However, to assess relative and absolute accuracies, judgments and performance measures must be on the same continuous or ordinal scale, which may not reflect the types of assessments used in academic settings (e.g., multiple-choice questions, ordinally graded essays, etc.) in this study. As such, a standardization process on students’ EOL judgments was performed.

Although the experimental manipulations in this study included changes to the content relevancy to the science question, as well as the behavior of the IVH, students did not have access to the text, diagram or agent when making their EOL judgment for a given trial. An F-test on the significance of the coefficients for content relevancy, IVH facial expression congruence, and type of science question in a multiple linear regression for content difficulty (see below) indicated none of the coefficients were significantly different from zero ($F(5, 12) = 0.63, p = .68$). As such, results suggest these factors did not significantly affect the difficulty of the content.
Standardizing EOL Judgments by Student. We standardized students’ EOL judgment scores by student. The resulting standardized measures of a student’s EOL represent how easy the student believed the multiple-choice problem to be relative to other problems the student had rated. For example, a standardized value of 1.5 meant that the student believed this problem was 1.5 standard deviations easier than their average EOL judgment. A negative value indicated that the student believed the problem was harder than their average EOL judgment.

This type of standardization is important because students could have had different interpretations of a problem being “easy” or “difficult”, which is reflected in their ratings on the 0–100 scale. For example, on the original 0–100 scale, one student’s average EOL judgment was 17.7, while another’s was 76.4. These two students demonstrate different interpretations of the 0–100 scale for their EOL judgments. This is important, considering that the first student, who thought the content was substantially more difficult to learn according to their raw EOL judgment values, correctly answered 12 of 18 (66.7%) questions while the second only correctly answered 9 of 18 (50%) questions. As such, this standardization allowed comparisons between students since the standardized values represent a student’s relative EOL.

Assessing Content Difficulty. Each trial included a multiple-choice question with four possible responses. For each multiple-choice question, there was one correct response, two partially correct responses, and one incorrect response. We calculated a weighted sum of a question’s ease from the total number of students who answered the multiple-choice questions according to these three response categories (i.e., correct = 1, partially correct = 0.5, incorrect = 0). Higher weighted sums indicate that more students answered correctly or partially correct (indicating easier to learn content), while lower weighted sums indicate that more students answered incorrectly or partially correct (indicating more difficult to learn content). These weighted sums were calculated for each of the 18 trials, where each trial’s weighted sum ranged from 0 to the total number of students who responded to those questions.

Determining EOL Judgment Error. Standardizing students’ EOL judgments allowed us to assess the accuracy of their EOL judgments against the measure of content difficulty. Specifically, we calculated accuracy in terms of students’ EOL judgment error, in contrast to traditional measures of relative and absolute accuracies. There were two error measures of interest: signed error and squared error.

We calculated the signed error as the difference between a student’s standardized EOL and the standardized problem difficulty (similar to the traditional measure of relative accuracy). Because the resulting value retained its positive or negative value, this allowed us to assess students’ EOL judgments in comparison to the actual difficulty of the problem. For example, if the signed error was positive, the student thought the problem was easier than it actually was, whereas if the signed error was negative, the student thought the problem was more difficult than it actually was.

The squared error was calculated as the signed error value squared. This allowed us to calculate of the magnitude of error in a student’s EOL judgment the problem difficulty
(similar to traditional measures of absolute accuracy). This measure was also motivated by noting that the sum of squares is a common regression error function.

We calculated both of these error measures on a student-trial basis such that our units of analysis were students' EOL judgments per question, as opposed to aggregating these judgments by student. As such, each time a student made a judgment, the error was calculated, for a total of 18 signed and squared errors per student (i.e., 18 trials $\times$ 48 students $= 864$ signed error values, and $18 \times 48 = 864$ squared error values).

3 Results

3.1 Are There Differences in Performance When Students Judge the Content as Easier to Learn vs. When Students Judge the Content as More Difficult to Learn?

The student-standardized EOL judgments were used to determine whether there were differences in performance per trial for students who judged the content as easier to learn vs. those who judged the content as more difficult to learn. More specifically, a one-way ANOVA was conducted to compare the effect of positive (judged to be easier to learn than average) or negative (judged to be harder to learn than average) student standardized EOL judgments on performance. There were 438 (50.6%) trials in which students judged the content to be easier to learn than their average EOL for the content and 426 (49.4%) trials in which students judged the content to be more difficult to learn than their average EOL judgment. There was a significant effect of students' EOL judgments on performance at the $\alpha = 0.05$ level for the positive (easier to learn) or negative (harder to learn) judgment groupings [$F(1, 862) = 4.02, p = 0.045$]. A post-hoc analysis using a Welch's (unequal variance) two sample t-test indicated that the performance for students who judged the content as easier to learn than average ($M = 0.74$, $SD = 0.34$) was significantly lower than students who judged the content as more difficult to learn than average ($M = 0.78$, $SD = 0.32$). Furthermore, a significant negative correlation ($r = -0.10$, $p = 0.003$) was observed between the standardized student EOLs and performance. This demonstrated that as students judged the content to be easier to learn relative to other content, the worse they performed on the multiple-choice questions. As such, results indicated that students who judged the content as more difficult to learn achieved higher performance on the multiple-choice questions.

3.2 Are There Differences in Performance for When Students Judge the Content as Easier to Learn Than It Actually Is vs. When Students Judge the Content as More Difficult to Learn Than It Actually Is?

The signed errors of EOL judgments were used to determine if there were differences in performance per trial among students who judged the content as easier to learn than it actually was vs. performance among students who judged the content as more difficult to learn than it actually was. A one-way ANOVA was conducted to compare the effect of positive (i.e., judged as easier to learn than it was) or negative (i.e., judged as harder to learn than it was) signed error judgement groupings on performance. There were 422
(48.8%) trials in which students judged the content as easier to learn than it actually was, and 442 (51.2%) trials in which students believed the content to be harder to learn than it actually was. There was a significant effect of the signed judgment error on performance at the α = 0.05 level for the positive and negative judgment groupings \[ F(1, 862) = 101.3, p < 0.001 \]. A post-hoc analysis using a Welch’s two sample t-test indicated that performance for students who judged the content as easier than it actually was (\( M = 0.65, SD = 0.36 \)) was significantly lower than students who judged the content as more difficult to learn than it actually was (\( M = 0.86, SD = 0.26 \)). Additionally, a significant negative correlation between the signed error and performance was observed (\( r = -0.37, p < .001 \)), such that as students who judged the content to be easier to learn than it actually was, the worse they performed on the multiple-choice question. As such, results indicated that students who judged the content as more difficult to learn than it actually was achieved higher performance on the multiple-choice questions (Table 1).

Table 1. Summary of multiple choice score by EOL grouping, which summarizes research questions 1 (top row) and 2 (bottom row).

<table>
<thead>
<tr>
<th></th>
<th>Positive group</th>
<th>Negative group</th>
<th>Overall correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n )</td>
<td>( M )</td>
<td>( SD )</td>
</tr>
<tr>
<td>Standardized EOL</td>
<td>438</td>
<td>0.74</td>
<td>0.34</td>
</tr>
<tr>
<td>Signed EOL Error</td>
<td>422</td>
<td>0.65</td>
<td>0.36</td>
</tr>
</tbody>
</table>

\( *p < .05 \). \( **p < .001 \).

3.3 Can We Use Ease of Learning (EOL) Judgments to Predict Student Performance?

For these analyses, we treated the performance prediction as a binary classification problem. This was done by treating partially correct answers as incorrect, resulting in a 61.5% performance correctness rate serving as the majority class (question answered correctly) baseline. We computed the 10-fold cross validation accuracy using a multi-layer perceptron model with layers of 15 and 5 rectified linear units implemented from the sklearn.neural_network package in Python [6] and using standardized EOL judgments, difficulty direction correct, and whether the standardized ease of learning is positive as features. The difficulty direction correct is a binary variable indicating whether the standardized EOL judgment and standardized ease of content have the same sign. This is an indicator of whether or not the student correctly assessed the content of being more or less difficult to learn than the average content and was correctly performed on 47.5% of trials. These predictors were chosen because they are almost independent of the difficulty of the content and reflect the usefulness of the student’s EOLs judgments without explicitly including the content’s difficulty. The three predictors used as input features, including the content’s difficulty used to calculate the errors, are calculated and standardized using only data from the training fold. The average accuracy across the 10-folds was 71.7%, which is a significant improvement over the majority class baseline of 61.5% (\( t = 5.88, p < 0.001 \)). This accuracy improvement indicates that these features
based primarily upon student EOL judgments are useful in predicting student performance.

4 Discussion

The study investigated how students' EOL judgments can influence and be predictive of performance. The results from these analyses significantly augment our understanding of how students' metacognitive judgments can be used to model performance during learning with ITSs and have implications designing future ITSs that emphasize the role of metacognition during complex learning.

Results from our first research question indicated that students who judged the content as being harder to learn outperformed students who judged the content as being easier to learn on their multiple-choice responses. These findings are compounded by results from research question 2, which indicated that students who judged the content as being harder to learn than it actually was, significantly outperformed students who judged the content to be easier to learn than it actually was. The significantly lower performance of students who judged the content to be easy suggests that students' overconfidence for these questions deleteriously impacted their performance. Contrary to published literature on EOLs, it is possible that these students did not accurately monitor their emerging understanding, select the appropriate cognitive strategies, and allocate sufficient effort necessary to successfully understand, and learn the multimedia materials, leading to poor performance [1–4]. Alternatively, students who had judged the content as being harder to learn achieved superior performance by accurately monitoring their understanding and selecting the appropriate strategies, allocating more effort than they needed to successfully understand the content. These results significantly extend previous research on EOL judgments by integrating different types of measures to indicate relative and absolute judgment accuracies (i.e., standardizing both EOL judgments and problem difficulty). Traditionally, research has addressed these measures of accuracy separately by calculating the absolute accuracy index for absolute accuracy and Goodman-Kruskal correlations for relative accuracy (see [5]).

Lastly, results from our third research question demonstrated the utility of EOLs in predicting student performance. Results indicated that including EOL judgments and their accuracy as predictors in a multi-layer perceptron model demonstrated statistically significant predictions of performance 71.7% of the time, improving prediction by 10.2% (relative improvement of 16.6% over the baseline). As such, it is possible that the context of learning with educationally relevant materials may facilitate more accurate EOL judgments than the other contexts where they have been examined. Furthermore, limited research has investigated including metacognitive judgments as features to use while building accurate student models. Therefore, our results provide evidence that prospective metacognitive judgments can provide ITSs with important student-based performance information with which the system can use to identify the accuracy of metacognitive monitoring processes during learning and intervene accordingly based on a sophisticated student model.
From practical and design perspectives, incorporating EOL-like features into ITSs is straightforward and imposes little burden on students. An ITS need only take one continuous input from a brief preview of a future problem, without showing the student any content, to predict their performance for learning that content. Results from our analyses indicated that students generally spent little time making these judgments ($M = 4.5$ s, $SD = 3.3$ s) relative to their overall time interacting with MetaTutorIVH in our study (average of 2.3% of the total time). As such, integrating this as a feature in ITSs can potentially provide the system pertinent performance information. For example, students could provide an EOL after being presented with their next topic. The ITS uses that students’ EOL judgment to model their performance and intervenes based on this prediction. Specifically, the IVH or other artificial agent (knowledgeable of the difficulty of the content) could provide scaffolding to the student in the form of suggestions to re-evaluate their metacognitive judgment, slow down and pay attention to the upcoming content, etc. Alternatively, the IVH could then prompt the student to engage in context-appropriate cognitive learning strategies by having the students summarize the presented material, make inferences about the content, and integrate the information in the text and diagrams to facilitate better conceptual understanding and deeper learning.

Acknowledgements. This research was supported by funding from the National Science Foundation (DRL#1431552). The authors would also like to thank members from the SMART Lab and IntelliMedia Group for their contributions to this project.

References

CHAPTER 4: TOWARD AFFECT SENSITIVE VIRTUAL HUMAN TUTORS: THE INFLUENCE OF FACIAL EXPRESSIONS ON LEARNING AND EMOTIONS

2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)

Toward Affect-Sensitive Virtual Human Tutors: The Influence of Facial Expressions on Learning and Emotion

Nicholas V. Mudrick, Michelle Taub, Roger Azevedo
Department of Psychology
North Carolina State University
Raleigh, NC, USA
{nmudrick, mtaub, receved}@ncsu.edu

Jonathan Rowe, James Lester
Department of Computer Science
North Carolina State University
Raleigh, NC, USA
{growe, lester}@ncsu.edu

Abstract—Affective support can play a central role in adaptive learning environments. Although virtual human tutors hold significant promise for providing affective support, a key open question is how a tutor’s facial expressions can influence learners’ performance. In this paper, we report on a study to examine the influence of a human tutor agent’s facial expressions on learners’ performance and emotions during learning. Results from the study suggest that learners’ performance is significantly better when a human tutor agent facially expresses emotions that are congruent with the content relevancy. Results also suggest that learners facially express significantly more confusion when the human tutor agent provides incongruent facial expressions. These results can inform the design of virtual humans as pedagogical agents can inform the design of virtual humans as pedagogical agents and designing intelligent learner-agent interactions.

1. Introduction

Research on learning with intelligent tutoring systems (ITSs) has consistently emphasized the importance of emotion, which has been shown to be critical during learning [1]. One component of ITSs that has been found to significantly influence learners’ emotions are virtual human tutors [2]. Virtual human tutors are embodied pedagogical agents (PaAs) that emulate the realistic appearance and behavior of humans. They leverage a broad range of verbal and non-verbal behaviors to scaffold student learning during naturalistic one-on-one tutoring interactions. Facial expressions have a key role in virtual human tutors’ repertoire of pedagogical behaviors. For example, a furrowed eyebrow could signal that a virtual human tutor is confused by a student’s recent problem-solving step, or a half smile could provide encouraging feedback to a student who has just overcome an impasse. However, there is a gap within the literature regarding how these facial expressions can influence learners’ performance and emotions during learning. As such, there exists a need to systematically examine specific components within human-PA interactions to assess how these expressions can influence learners’ performance and emotions during learning.

In this paper, we examine the influence of a human tutor agent’s facial expressions on performance and emotions during learning with biology content with varying levels of relevancy with the MetaTutor Learning Environment. Within the MetaTutor Learning Environment, we manipulated the relevancy of the content to the question learners needed to answer so they would look towards the human tutor agent for information regarding their own assessments and performance. The human tutor agent then provided facial expressions congruent (i.e., joy) or incongruent (i.e., confusion) with the content relevancy or a neutral facial expression (included as a comparison). For the human tutor agent, we used video recordings of a human's expressions to investigate the effects of realistic facial displays of emotion on student learning. As there exists little research to draw on, using video recordings allowed us to assess the specific influence of realistic facial expressions on learners’ performance and emotions to inform the design of a virtual human tutor’s behavior. Investigating these effects is a formative step in the design of an intelligently interactive, autonomous virtual human tutor with naturalistic verbal and non-verbal behaviors. As such, our study contributes to an evidence base informing the design of intelligent virtual human tutors and their behaviors to facilitate more intelligent learner-agent interactions and influencing learning.

1.1. Related Work

The behavioral realism of realistic virtual humans has prompted growing interest in their use as PaAs. Leveraging their realism can help us understand students’ behaviors during learning, as virtual human behaviors can be finely controlled, repeated, and closely resemble actual humans [3-4]. Additionally, with their behavioral realism, more naturalistic virtual agents have been argued to facilitate increased perceptions of their believability [4]. Research examining the emotional expressions of virtual agents has identified that facial expressions and gestures achieve high recognition rates [5] and higher ratings of authenticity in their expressions than human actors [6]. Furthermore, research has demonstrated the timing of these expressions to significantly moderate individuals’ perceptions of these expressions during interactions [7]. As such, research has identified that facial expressions of a virtual human alone are
enough to influence participants’ perceptions of their emotions [3]. However, there has been limited research on the impact of virtual humans’ non-verbal expressions of emotion in ITSS, including their effects on student learning and emotions. We address this gap in the literature by examining the influence of a human tutor agent’s facial expressions on learner’s emotions and performance.

Research on the system dynamics of ITSS stems from the assumption that PAs can facilitate increased learning and performance by supporting and instructing learners through the provision of scaffolding and feedback [8]. Within the literature on PAs, their design, purpose, and feedback varies substantially, and they have been shown to influence learners’ emotions and learning differently. For example, research has shown confusion induced through written dialogues between the learner and the PAs resulted in students achieving higher levels of comprehension [9]. Other research has revealed that the frequency of spoken prompts and feedback provided by PAs can lead to learner frustration and boredom, which negatively influenced learning [10]. Alternatively, research has identified that the content of feedback can influence learning differently [11]. Lastly, research has also suggested that PA feedback in the form of facial expression and gesture can lead learners to perform better on transfer tests (i.e., embodiment effect; [12]).

The majority of this research has used cartoon-like PAs and may not be representative of the influence of realistic intelligent virtual agents that can leverage real-time 3D rendering and naturalistic animations to deliver non-verbal feedback [13]. Furthermore, none of these studies specify the content, facial expressions, or gestures by the PAs. Given the variety of the different populations, roles agents played, learning contexts, and content, it is difficult to generalize their effects. As such, systematically examining the influence of specific components of the human-agent interaction (e.g., the influence of specific facial expressions of emotions) on learners’ performance is needed. However, given the lack of research regarding the specific influence of facial expressions on learners’ performance and emotions, more evidence is needed to inform the design of future virtual human tutors. Therefore, the study reported in this paper uses realistic, natural facial expressions provided by a human tutor as a first step toward developing virtual human tutors for more intelligent interactions during learning with ITSS.

1.2. Theoretical Frameworks

To assess the influence of content relevancy and human tutor facial expression congruency on learners’ performance and emotions during learning, multiple theoretical frameworks are needed. Specifically, we use the Emotions as Social Information (EASI) model [14] and Dynamics of Affective States Model [15] to explain the influence of the human tutor agent’s facial expressions on the emotions and learning outcomes examined in this study. The EASI model suggests that facial expressions of emotion serve to provide information about a situation (e.g., congruence between the relevancy of a text or diagram to a question), but this information is dependent on how accurately the learner processes the expression (i.e., allocating effort to understand why the expresser is facially expressing an emotion) [14].

The Dynamics of Affective States Model argues that when learning is interrupted by an impasse, contradiction, or unexpected event, students experience confusion [15]. If their confusion is left unresolved, students will transition into frustration and then boredom. However, if students resolve their confusion through effortful cognitive processing, they will transition back into a state of deep learning.

Based on these models, we hypothesize that learners monitor and perceive the congruent information provided by a human tutor agent’s facial expressions, and learners use that information as affirmation of their own content judgments, thereby influencing their performance on the multiple-choice questions. Additionally, we hypothesize that when learners are presented with incongruent or neutral facial expressions, they experience confusion based on the contradictory information to the relevancy of the content.

2. Methods

2.1. The MetaTutor Learning Environment

We conducted a study that involved students interacting with the MetaTutor Learning Environment to answer a series of multiple-choice questions regarding biology concepts. The MetaTutor Learning Environment was designed to examine the influence of a virtual human tutor agent on students’ cognitive learning strategies, metacognitive judgments, and emotional responses during learning. The virtual human tutor agent resided in the upper-right section of MetaTutor’s user interface while students read text passages and diagrams about human body systems (Figure 1). In this study, the virtual human tutor agent was realized with recorded video segments of a human tutor agent. These video recordings were produced with a professional actress who facially expressed joy, confusion, and neutral based on the strategic expression of specific action units (AUs) in the Facial Action Coding System (FACS; [16]). The tutor’s facial expressions were also validated by a trained and certified FACS coder. The facial expressions of joy, confusion, and neutral were selected in collaboration with a biology professor who indicated these emotions arise during one-on-one tutoring sessions with college students, as well as the Dynamics of Affective States Model that argues these emotions play a critical role in student learning [15]. In this study, the tutor’s behavior consisted of facial expressions of specific emotions, which were dependent on the relevancy of biology content that was presented on-screen to students. For example, the tutor displayed a joyful facial expression when the on-screen content was fully relevant to the question that the student was attempting to answer, or a confused facial expression when the on-screen content was weakly relevant.

Students’ interactions with the MetaTutor Learning Environment consisted of 9 counter-balanced, linearly structured, self-paced trials that interleaved science questions, multimedia science content, and metacognitive
judgment prompts. The 9 trials had an identical format. In each trial, learners were first presented with a science question, and they were asked to submit an ease of learning judgment, “How easy do you think it will be to learn the information needed to answer this question?” on a 0 to 100% scale. Learners were next presented with a text passage, diagram, and the science question, as well as an idle video of the human tutor agent expressing a neutral facial expression. After 30 seconds learners were prompted to assess the relevancy of the text and the diagram: “Do you feel the text/diagram on this page is relevant to the question being asked?” Learners made two content evaluation judgments (CEs; [17]) on a Likert scale (ranging from 1-3) to the following statements: “The text/diagram is relevant”, “The text/diagram is somewhat relevant”, and “The text/diagram is not relevant.” Upon making their text and diagram CEs, the on-screen tutor expressed either a congruent, incongruent, or neutral facial expression based on the relevancy of the content. The tutor’s facial expression lasted 10 seconds from start to finish. Following the facial expression, the human tutor agent video returned to a looping neutral expression.

Learners were permitted to read the text and inspect the diagram at their own pace. After they were finished examining the content, learners were prompted to answer the science question by choosing the correct response from 4 options. After submitting their answer, learners were prompted to make a retrospective confidence judgment (RCJ) by answering, “How confident are you that the answer you provided is correct?” Learners made their judgment on a 50 to 100% scale (where a score of 50% would indicate learners believed they had a 50/50 chance of submitting the correct answer). After submitting their response, learners were prompted to justify their answer by typing their response into a text box to ensure that learners had not simply skimmed the material and guessed at their answer. Learners were then asked to make another RCJ based on their justification.

2.2. Participants

Forty-four (n = 44) undergraduate students enrolled at a large mid-Atlantic North American university participated in this study. Participants’ ages ranged from 18-24 (M = 20.10, SD = 1.61). Scores from the 18-item science pre-test indicated that participants had some prior knowledge of the science domains presented in the experiment (M = 11.56 [64.22%], SD = 2.04). Participants were monetarily compensated $30 for their participation.

2.3. Design

The study used a 3x3 within-subjects design resulting in 9 trials. Our study investigated the impact of two factors. The first factor was content relevancy, which refers to the relationship between a question and the text and/or diagram that accompanied the question. The second factor was human tutor agent’s facial expression congruence, which refers to the relationship between the human tutor agent’s facial expression and the relevancy of the content. Specifically, learners interacted with 3 levels of content relevancy to the question they needed to answer: high relevancy (i.e., the text and diagram were completely relevant), low text relevancy (i.e., the text contained information that was not required for answering the question, but the diagram was still fully relevant to the question asked), and low diagram relevancy (i.e., the diagram contained information that was not required for answering the question, but the text was still fully relevant to the question asked). For the three levels of congruency, the human tutor agent facially expressed different emotions depending on the relevancy of the content: congruent (i.e., the emotion matches the relevancy of the content), incongruent (i.e., the emotion does not match the relevancy of the content), and neutral (i.e., the lack of an emotional expression). Based on these manipulations, each learner completed 9 counter-balanced, randomized trials, with different combinations of content relevancy and human tutor agent facial expression congruency (see Table 1).

![Figure 1](image1.png)

Figure 1. Example screenshot of the question, text, diagram, and human tutor agent with a congruent facial expression of joy.

![Figure 2](image2.png)

Figure 2. Human tutor agent facial expressions of emotion: joy (left), confusion (middle), and neutral (right).

| TABLE I. HUMAN TUTOR AGENT FACIAL EXPRESSIONS BY STUDY CONDITION. |
|-----------------|-----------------|-----------------|
| Content Relevancy | Fully Relevant | Text Less Relevant | Diagram Less Relevant |
| Congruent | Joy | Confusion | Confusion |

186
2.4. Measures and Multimodal Setup

The materials and equipment used in this study consisted of the following: an SMI RED 250 eye tracker, a Logitech HD Pro webcam 920, and a Shimmer 3+ wireless bracelet to capture participants’ electrodermal activity (EDA). The SMI RED 250 eye tracker collected learners’ eye movements at 120 Hz, while the webcam recorded learners’ facial expressions of emotions at a rate of 30 Hz at 1080p. Participants’ EDA was logged with a Shimmer 3 that collected at a rate of 128 Hz. Lastly, learners’ performance was assessed with 4-item multiple-choice questions administered near the end of each trial.

2.5. Procedure

After participants entered the lab and completed an informed consent form, the eye tracker was calibrated by the researcher. Participants were asked to face a blank monitor screen and hold a neutral facial expression for 6 seconds to establish a baseline. Afterward, participants were asked to sit still for 5 minutes to establish a baseline for their EDA. After establishing these baselines, participants were asked to complete a computerized demographic questionnaire and an 18-item pretest that assessed their basic biology content knowledge. After participants completed the pretest, participants completed the 9 previously described trials in MetaTutor. The average interaction with the MetaTutor Learning Environment lasted approximately 60 minutes (M = 56.53 min, SD = 15.23).

2.6. Data Sources

1) Multiple-choice response accuracy.

Learner responses to multiple-choice questions were coded for accuracy. The coding scheme was created in collaboration with a biology professor. Fully correct responses were given a score of 1, partially correct responses were coded as .5 for containing relevant but not correct information, and incorrect responses were coded as 0. Each multiple-choice question had one correct answer, two partially correct answers, and one incorrect answer.

2) Facial expressions of emotion.

The videos of each learner’s facial expressions captured during the experimental session were analyzed using Attention Tool FACET developed by Motions [18]. FACET analyzed each video frame and classified 19 action units (AU) and nine emotions (i.e., joy, anger, disgust, fear, surprise, sadness, contempt, confusion, frustration, and neutral) and calculated evidence scores for each AU and emotion at 30 Hz. The evidence scores were formatted as probabilities, representing the log odds of the presence of a facial expression as coded by an expert human coder. Though each AU and emotion were analyzed, we focused primarily on learners’ facial expressions of joy, confusion, and neutral.

An absolute threshold of zero was used during analyses of the evidence score data; in other words, any negative evidence scores (i.e., the evidence of a facial expression not being present) were identified and set to zero. This threshold was used because we were primarily interested in the presence of specific emotions during the experiment, as opposed to the absence of specific emotions. The positive evidence score values were then smoothed using a symmetrical moving average filter configured with a fixed window of 11 data points. This filter is useful for analyzing temporal data (e.g., facial expressions over time) as it smooths out random outliers while retaining the step response of the signal. As such, each smoothed data point was calculated by averaging the raw data point with the previous five and next five raw data points in the original time-ordered data stream.

To investigate the influence of the human tutor agent’s facial expressions on learners’ emotions, we distinguished three time segments of learners’ FACET data: 1) the period before the human tutor agent’s facial expression (M = 84.08 s, SD = 43.31), 2) the period during the human tutor agent’s facial expression (10 s), and 3) the period following the tutor’s facial expression until the learner responded to the multiple-choice question (M = 6.61 s, SD = 11.43). Three mean scores were created for learners’ facial expressions of confusion, frustration, and joy for each level of relevancy and congruency. This produced 27 mean values per emotion (i.e., 3 relevancy conditions x 3 congruency conditions x 3 time segments).

3. Results

3.1. Do human tutor agent facial expression congruency and content relevancy interact to affect learners’ multiple choice question accuracy?

A two-way RM-ANOVA was conducted to determine the influence of congruency type and relevancy type on multiple choice response accuracy. There was a significant interaction between congruency and relevancy on response accuracy, $F(1, 172) = 17.88, p < .001$, $\eta^2_p = 1.00$. Therefore, simple main effects were conducted. Response accuracy was not significantly different for congruent (M = .72, SD = .38), incongruent (M = .56, SD = .16), or neutral (M = .69, SD = .38) facial expressions for content with high relevancy. Congruence type significantly influenced response accuracy for low text relevant content (F(2, 86) = 25.41, $p < .001$, $\eta^2_p = .37$). Specifically, question responses were more accurate for congruent (M = .97, SD = .17) than incongruent (M = .71, SD = .10) and neutral facial expressions (M = .48, SD = .05).

<table>
<thead>
<tr>
<th>Incongruent</th>
<th>Confusion</th>
<th>Joy</th>
<th>Joy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>Neutral</td>
<td>Neutral</td>
<td>Neutral</td>
</tr>
</tbody>
</table>

1 Although eye-tracking and EDA data were collected, they were not analyzed for this paper.
Additionally, incongruent facial expressions coincided with higher question response accuracy than neutral facial expressions. Response accuracy was significantly different for low diagram relevance content. Mauchly’s test indicated that the assumption of sphericity had been violated ($\chi^2 = 11.17, p = .004$, as such Greenhouse-Geisser values are reported. Results revealed that congruency significantly influenced response accuracy, $F(1.62, 69.72) = 10.82, p < .001$), indicating that responses were more accurate for incongruent facial expressions ($M = .86, SD = .25$) and neutral facial expressions ($M = .78, SD = .05$) than congruent facial expressions ($M = .55, SD = .48$). There were no significant differences in response accuracy for incongruent and neutral expressions for low diagram relevance content.

3.2. Do congruency type, relevancy type, and timing interact to influence learners’ facial expressions of confusion, frustration, and joy?

A three-way RM-ANOVA was conducted to determine if congruency type, relevancy type, and timing interacted to influence student’s facial expressions of confusion. Results revealed a significant three-way interaction between all the independent variables. Mauchly’s test revealed that the assumption of sphericity had been violated ($\chi^2(35) = 85.06, p < .001$), as such Greenhouse-Geisser values are reported. $F(3.35, 229.92) = 2.23, p < .05, \eta^2_g = .74$. We did not find evidence of a simple two-way interaction between congruency and content relevancy prior to the human tutor agent’s facial expression ($F(3.06, 131.48) = .55, p > .05$), nor did we find a two-way interaction after the human tutor agent’s facial expression $F(2.86) = 2.82, p > .05$. There was a significant two-way interaction between congruency and content relevancy during the human tutor agent’s facial expression $F(3.38, 145.29) = 3.68, p < .05, \eta^2_g = .83$. There was a simple main effect of congruency for low text relevance content $F(2, 86) = 4.60, p < .05, \eta^2_g = .07$, but not for highly relevant or low diagram relevance content ($p’s > .05$). Pairwise comparisons revealed that learner confusion was highest during neutral expressions by the tutor ($M = .68, SD = .08$) and significantly different from learner confusion during congruent expressions by the tutor ($M = .49, SD = .08$). We did not find evidence of a significant difference in learner confusion between neutral tutor facial expressions and incongruent tutor facial expressions ($M = .61, SD = .09$). There was no significant difference in learner confusion between incongruent and congruent tutor facial expressions for low diagram relevant content.

Two additional three-way RM-ANOVAs were conducted to determine the influence of congruency type, relevancy type, and time on learners’ facial expressions of frustration and joy. Results revealed Mauchly’s test was significant and the assumption of sphericity was violated for both facial expressions of frustration, ($\chi^2 (35) = 130.29, p = .000$), and joy ($\chi^2 (35) = 111.26, p < .001$). Greenhouse-Geisser statistics revealed there was no significant interaction between the three independent variables on frustration ($F(4.70, 202.13) = 1.86, p = .108, \eta^2_g = .061$) or joy ($F(4.61, 197.94) = 0.64, p > .05, \eta^2_g = .22$).

4. Discussion and Future Directions

The study’s objective was to systematically examine the influence of tutor facial expression congruency, content relevancy, and timing on learner performance and emotion. The set of examined variables in this study can significantly augment current computational models of affect dynamics and intelligent human-agent interactions in complex learning situations with ITSs [6, 8, 9]. Additionally, our findings suggest that contextual congruence is important for models of emotion in virtual human tutors.

Results from our first research question indicated that learners’ performance was highest for low text relevant content when the human tutor agent expressed confusion, congruent with the relevancy of the content. These results were consistent with our hypothesis, which suggested that learners would use the information provided by the human tutor agent’s facial expression as affirmation of their own evaluation of the content, thereby influencing their performance [20]. Results from our second research question indicated that learner confusion was highest when the learner interacted with low text relevant content while the tutor expressed a neutral facial expression. We did not find evidence of a significant relationship between content relevancy and tutor facial expression congruency on learner joy or frustration. This may have been due to the nature of the task: although there is evidence suggesting that joy is occasionally present during learning with ITSs, research also suggests that joy varies considerably between learning tasks [1]. The Dynamics of Affective States Model suggests that frustration occurs after periods of unresolved confusion. The trial-by-trial nature of the learning task with MetaTutor, may not have afforded the opportunity to experience lengthy periods of confusion [15]. The results on student confusion were consistent with our hypotheses, which suggested that the contradiction between the information-neutral facial expression and content relevancy would increase learners’ feelings of confusion. Lastly, as results indicated that learners expressed confusion during the human tutor agent’s facial expressions, these findings support prior research suggesting that the timing of affective expressions significantly influences human-virtual agent interactions [7].

This study examined a human tutor agent’s facial expressions of emotion congruent with multimedia biology content. However, the results have important implications for contextual congruency of virtual tutor emotion expressions in other contexts, such as with the learners’ affective states or the tutor’s own internal state. For example, the tutor’s facial expression could mimic a learner’s expressions of emotion (i.e., facially expresses confusion as the learner does) to facilitate an increase in the learner’s awareness of their own emotions. Alternatively, a virtual human tutor could express emotions representative of their own appraisal of the learner’s actions (e.g., confusion when a learner repeatedly makes poor notes about irrelevant content). Additionally, as
our results indicated the importance of timing of the facial expressions, designing future affect-sensitive virtual human tutors will need to be sensitive to the timing of their non-verbal feedback by accounting for how these expressions influence learners’ own emotions. For example, designers will need to identify appropriate instances during which facial expressions may have deleterious influences on learning, such as causing a learner excessive confusion. Alternatively, identifying instances where leveraging the emotional influence of a specific facial expression will be needed, but may be challenging given current research in affective computing. For example, if the learner facially expresses confusion, the virtual human tutor can intervene with a facial expression of concern to alleviate or reframe the learner. The timely modeling and computational demands required for such precise intervention is likely to challenge yet advance current systems. Lastly, our results indicating information-neutral facial expressions also increased learners’ confusion emphasize the importance of providing learners’ informational feedback in the form of facial expressions. Promising future directions include investigating affect-sensitive virtual tutor agents that facially express emotions that are dependent on different contexts and time and are self-modifying based on a human’s affective reactions to the agents’ expressions thus supporting intelligent interactions and leading to optimal learning and performance.

Investigating the impact of emotive facial expressions by virtual human tutors raises several challenges: it requires creating believable expressions of emotion that are relevant to learning, it calls for granular sensor-rich measurement of learner affect (i.e., facial expressions), and it demands consideration of the temporal dynamics of emotion among virtual humans as well as learners. In this study, we have leveraged videos of a trained human actor to systematically examine the influence of a human tutor agent’s facial expressions on learner performance and emotions in the MetaTutor Learning Environment. Our results imply that contextual and temporal congruence are important features for the development of future virtual human tutors.

Acknowledgments

This research is supported by the National Science Foundation (NSF DRL 1431532). The authors thank the members of the SMART Lab and the IntelliMedia Group.

References


CHAPTER 5: METAMENTOR: AN INTERACTIVE SYSTEM THAT USES VISUALIZATIONS OF STUDENTS’ REAL-TIME COGNITIVE, AFFECTIVE, METACOGNITIVE, AND MOTIVATIONAL SELF-REGULATORY PROCESSES TO STUDY HUMAN TUTORS’ DECISION MAKING

Abstract

In this chapter, we argue that the new generation of intelligent tutoring systems (ITSs) will need to detect, model, and respond to all of a student’s cognitive, affective, metacognitive, and motivational (CAMM) processes in real time to successfully foster their self-regulated learning (SRL) with these systems. Additionally, we argue that multimodal, multichannel process data (i.e., log files, eye tracking, facial expressions of emotion, and electrodermal activity) is needed to identify behavioral signatures of a student’s real-time CAMM SRL processes during learning. As such, we have designed a new system called MetaMentor to investigate how a student’s CAMM SRL processes unfold and interact in real-time. In this chapter, we outline the development of MetaMentor by presenting a brief review of traditional ITS research, our past research with MetaTutor, and the theoretical underpinnings of the design of MetaMentor. We discuss how we can use our findings from research with MetaMentor to design intelligent, adaptive, CAMM-sensitive ITSs, and how our findings can be applied to other advanced learning technologies to augment conceptual and theoretical frameworks of CAMM SRL processes.

Keywords: cognition, affect, metacognition, motivation, self-regulated learning, intelligent tutoring systems, multimodal, multichannel process data, data visualizations

Introduction

Intelligent Tutoring Systems (ITSs) are computer-based systems designed to detect, model, foster, and adapt to a student’s declarative, procedural, and conditional knowledge and skills during learning of complex topics (Aleven & Koedinger, 2016). Traditionally, ITSs have been designed to emulate effective tutoring strategies that an expert human tutor uses during one-
to-one human tutoring (Kulik & Fletcher, 2016; Ma, Adesope, Nesbit, & Lie, 2014; Steenbergen-Hu & Cooper, 2014; VanLehn, 2011). By emulating the instructional strategies of an expert human tutor, ITSs are able to detect and respond to a student’s learning behaviors (e.g., study allocation, accuracy of metacognitive judgments, effectiveness of cognitive strategies, etc.) (Azevedo, Mudrick, Taub & Bradbury, in press; Azevedo, Taub, & Mudrick, 2018), provide individualized feedback and scaffolding on a student’s performance (Graesser, Rus, & Hu, 2017), and contribute to their increased learning outcomes across many different domains (e.g., biology, chemistry, math, research methods, medicine, physics, military, etc.) (Azevedo, et al., 2018, in press; Graesser, et al., 2017; Sottilare, Graesser, Hu, & Goodwin, 2017; VanLehn, 2011).

Research with ITSs also aims to understand how these systems can impact a student’s self-regulated learning (SRL), as effective SRL can significantly enhance their learning outcomes (Aleven & Koedinger, 2016; Azevedo et al., 2018, in press; Lester, Lobene, Mott, & Rowe, 2014). SRL involves the student playing an active role during learning by accurately monitoring and controlling their cognitive, affective, metacognitive, and motivational (CAMM) processes (Azevedo et al., in press; Schunk & Greene, 2018; Winne, 2018; Winne & Hadwin, 1998, 2008). However, research indicates that the average student does not always engage in monitoring and controlling of their CAMM processes and therefore does not effectively self-regulate their learning with ITSs (Azevedo et al., in press; Schunk & Greene, 2018). As such, researchers have begun to use multimodal, multichannel process measures (eye tracking, facial expressions of emotions, log files, concurrent verbalizations, and electrodermal activity) to detect and understand how a student’s CAMM processes impact their SRL with ITSs (Azevedo et al., 2018, in press; Taub & Azevedo, 2018; Taub, Mudrick, & Azevedo, 2018).
While this extensive body of work has improved our understanding of key CAMM SRL processes, it has not yet addressed how these processes can unfold in real-time during learning with ITSs. Traditionally, researchers who use multimodal, multichannel process measures primarily rely on post-hoc analyses to understand the impact of specific CAMM processes on a student’s ability to engage in effective SRL, as well as their learning outcomes with ITSs (Azevedo et al., 2018, in press; Taub et al., 2018). Furthermore, some of the systems are designed using predefined time- and event-based production rules that prompt and scaffold specific CAMM SRL processes during learning (e.g., MetaTutor, Betty’s Brain, nStudy) (Azevedo, et al., 2013; Biswas, Segedy, & Bunchongchit, 2016; Winne, 2017). Many of these systems are able to detect how a student’s CAMM processes unfold and interact during learning with ITSs but are unable to use this information in real-time to foster all of a student’s CAMM SRL processes and domain learning. Additionally, many contemporary ITSs are designed to scaffold certain CAMM processes (e.g., cognitive strategies and metacognitive processes vs. learner-center emotions) and do not comprehensively address the dynamic interplay of all of a student’s CAMM SRL processes during learning. Therefore, we argue that the new generation of ITSs will need to detect, model, and respond to all of a student’s CAMM processes in real-time to successfully foster their SRL with ITSs.

We have developed a new system called MetaMentor to investigate how a student’s CAMM SRL processes unfold in real-time during learning with ITSs. MetaMentor combines the previous tradition of designing ITSs to emulate the effective instructional strategies of an expert human tutor with recent research on the importance of using CAMM SRL processes during learning with ITSs. Specifically, MetaMentor relies on the ability of an expert human tutor to detect, understand, interpret, and act on complex visual representations of a student’s CAMM
SRL processes abstracted from multimodal, multichannel process data (eye tracking, facial expressions of emotions, and log files) as a student interacts with the non-adaptive system. We are still in the initial stages of our research with MetaMentor, so we have not begun conducting empirical studies. However, our goal is to build MetaMentor as an adaptive ITS capable of responding to a student’s real-time CAMM processes to foster their SRL. Furthermore, we also aim to use the system to train human tutors to correctly detect, infer, understand, and adapt to the student’s SRL, based on data visualizations of their real-time CAMM processes.

This chapter outlines the development of MetaMentor by presenting a brief review of ITS research, our past research with MetaTutor and the importance of CAMM SRL process data, and the theoretical underpinnings that contributed to the MetaMentor design. We conclude this chapter by presenting potential challenges with our future research on MetaMentor and discussing possible future applications of MetaMentor to identify how a student’s CAMM SRL processes unfold during learning with other advanced learning technologies (e.g., serious games, simulations, and augmented and virtual reality).

**Using Expert Human Tutors to Inform the Design of ITSs**

Designing ITSs to emulate the instructional strategies of an expert human tutor is based on empirical evidence demonstrating that individualized instruction and tailored feedback more effectively foster student learning when compared to other instructional methods (e.g., traditional classroom instruction) (Bloom, 1984; Chi, Siler, Jeong, & Yamauchi, 2001; Graesser et al., 2017). Research using this approach indicates that these ITSs can foster a student’s cognitive and metacognitive SRL processes (Azevedo et al., 2013, 2016, 2018, in press), and significantly improve their learning outcomes (Craig, Graesser, & Perez, 2018; Ma, Adesope, Nesbit, & Liu, 2014; Steenbergen-Hu & Cooper, 2014; VanLehn, 2011). As such, many ITSs have been
developed to emulate the strategies (e.g., assessing student understanding, delivering immediate feedback and adaptive scaffolding, etc.) an expert human tutor may use to foster student learning for different topics and domains (Ma et al., 2014; VanLehn, 2011). The following section briefly discusses the history of ITS researchers incorporating effective tutoring strategies to design intelligent systems.

Assessing student understanding. An expert human tutor must accurately monitor and assess a student’s understanding during learning, problem solving, and reasoning to provide tailored scaffolding and adaptive feedback to the student (Chi, Siler, & Jeong, 2004; Graesser et al., 2017). Research demonstrates that a tutor’s accurate assessments of the student’s understanding can significantly impact how effective the tutor is in fostering the student’s learning (see Chi, et al., 2001, 2004; Graesser et al., 2017). As such, research on a student’s learning using ITSs has aimed to design ITSs capable of accurately evaluating their understanding during learning. This research has used several different approaches, like model tracing, whereby ITSs track the student’s conceptual understanding by identifying the errors underlying their specific problem-solving steps (Anderson, Corbett, Koedinger, & Pelletier, 1995). Additionally, some ITSs use expectation and misconception tailoring by interpreting the dialogue between the student and the ITS to identify the student’s misconceptions about the content (Nye, Graesser, & Hu, 2017). Furthermore, other ITSs model and track the student’s understanding by constraining the amount of declarative knowledge concepts the student must learn before progressing them on to other learning activities (e.g., constraint-based modeling; Mitrovic, Martin, & Suraweera, 2007). Based on the development of an accurate student model, these ITSs then provide feedback to the student during specific problem-solving steps or learning activities.
Feedback, scaffolding, and modelling. ITSs have also been designed to emulate how an expert human tutor delivers scaffolding, feedback, and model specific learning strategies to the student (Graesser et al., 2017). For example, research identifies that verbal communication plays a central role during one-to-one human tutoring and can influence the effectiveness of an expert human tutor’s feedback (Graesser et al., 2017). Specifically, an expert human tutor typically structures a tutoring session by verbally prompting the student with questions, allowing the student to answer those questions, then providing feedback to support the student’s learning (Chi et al., 2001, 2004). To mimic these interactions, many ITSs are embedded with pedagogical agents (PAs) or intelligent virtual human (IVH) tutors that can foster the use of effective SRL processes (e.g., cognitive strategies and metacognitive judgments) and provide feedback on the student’s performance during learning (Craig & Schroeder, 2018; Johnson & Lester, 2016; Kim & Baylor, 2016; Schroeder, Adesope, & Ma, 2014; Taub, Martin, Azevedo, & Mudrick, 2016). Specifically, PAs and IVH tutors serve as external regulating agents and emulate the scaffolding and feedback an expert human tutor may provide during learning (Taub et al., 2016).

Research has also identified that an expert human tutor structures their tutoring sessions to successfully increase a student’s learning outcomes (Graesser et al., 2017). For example, the reciprocal teaching paradigm can effectively increase a student’s performance during learning (Belland, Walker, Kim, & Lefler, 2017). During reciprocal teaching, the expert human tutor models effective learning strategies by situating the specific learning strategy within the current content the student is examining. Then the student uses that strategy, and the expert human tutor offers feedback regarding the student’s effectiveness in enacting that strategy. For example, the expert human tutor can model how to create a concise and accurate summary of the content the student is currently reading (e.g., the function of the circulatory system), prompt the student to
write another summary about other (but related) content (e.g., the structure of the heart), and provide feedback on the conciseness of the student’s summaries. Some ITSs, like MetaTutor, embody this principle by providing the student examples of an accurate summary (e.g., summarizing the presented content). Then the embedded PAs provides feedback on the accuracy of the student’s summaries (Azevedo et al., 2013, 2016, 2018, in press).

As the approach to emulate expert human tutor scaffolding, feedback, etc. has historically been effective in designing ITSs to scaffold and support student learning with other ITSs, we also use this approach to understand how an expert human tutor can detect and respond to student’s CAMM processes to model and foster a student’s SRL in real-time. Additionally, we aim to combine our approach with our previous research with MetaTutor which has identified the importance of understanding CAMM SRL processes with multimodal, multichannel process data, that we will present next.

**Research with MetaTutor: The Impact of CAMM SRL Processes on Student Learning**

MetaTutor is a hypermedia-based ITS that encourages the student’s use of cognitive and metacognitive SRL processes while they complete sub-goals and an overall goal of learning as much as they can about the human circulatory system (Azevedo, Witherspoon, Chauncey, & Graesser, 2009). MetaTutor does this in two different ways. First, the interface’s design contains nine elements that each support planning, monitoring, and strategy use by allowing the student to self-initiate these processes. This includes a table of contents, timer, sub-goal progress bar, SRL palette, notebook, text content, corresponding diagram, and PAs (Taub, Mudrick, & Azevedo, 2018). Second, the embedded PAs provide external regulation by prompting the student to engage in cognitive and metacognitive SRL processes and then providing feedback on their performance. In addition, when we run our studies, we collect self-report data on the student’s
emotions and motivation, which allows us to examine CAMM processes and their impact on overall learning.

We have conducted several analyses of CAMM SRL data during learning with MetaTutor using process, product, and self-report data (e.g., Azevedo et al., 2016). Trevors, Duffy, and Azevedo (2014) used log-file data to examine note taking during learning with MetaTutor, and how levels of prior knowledge impacted the frequency of taking notes. Their results revealed that when students were provided with external regulation from the PAs, students with low prior knowledge took more notes than students with high prior knowledge. Taub and Azevedo (2018) used log-file and eye-tracking data to examine cognitive and metacognitive SRL processes by examining how students with different levels of prior knowledge fixated on different interface elements and by examining sequences of engaging in accurate and inaccurate cognitive and metacognitive processes. Their results revealed that students with high prior knowledge engaged in higher frequencies of fixating from the text content to other areas on the interface and engaged in more sequences containing accurate cognitive and metacognitive processes, whereas students with low prior knowledge had greater frequencies of sequences containing inaccurate metacognitive processes. Both results reveal how we can examine different cognitive and metacognitive SRL processes during learning with MetaTutor, and how different groups of students engage in these processes differently.

To examine affective processes, Jaques, Conati, Harley, and Azevedo (2014) investigated gaze transitions using eye-tracking data to predict different emotions during learning with MetaTutor. Using eye-movement transitions from one area of the screen to another, the authors were able to predict if students were bored or curious, which were defined as students’ levels of engagement during learning (Jaques et al., 2014). To examine motivation, Duffy and Azevedo
(2015) examined how students in different experimental conditions and with different goal orientations performed on cognitive and metacognitive SRL processes during learning with MetaTutor. Results revealed that students who received external regulating prompts and feedback from the PAs and who were performance-approach oriented had the highest sub-goal relevancy scores, while mastery-approach oriented students in the prompt and feedback condition had the lowest sub-goal relevancy scores. These studies demonstrate how affect and motivation do impact SRL and performance during learning with MetaTutor, and how we need to consider these processes when designing ITSs that foster SRL. Furthermore, this research emphasizes the need to use multimodal, multichannel process data to detect, understand, and interpret how CAMM processes can influence a student’s SRL with ITSs. Although these studies have substantially improved our understanding of these processes, our previous investigations of CAMM SRL processes with MetaTutor contain some limitations, which will be discussed below.

Limitations: Fostering CAMM SRL Processes During Learning with MetaTutor

The studies discussed above demonstrate how we can use multimodal, multichannel process data to examine CAMM SRL processes during learning with MetaTutor. However, these findings also reveal that though we were able to measure all CAMM processes, MetaTutor only detects and fosters the use of cognitive and metacognitive SRL processes via prompts and feedback from the PAs based on a series of preset time- (e.g., prompting the student after spending 30 seconds on a page) and activity-based (e.g., after the student engages in specific cognitive and metacognitive processes) production rules. Therefore, the system does not adaptively prompt a student to engage in emotion regulation strategies that could help them reduce high levels of frustration or other negative emotions (e.g., anger). Furthermore, the system also cannot address the student’s current motivational states in real-time. For example,
research has indicated that adaptively altering the content during learning can improve a student’s self-efficacy for the task (Graesser et al., 2017). However, MetaTutor is unable to detect and respond to a student’s motivational processes and therefore cannot adaptively change the content to improve their feelings of self-efficacy.

As the nature of the prompting relies on a set of time- and event-based production rules, MetaTutor is unable to adapt to a student’s CAMM processes in real-time and can therefore potentially dysregulate the student’s SRL by inaccurately prompting the student to engage in cognitive and metacognitive processes at the wrong time. For example, if a student with low prior knowledge spends a long time reading the content, the PA will prompt the student to engage in a content evaluation (i.e., a judgment to evaluate the relevancy of the information to the student’s current sub-goal) without considering whether the student actually needs help to determine if the content is relevant to their sub-goal or is actually confused by the content. It is possible that the student was spending a long time reading the content because they were already metacognitively monitoring the relevancy of the instructional content or were taking the time to resolve their confusion by re-reading the content, and the prompting from the PA may be disruptive for the student. This raises the issue that covert learning behaviors and CAMM SRL processes that are not detected during learning may lead to inappropriate scaffolding and feedback and therefore disrupt or interfere with a student’s learning. If the feedback was more individualized (and based on the student’s real-time CAMM processes), before having the student engage in this metacognitive process, the PA could ask the student if they need help to assess the relevancy of the content because they look confused. Therefore, if we develop pedagogical interventions (e.g., prompting) that are not based a limited set of predefined production rules, but instead are based on each of the student’s CAMM multimodal,
multichannel process data signatures, we can provide more individualized, comprehensive, and effective feedback to the student during learning.

Despite advances made in SRL and ITSs, we acknowledge that the limitations of our past work with MetaTutor to detect and adaptively respond to a student’s CAMM SRL processes are also mirrored in the general ITS community. For example, though some exceptions exist (e.g., Arroyo et al., 2014), many contemporary ITSs focus on fostering a student’s cognitive and metacognitive processes during learning rather than providing guidance or feedback on student emotions and motivation. Furthermore, research investigating the detection, measurement, and scaffolding of real-time motivational processes (e.g., goal orientation, interest, task value) within the ITS community is relatively limited (for exceptions see Arroyo et al., 2014; Cloude, Taub, & Azevedo, 2018; du Boulay & Luckin, 2016; Duffy & Azevedo, 2015) and has primarily examined how individual differences in a student’s motivational orientation can influence their overall learning outcomes with ITSs. Therefore, a more comprehensive approach to investigate how a student monitors and controls all of their CAMM processes during learning with ITSs is needed, which we present below through our description of our newly developed system, MetaMentor.

**MetaMentor: An Interactive System Developed to Identify and Foster a Student’s CAMM SRL Processes in Real Time**

Examining a student’s real-time CAMM SRL processes during learning with ITSs entails using multimodal, multichannel process data to understand the temporal unfolding and interaction of these processes during learning, problem solving, and reasoning. To date, limited research has sought to investigate a student’s real-time CAMM SRL processes with multimodal, multichannel data *during* learning with ITSs. Even less research has examined how an expert
human tutor can interpret, understand, infer, and act on these processes to foster a student’s learning in real-time. Furthermore, little research has attempted to computationally model system features and rules based on an expert human tutor’s understanding and inferences of these processes. Due to the complexity of examining these CAMM SRL processes, the limited research that has been conducted, and our aim to design a system capable of automatically adapting to a student’s CAMM SRL processes, we have developed MetaMentor, that is described below.

MetaMentor is a multifaceted computer-based system that presents complex instructional- and training-material to the student and shows visualizations of the student’s multimodal, multichannel process data (e.g., screen recording of human-machine interactions, eye-movement behaviors, facial expressions of emotions, real-time log files, performance data on quizzes and other embedded assessments, and metacognitive judgements) to an expert human tutor in real-time. Specifically, MetaMentor is (a) a learning tool designed to help the student self-regulate during their learning of complex topics, (b) a tutoring and training tool for the human tutor to identify, understand, and interpret multimodal, multichannel data signatures of the student’s CAMM SRL processes from (a) and pedagogically intervene in real-time, and (c) a research tool designed to investigate how the student reacts to individualized scaffolding and feedback based on (b) during learning and how the expert human tutor acts on complex visual representations of multimodal, multichannel data on the student’s CAMM SRL processes to provide adaptive support, modeling, scaffolding, and feedback.

Currently, MetaMentor is not an automatically adaptive ITSs, therefore it is designed with an expert human tutor in-the-loop that provides real-time feedback and instruction to the student through an embedded IVH (discussed below). This means that there are two different
interfaces of the application with different features presented to the student and the expert human tutor. The following section describes the theoretical frameworks underlying the design of the features embedded within the student and human tutor interfaces of MetaMentor.

**MetaMentor: Student Interface**

*Interface design.* The design of the student’s MetaMentor interface is based on theories of multimedia learning (Mayer, 2014), macro- and micro-level cognitive learning strategies and metacognitive monitoring processes (Greene & Azevedo, 2009), SRL (Winne, 2018; Winne & Hadwin, 1998, 2008), and extensive previous research conducted with MetaTutor (Azevedo et al., 2013, 2016; 2018; in press; Harley, Taub, Azevedo, & Bouchet, 2018; Taub & Azevedo, 2018). Specifically, the design and use of the multimedia content (i.e., text and diagrams) correspond to several multimedia learning principles suggested by the Cognitive Theory of Multimedia Learning (Mayer, 2014). For example, the placement of the text and diagrams within MetaMentor correspond to the temporal contiguity principle, which suggests that the simultaneous presentation of text and diagrams reduces extraneous cognitive load (Mayer, 2014; Mayer & Fiorella, 2014)\(^4\). The macro- and micro-level cognitive learning strategies and metacognitive monitoring processes embedded within the interface (i.e., the buttons to engage in these processes embedded in the interface, see description below) stem from extensive work conducted by Azevedo and colleagues (2009, 2011, 2013, 2016, 2018, in press) that details the importance of micro-level SRL processes during learning. The inclusion of the other interface

\(^4\) Although a full review of these principles is outside the scope of the current chapter, please see Taub et al., 2018 for an extensive review of the multimedia learning principles used to design MetaTutor.
features (e.g., sub-goal progress bars, self-paced and open content presentation), correspond to specific phases a student engages in during successful SRL (Winne & Hadwin, 1998, 2008).

**Interface description.** The student’s interface contains several features that allow the student to enact cognitive and metacognitive SRL processes during learning about the human circulatory system (see Figure 1). On the top left of the interface is a timer that shows the student how much time is left in the learning session. To the right of the timer are two sub-goal progress bars that indicate how much content the student has learned for each of the two sub-goals they have set. Both the timer and sub-goal progress bars are included for the student to monitor their progress of their learning goals (MPTG), which is a metacognitive judgment that allows the student to monitor and regulate how much time they allocate toward learning for each sub-goal (Greene & Azevedo, 2009).

The table of contents, which is located underneath the timer, displays the titles of 50 pages of static text and diagrams that depict the core components of the human circulatory system. By allowing the student a wide range of topics, some content presented may not be fully relevant to the student’s current sub-goal, which encourages the student to engage in another metacognitive judgment, content evaluation (CE). CEs are key monitoring judgments the student can make when they assess the relevancy of content to their current goal, and accurate CEs can influence the efficiency of a student’s effort during learning (Mudrick, Taub, & Azevedo, 2017). Additionally, the static text and diagrams are placed in the center of the student’s interface and

---

5 The basis for this design and using this specific content is derived from research detailing the effectiveness of previous experiments with MetaTutor (Azevedo et al., 2016, 2018, in press; Taub et al., 2014, 2016, 2018).
foster the cognitive learning strategy of coordination of informational sources (COIS) between the text and diagrams. This facilitates the integration of key concepts and increased understanding of content (Greene & Azevedo, 2009; Mayer, 2014).

Figure 1. Example screenshot of the student’s interface during learning with MetaMentor illustrating the timer (top left), overall learning goals for the session and self-set goal (top middle), intelligent virtual human (top right), SRL palette allowing student to behaviorally enact several planning, cognitive strategies, and metacognitive judgments (right side), multimedia content (middle center), and table of contents (left side).

On the right side of the interface is the SRL palette, which allows the student to engage in several cognitive and metacognitive learning processes shown to successfully increase learning (Azevedo et al., 2016, 2018, in press; Taub & Azevedo, 2018). The student can activate prior knowledge of their current sub-goal (PKA) by clicking “Telling what I already know about this content.” Specifically, this button prompts a text box pop-up where the student can type in what they already know about their current sub-goal. This can influence how the student directs their
attention and effort to content they may not know or understand (as opposed to reviewing content they have seen before).

Underneath the PKA button, the student can also engage in several metacognitive judgments. The first judgment is called a judgment of learning (JOL) where the student can monitor how well they think they understand the presented information. The second judgment is feeling of knowing (FOK) where the student can assess if they have already seen the information before. When the student clicks either the JOL or FOK buttons, they are prompted to make a judgment on a scale from 0-100 percent to rate their confidence in how well they understand the information they are viewing (for JOLs) or if they feel they have seen the content before (for FOKs). The student is then prompted to take a 3-item, 4-option multiple-choice quiz. After taking these quizzes, the student is shown their scores to assess the calibration of their metacognitive judgments (the difference between their metacognitive judgment and actual performance—i.e., accuracy) and to regulate their learning strategies accordingly. Under the FOK button is the CE button where the student can assess the relevancy of the text and diagram they are currently viewing to their current sub-goal on a 3-point Likert scale with the statements, “I feel that the text and diagram are relevant,” “I feel that the text and diagram are somewhat relevant,” and, “I feel that the text and diagram are not relevant.”

Beneath the metacognitive judgment buttons are the cognitive learning strategy buttons. Clicking the take notes (TN) button generates a pop-up text box where the student can take notes on the multimedia content, clicking on the summarizing button (SUMM) will prompt them to summarize the information presented on the screen in a pop-up text box, and clicking on the inference button will prompt the student to make an inference from the presented information.
The student can choose to review their notes, summaries, and inferences at any time during learning.

Last, on the top right side of the interface is an IVH that can express different learning-centered and basic emotions (i.e., confusion, frustration, boredom, joy, anger, described in more detail below). Embedded within the IVH’s display are pop-up messages provided by the tutor (see section below) that can be coupled with the IVH’s facial expressions to provide feedback and scaffolding on, as well as prompts for the student to engage in, specific CAMM SRL processes during learning.

**MetaMentor: Human Tutor Interface**

*Interface design.* The human tutor interface of MetaMentor contains multiple visualizations of the student’s temporal unfolding CAMM SRL processes derived from multimodal, multichannel data as they learn (see Figure 2). The interface design integrates several theoretical frameworks and emerging research on the use of multimodal, multichannel process data to detect and understand CAMM SRL processes during learning with ITSs (Azevedo et al., 2018, in press; Winne, 2017, 2018). Specifically, we developed the student’s screen mirror with the student’s eye-movement behavior to indicate where they are directing their attention to learning content. Research has indicated that eye movements traditionally indicate the moment-to-moment cognitive processes that occur during learning (Rayner, 2009). Furthermore, research on the use of eye-movement modeling examples (EMMEs) has demonstrated the effectiveness of these visualizations to direct another individual’s attention to relevant information (Jarodzka, Holmqvist, & Gruber, 2017). As such, these findings suggest that the expert human tutor will attend to and interpret the student’s cognitive processes through this visualization.
We also included a webcam video stream of the student’s face in the interface to indicate when the student expresses specific emotions during the learning session. Presenting the video stream of the student’s facial expressions stems from theories of human emotion, which indicate that facial expressions convey an individual’s underlying affective states, and this data channel has been used extensively to identify how a student’s emotions can influence their learning outcomes with an ITS (Baker et al., 2010; D’Mello, 2013; Ekman, 1973; Fernández-Dols & Russell, 2017). In addition to the video stream, we designed the tutor interface to play the student’s audio recordings captured during the learning session. Including the audio recording of the student’s verbalizations to the expert human tutor is based on extensive empirical research that has investigated CAMM SRL processes from students’ think-alouds during learning with ITSs (Azevedo et al., 2018, in press; Greene & Azevedo, 2009; Miele & Scholer, 2018).

Finally, research on the use of multimodal, multichannel CAMM SRL process data has traditionally relied on the use of log files to contextualize and understand other data channels (e.g., the student facially expressing more confusion when they opened a page containing irrelevant information) (see Azevedo et al., 2013, 2018, in press; Taub et al., 2016, 2018). However, most of these analyses are conducted post-hoc, after the learning session. Research investigating the interpretation of this data channel and its integration with others (e.g., eye tracking, facial expressions, audio recordings) in real-time has yet to be conducted. As such, the activity log was designed based on a real-time, temporally unfolding log file to determine if presenting this information (e.g., time the student spends on the page, accuracy of their metacognitive judgments, quality of the student’s notes and summaries) to the expert human tutor in real-time contributes to their understanding of the student’s CAMM SRL processes.
Figure 2. Example screenshot of the tutor’s interface illustrating the student’s screen recording (top left), student’s facial expressions and concurrent verbalizations (top right), activity log and dialogue box to submit pedagogical intervention to the student (bottom right), buttons to indicate the presence of the student’s cognitive and metacognitive processes (bottom left), and tutor emote buttons to trigger IVH’s facial expressions to be show to the student (bottom middle).

Interface description. Based on the integration of multiple theoretical frameworks and research with multimodal, multichannel process data to investigate CAMM SRL processes during learning, we designed the tutor interface to include multiple visualizations to facilitate their detection, interpretation, understanding, and actions based on these processes. On the top left of the interface is a mirrored display of the student screen that is customizable depending on the tutor’s preferences. For example, the tutor has the option to turn on or off an overlaid display of the student’s computer mouse movements or their eye-movement behavior (captured by the eye tracker on the student’s workstation; see section below). To the right of the screen mirror is a video stream of the student’s facial expressions that also plays their verbalizations to the tutor. Below the video stream is the activity log, which shows the student’s time-stamped interactions
with their side of the interface (e.g., scores on quizzes, metacognitive judgment values, and time spent interacting with a specific page). This feature was designed to emulate a typical log file, capturing the student’s real-time actions during learning. For example, the activity log displays the name of the page the student is currently viewing, the amount of time the student spent on that page, and the score of their quiz initiated by clicking the cognitive and metacognitive buttons on the SRL palette. Additionally, the activity log displays the notes, summaries, and inferences that the student makes in real-time, which allows the expert human tutor to assess the quantity and more importantly the quality of these notes, summaries, and inferences and how the quality may change during learning.

Beneath the student’s screen mirror are two groupings of clickable buttons related to the student’s cognitive and metacognitive SRL processes. Underneath these buttons are the valence buttons that are indicated by a plus sign (+) or a minus sign (-), that allow the tutor to denote the accuracy of the student’s enactment of these cognitive and metacognitive SRL processes. Specifically, these buttons allow the tutor to justify their actions during interactions with the student (i.e., feedback given to the student through the IVH’s facial expressions and chat log that is based on the tutor’s assessments of the student’s use of cognitive and metacognitive processes). Second, these buttons also indicate if the expert human tutor shows a preference (i.e., focusing more on the student’s note-taking), flexibility (i.e., can discriminate between cognitive and metacognitive processes), and expertise (i.e., can detect and respond to both cognitive and metacognitive processes) for coding specific cognitive strategies or metacognitive SRL processes based on the student’s behaviors. Ultimately, the overall goal for these buttons is to allow the expert human tutor to code the student’s real-time cognitive and metacognitive SRL processes for the development of adaptive features that we can include in future versions of MetaMentor.
(depending on the expert human tutor’s accuracy in identifying these behaviors). For example, depending on the student’s onscreen behavior (indicated by the screen mirror) and the activity log, the tutor will code for the student’s cognitive and metacognitive behaviors. When the tutor views the student’s screen mirror with their overlaid eye-movement behavior, and they see that the student is continuously viewing the timer on the top left of their interface, we hypothesize the tutor will click the “Monitoring Progress” button because that eye-movement behavior suggests that the student is monitoring their progress towards their goals. Furthermore, when the tutor hears the student verbalize that they feel like they have seen the information they are currently reading before, the tutor will click the “Feeling of Knowing” button and the plus sign because the student’s verbalizations suggest they are performing accurate FOKs. Lastly, when the activity log indicates that the student made a low JOL, but answered all of the quiz questions correctly, the tutor should click the “Judgment of Learning” and the negative sign buttons, because this suggests that the student’s JOL is inaccurate.

To the right of the tutor cognitive learning strategy buttons are the tutor emote buttons, which provide the tutor with a list of emotion options they can prompt the IVH to express to the student based on the student’s learning behaviors. Specifically, the tutor can choose the IVH to facially express anger, boredom, confusion, frustration, or joy. We included these options based on extensive research on emotions a student experiences during learning with an ITS (Baker et al., 2010; D’Mello, 2013; D’Mello, Lehman, Graesser, & Pekrun, 2014), research on one-to-one human tutoring interactions (Graesser et al., 2017), and suggestions from a subject matter expert in tutoring complex science content. On the bottom right of the tutor interface is the chat log, which allows the tutor to provide scaffolding and feedback by typing written messages based on the student screen mirror and webcam stream of their facial expressions and concurrent audio
recordings. For example, when the student clicks on the CE button to evaluate the relevancy of the content and judges the content to be irrelevant but continues to spend time reading it, the tutor will prompt the IVH to express confusion to signal to the student that they do not understand why they are continuing to read irrelevant content. In addition, if the student continues to navigate to less relevant content, then the tutor could couple their facial expression of confusion with an instructional message stating, “*Maybe you should examine content relevant to your current sub-goal*” to prompt the student to look for more relevant content. If the student does not comply with these prompts, the tutor could prompt the IVH to express frustration or anger to more forcefully suggest that the student should alter their current learning strategies.

*Standalone tutor interface description.* An expert human tutor or researcher can also use the tutor interface as a standalone program for training and research purposes to understand a student’s CAMM SRL processes during learning with ITSs. With this interface of MetaMentor, the student screen mirror and webcam stream of their facial expressions can be either pre-recorded from previous student’s experimental sessions with the MetaMentor student interface, a human actor portraying specific examples of a student using CAMM SRL processes, or recordings of a student’s learning session with a variety of advanced learning technologies (e.g., a serious game, or a recording of a virtual reality learning experience). While the cognitive and metacognitive process coding buttons remain the same as the “live” interaction version, the tutor emote buttons serve as an emotion coding tool that the tutor or researcher can use to classify the student’s facial expressions. Additionally, the emotion coding feature is expanded to include basic emotions (i.e., joy, sadness, surprise, fear, anger, disgust, and contempt) in addition to the learner-centered emotions (i.e., confusion, frustration, and boredom) already included in the interactive version. This feature was designed to allow the tutor or researcher to detect, classify,
interpret, and code emotions the student experienced during learning across different learning environments that may induce the student to experience emotions not typically observed during learning with ITSs. Lastly, the chat log is also repurposed as a note-taking tool that the tutor or researcher can use to take notes or justify why they chose the combination of CAMM SRL processes based on the screen mirror and webcam stream.

Example Learning Session: Towards Developing an Adaptive MetaMentor

During a typical learning session with MetaMentor, we collect several multimodal, multichannel process data from both the student and the expert human tutor to investigate the student’s and tutor’s CAMM SRL processes during learning, as well as the tutor’s interpretations of these processes in real-time (see Figure 3). The following section describes how we hypothesize an expert human tutor to detect, understand, interpret, and act on a student’s CAMM SRL processes, based on the specific visualizations presented in the tutor interface (i.e., screen mirror with overlaid eye-movement behavior, video stream of facial expressions, and activity log). For sample hypotheses of specific CAMM SRL processes depicted by each visualization, please see Table 1. We also discuss how we will use the expert human tutor’s inferences and actions to identify CAMM SRL data signatures that the future version of MetaMentor will detect and respond to in real-time.

*Screen mirror.* During a learning session, we hypothesize the expert human tutor to use the visualizations of the student’s CAMM SRL processes to inform how they provide the student with immediate feedback and scaffolding about their SRL knowledge and skills, as well as the domain content. For example, we expect the expert human tutor to use the screen mirror with the student’s overlaid eye-movement behavior to identify key cognitive strategies they may use during learning. For example, when the tutor observes that the student is not integrating the text
and diagram, as evidenced by the student only fixating on the text and not the diagram, the tutor will infer that the student is not engaging in the effective learning strategy to coordinate informational sources. The tutor can then click the “COIS” button and prompt the student (using the chat box) to also examine the diagram in addition to the text. Then we will use this information to design the system to automatically identify a student’s COIS behaviors from their eye movements and the corresponding prompts (based on what was typed in the chat box) for future versions of MetaMentor.

Figure 3. Example experimental setup with instrumented student (left) and expert human tutor (right). We emphasize that the physical distance between student and human tutor is optimal to avoid distractions and other socially desirable behaviors.
### Table 1.

**Hypothesized CAMM SRL Processes Depicted from Each Visualization**

<table>
<thead>
<tr>
<th>Visualizations</th>
<th>Cognitive</th>
<th>Affective</th>
<th>Metacognitive</th>
<th>Motivational</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen Mirror</td>
<td>Gaze behavior the text and diagram indicative of COIS</td>
<td>Gaze behavior patterns related to confusion and frustration</td>
<td>Selection of JOLs, FOKs, and CEs</td>
<td>Low quiz score contributing to decrease in task value</td>
</tr>
<tr>
<td>Webcam Video Stream</td>
<td>Activation of AU4 (brow lowerer) indicative of cognitive effort</td>
<td>Facial expressions of boredom</td>
<td>Activation of AU7 (lid tightener) indicative of FOKs and CEs</td>
<td>Facial expressions of boredom contributing to lowered interest</td>
</tr>
<tr>
<td>Audio Channel</td>
<td>Verbalizations of PKAs</td>
<td>Verbalizations of confusion</td>
<td>Verbalizations of JOLs</td>
<td>Verbalizations of high self-efficacy</td>
</tr>
<tr>
<td>Activity Log</td>
<td>Study-time allocation to specific content</td>
<td>Time spent on a diagram causing student confusion</td>
<td>Accuracy of FOKs</td>
<td>Longer time spent on specific content indicative of increased persistence</td>
</tr>
</tbody>
</table>

*Note. These examples are by no means a comprehensive list of the potential CAMM SRL processes to be depicted by each visualization.*

Webcam video stream. We also expect the tutor to use the webcam video stream of the student’s facial expressions to identify the emotions the student is experiencing during the learning session. For example, when the student facially expresses confusion for an extended period of time, the tutor will infer that the student need assistance in regulating their confusion. We hypothesize that the tutor will then prompt the student to engage in an effective emotion regulation strategy, such as cognitive reappraisal (see Gross, 2015), to help externally regulate their confusion. In this case, the tutor could prompt the student to reappraise the source of their
confusion by helping the student identify discrepancies in their understanding of the learning materials by prompting them to re-read the text and re-inspect the diagram. Based on the tutor’s accurate inference of the student’s facial expressions, and the corresponding prompt to externally regulate the student’s confusion, we will design a sensor to automatically detect when a student experiences confusion during learning, and then (depending on several variables related to occurrences, frequency of expression, evidence and intensity of expression, emotion regulation efficacy, etc.) prompt them to engage in specific emotion regulation strategies to alleviate their confusion.

Audio channel. We expect that the tutor may use the audio channel of the student’s concurrent verbalizations to identify instances of the student’s strategy use, metacognitive monitoring, affective responses, and motivational processes. For example, if the student states, “I’m running out of time to complete this sub-goal. I need to look for the right information quickly!” the tutor will infer that the student is monitoring their progress toward their current goal (MPTG), click the “MPTG” button, and suggest to the student to engage in content evaluation (CE) judgments to help them identify information that is relevant to their current learning sub-goal. Additionally, when the tutor hears the student say, “This is too hard. I can’t do this!”, the tutor will then infer that the student may have low self-efficacy for the specific tasks and should prompt them to engage in proximal goal setting by stating, “It sometimes helps to break this task down into sub-goals. Let’s set a sub-goal of finishing this page and look at your progress on the sub-goal bar” (Miele & Scholer, 2018). As we hypothesize these prompts to be effective in improving the student’s metacognitive monitoring processes and increasing the student’s motivation, we will design the system to not only identify instances of a student’s
metacognitive monitoring and motivation based on the student’s verbalizations, but also corresponding, real-time prompts to address these processes.

Activity log. Another option for the expert human tutor to view is the activity log, which provides critical information regarding the quality of the student’s notes and summaries or the accuracy (identified from their quiz scores) of their metacognitive monitoring judgments. When the tutor recognizes that student is simply copying the content from the page verbatim, which is an inefficient learning strategy, we expect the tutor to prompt the IVH to express confusion to signal to the student that they may want to restate the content in their own words. Alternatively, when the tutor recognizes that the student’s summaries are concise synopses of the presented information, we hypothesize the tutor will prompt the IVH to express joy, coupled with a positive and motivating statement: “Good job on your summary!”.

Future adaptations to MetaMentor will then be developed to monitor the quality of the student’s notes (based on the expert human tutor’s evaluations) and automatically program the IVH to provide these types of responses to the student in real-time.

The activity log also displays the scores of the student’s JOL and FOK quizzes and the numerical value of their metacognitive judgments. The tutor will use these to identify how calibrated the student’s monitoring judgments are for the current learning content. If the student performs well on a JOL quiz and makes a judgment to indicate that they understand the information they are viewing, the tutor could prompt the IVH to facially express joy and suggest to the student to move on to other content. Alternatively, when the student is continuously making over-confident JOLs and disregarding previous feedback, we expect the tutor to prompt the IVH to facially express anger or frustration and suggest that the student should spend more time reading and understanding the presented information. As such, we will design the system to
take into account a student’s metacognitive judgment accuracies over the learning session and the expert human tutor’s inferences based on these judgments to inform how the IVH will automatically respond to the student’s metacognitive monitoring processes over the course of the learning session.

With this system, we hope to identify signatures of a student’s CAMM SRL processes during learning and build an automatically adaptive system with the data collected from these learning sessions. However, we also expect to face several challenges during our future research with MetaMentor and developing these adaptive features for future versions. As such, the next section discusses some of the potential challenges we may face when using an expert human tutor’s inferences of a student’s CAMM SRL processes to design an automatically adaptive version of MetaMentor.

**Possible Challenges and Open Questions for Future Research with MetaMentor**

Due to the complexity of the system there are several possible issues that we may encounter in our research with MetaMentor. The following section will discuss some of these challenges and present some open questions to be answered by our future empirical studies with MetaMentor, such as the need to identify valid behavioral signatures of CAMM SRL processes and the most appropriate representations of these processes to present to an expert human tutor. Addressing these challenges and answering these questions will augment our understanding of the multimodal, multichannel data signatures of CAMM SRL processes, support our hypotheses embedded in the design of the system, and even optimize the expert human tutor’s performance by reducing their extraneous cognitive load when using the system.

To effectively create an adaptive system based on a student’s CAMM SRL processes, we will need to identify examples of valid, *ground truth-based* signatures (based on the
visualizations embedded in the tutor interface) representative of effective and ineffective CAMM SRL processes. This will require extensive research examining the underlying factors contributing to these behavioral enactments of CAMM SRL processes. For example, does a student’s neutral facial expression represent that they are currently engaged in learning the content (Baker et al., 2010; D’Mello & Graesser, 2012) or does it indicate that the student may be disengaged from the learning task, and ultimately lacking the motivation (e.g., decrease in their value for the task) to continue learning? Alternatively, does a student’s gaze behavior, indicating several fixations between the text and diagram, mean that they are using an effective cognitive strategy (i.e., COIS) or does it mean that they are confused by the content and hoping to find specific information in the text or diagram to alleviate that confusion? Is the differentiation between inferring between a cognitive strategy vs. affective states based on the use of a single or multiple data channels?

Another challenge will be to distinguish the most appropriate visualizations for each data channel, each specific CAMM SRL process (e.g., JOLs or FOKs, summaries or notes, facial expressions of confusion or joy, etc.), and the integration of multiple CAMM SRL processes (e.g., a student’s facial expressions of confusion coupled with multiple fixations between the text and diagram suggesting that the student is becoming frustrated with the learning task). For example, eye-tracking data can be visualized in multiple ways that range from dynamic video recordings of eye-movement behaviors (e.g., EMMEs; Jarodzka et al., 2017) to static heat maps (i.e., visualizations of cumulative visual attention, over time; Azevedo et al., 2018). Both of these visualizations can represent cognitive processes, but the expert human tutor may need to identify the most appropriate visualization given their current pedagogical decision. For example, the tutor may prefer a dynamic recording of the student’s gaze behavior to identify what the student
is reading at that moment but may need to use a heat map visualization to investigate if the student is not consistently pay attention to diagrams throughout the learning session.

Furthermore, emotions can be represented by a real-time webcam video stream of the student’s facial expressions, but they can also be numerically and graphically represented over time with facial expression recognition software (see iMotions, 2018). It is possible that the expert human tutor may not be able to accurately interpret the student’s facial expression and therefore may need to rely on a graphical representation of these emotions to identify what the student is currently feeling. As such, there may be individual differences (e.g., expertise, personality, pedagogical training, visuospatial ability) that can impact the preferences of the expert human tutor for specific types of data visualizations, and therefore their accuracy in using the student’s multimodal, multichannel process data for instructional decisions.

Accurate detection, understanding, and actions based on these visualizations are imperative, as it is possible that the expert human tutor’s actions based on inaccurate understandings of these visualizations may dysregulate the student and disrupt their learning. Therefore, these data channels must be appropriately visualized, presented to, and personally customized for the expert human tutor to facilitate their effective, accurate, and timely interventions.

Each of these data channels have extensive histories of empirical research where they have been used to detect and interpret specific CAMM processes. For example, eye tracking has been traditionally used to measure cognitive processes (e.g., Rayner, 2009), facial expressions have been used to measure emotions and affect (e.g., Fernández-Dols & Russell, 2017), and log files have been analyzed to understand cognitive and metacognitive processes during learning (e.g., Azevedo et al., 2018; Winne, 2017, 2018). However, it is possible that these traditional
interpretations may not always be the most relevant to understand specific CAMM SRL processes during learning. For example, previous research with MetaTutor has identified eye-movement patterns that represent specific emotional states (i.e., confusion, disengagement) (Jaques et al., 2014). Additionally, research indicates that facial expressions may not always correspond to an emotional state and could instead represent mental effort (e.g., Grafsgaard, Wiggins, Boyer, Wiebe, & Lester, 2013) or metacognitive monitoring processes during learning (e.g., Taub, Azevedo, & Mudrick, 2018). Our future research will need to investigate how an expert human tutor will attend to and integrate visualizations to identify complex interactions of CAMM SRL processes during learning. For example, we will need to answer questions related to: what CAMM SRL data should be presented (e.g., all CAMM vs. some CAMM processes)? How should these representations be presented (traditional formats such as histograms or some new data visualization method more appropriate to illustrating incomplete, multivariate data)?

Another challenge that we may face will be to reduce the extraneous cognitive load the expert human tutor may experience when observing, monitoring, inferring, understanding, interpreting, and acting on these visualizations of a student’s CAMM SRL processes in real time. Successful interpretation and coding of processes while also prompting the IVH to provide a specific facial expression and typing an individualized chat log message will require significant effort. Therefore, the potential for dysregulating the student during learning may be high because ill-timed scaffolding or feedback may confuse or interrupt the student’s learning. Additionally, reducing the cognitive load of the tutor may allow the tutor to identify and act on changes the student makes over a learning session. For example, research suggests that an expert human tutor changes their style of tutoring over the course of a lesson, which is an adaptive and effective tutoring strategy that can significantly improve student learning (Graesser et al., 2017). If the
expert human tutor’s focus is to interpret and understand the visualizations of CAMM SRL data channels, this could detract from their ability to modify and adapt tutoring strategies for the student during the lesson. Therefore, we will need to address questions such as: how should we maximize and optimize the human tutor’s performance without introducing extraneous cognitive load and does this require reducing the number of data visualizations based on contextual (e.g., specific CAMM behaviors during learning) or instructional (e.g., learning objectives, goals of the tutor’s instruction) factors? Reducing the tutor’s cognitive load may allow them to detect qualitative changes in a student’s CAMM SRL processes during the learning session and adapt their feedback, scaffolds, and prompts to the changes in the student’s SRL.

Though we expect some challenges with our future research on MetaMentor, we believe it has transformative potential for how we understand the complex interplay between CAMM SRL processes during learning with ITSs. We hope to extend this platform to other advanced learning technologies to augment our understanding of how a student self-regulates their CAMM processes during learning across domains and learning technologies.

**Conclusion and Future Directions**

In this chapter, we discussed the theoretical and pedagogical origins of MetaMentor. We presented a brief review of contemporary ITS research, our past research with MetaTutor, and the theoretical underpinnings of the MetaMentor design. We argued that advances in ITSs and learning and cognitive science will be based on a comprehensive understanding of both a student’s and human tutor’s real-time CAMM SRL processes during human tutoring sessions. The focus on real-time use of both student’s and human tutor’s multimodal multichannel data is critical in advancing our current theoretical models of learning, instruction, tutoring, and SRL by augmenting our understanding of the individual and combined impact of CAMM processes on a
student’s SRL. Specifically, this approach can generate more conceptually and empirically-based evidence of their sequence, duration, occurrence, relevance, role, intensity, etc. during human tutoring with MetaMentor. These data will force interdisciplinary researchers to use and design new sensors and devices capable to detecting, measuring, and illustrating these processes in real-time for different purposes ranging from open learner models to foster a student’s metacognitive monitoring, to intelligent dashboards capable enhancing a tutor’s understanding of a student’s SRL behaviors and therefore allowing for optimal scaffolding and feedback in real-time.

Our future research with MetaMentor will begin to address some of the questions presented in the challenges discussed above. Specifically, we will need to investigate how multiple expert human tutors can accurately detect and interpret the visualizations to augment our understanding of these processes, and thereby provide evidence for our hypotheses and expectations of the design of this system. As such, we plan to create short video examples indicative of specific CAMM SRL processes to use with the standalone tutor interface to assess if multiple tutors can identify specific CAMM SRL processes. Specifically, we will create examples that we believe should indicate effective and ineffective CAMM SRL processes (e.g., eye-movement behaviors indicative of COIS, facial expressions of confusion negatively influencing content understanding, verbalizations of motivational state changes, etc.) to determine if expert human tutors also agree with our hypotheses. We plan to run studies to investigate the levels of convergence of multiple expert human tutors’ and researchers’ inferences and judgments to address if: (1) there are behavioral signatures of CAMM SRL processes, (2) if they can be reliably detected by expert human tutors and researchers, to eventually integrate (1-2) in an adaptive MetaMentor capable of responding to these signatures to effectively foster a student’s SRL during learning. Identifying the convergence of both
researchers’ and tutors’ interpretations of behavioral signatures of CAMM SRL processes will augment our conceptual and theoretical understanding of using multimodal, multichannel data to understand these processes during learning with ITSs.

Finally, MetaMentor is designed as a plug and play system, so any researcher, educator, teacher, or training specialist can literally plug in their multimodal multichannel data of some human-machine interaction to be analyzed by a new set of human tutors. For example, imagine multimodal multichannel data (e.g., motion capture data of physical location in the environment, verbalizations of cognitive and metacognitive processes, EDA, etc.) from a high-school student learning about photosynthesis using a virtual reality system. These data streams can be fed into MetaMentor and be used to study novice and expert human tutors’ ability to observe, understand, and infer the student’s SRL behaviors and assess their accuracy and impact of their pedagogical interventions (e.g., inability to accurately infer certain metacognitive processes, etc.). As such, our goal is to develop and test MetaMentor to have broad impact and contribute the cognitive, learning, computational, engineering, and instructional sciences, and advanced learning technologies.

**Acknowledgments**

Please address all correspondence to Nicholas V. Mudrick, North Carolina State University, 2310 Stinson Drive, 640 Poe Hall, Raleigh, NC 27695, USA, nvmudric@ncsu.edu. This chapter was supported by funding from the National Science Foundation (DRL#1431552; DRL#1660878, DRL#1661202, CMMI#1744351) and the Social Sciences and Humanities Research Council of Canada (SSHRC 895-2011-1006). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or Social Sciences and Humanities.
Research Council of Canada. The authors would also like to thank members of the SMART Lab and IntelliMedia Group at NCSU for their assistance and contributions.
References


Lester, J., Lobene, E., Mott, B., & Rowe, J. Serious games with GIFT: Instructional strategies, game design, and natural language in the generalized intelligent framework for tutoring.


Handbook of research on 3-D Visual environments and hypermedia for ubiquitous learning (pp. 362-386). Hershey, PA: IGI Global.


CHAPTER 6: INTEGRATED DISCUSSION

The goal of the dissertation was to detail the conceptual, theoretical, methodological, and design components for assessing CAM SRL processes with multimodal, multichannel process data during learning with ITSs. Specifically, this dissertation demonstrated that multimodal, multichannel process data can augment our conceptual and theoretical understanding of CAM SRL processes by measuring and identifying how these processes interact and unfold in real-time. Many contemporary frameworks and theories of cognition, affect and emotions, metacognition, and SRL are process specific (e.g., focusing on cognition and metacognition and not addressing learner-centered emotions, focusing on emotions and not explaining metacognition, etc.) and are limited in their ability to explain and predict the occurrence of and interactions between CAM processes during SRL. Therefore, multimodal, multichannel process measures can be used to augment these frameworks by not only concurrently measuring all of a learners’ CAM SRL processes, but also identifying specific instances of optimal and efficient SRL behaviors. These identified behaviors can then help explain how these processes interact and contribute to learning outcomes. Furthermore, this dissertation also provided examples of how researchers can use specific experimental designs (e.g., within-subjects and Wizard of Oz experimental designs) to measure and understand the impact of specific CAM processes on learners’ SRL, thereby contributing to theories of these processes, and identifying features to embed in future ITSs. As such, this dissertation presented several critical issues underlying conceptual, theoretical, methodological, and design considerations related to the development of future CAM-sensitive ITSs.

Chapters 1 and 2 outlined the contributions multimodal, multichannel process data can have to augment underspecified theoretical frameworks of specific CAM processes related to
effective SRL. Chapters 3 through 5 demonstrated how to use different experimental approaches to identify behavioral signatures of CAM SRL processes to augment these theories and frameworks. In our handbook chapter (Chapter 2), we presented several conceptual issues related to the inability of contemporary theoretical models and frameworks of CAM processes to predict, describe, and explain the occurrence and influence of key CAM processes on learners’ ability to self-regulate their learning. For example, one conceptual issue related to these models pertains to how CAM processes unfold during learning. Many models do not specify if CAM SRL processes unfold as discrete events (e.g., occurring serially, learners’ experiences of confusion following miscalibrated and low judgments of learning) or in parallel (e.g., co-occurring, learners’ experiences of confusion occur as they make miscalibrated and low judgments of learning) during effective SRL. As emphasized by Chapters 3 and 4, constraining the learning context with within-subjects designs and concurrently collecting multimodal, multichannel process data allows researchers to be more certain of the occurrence of specific CAM processes. If we collect multiple measures of learners’ CAM processes, we can align their learning behaviors indicative of effective SRL with specific sequences and signatures present in these data channels to identify if they occur serially or in parallel. Additionally, Chapter 5 outlined an alternative empirical approach to identify the interrelationships between CAM processes and how they influence each other during learning in real-time. Taken together, this dissertation emphasizes the utility of collecting multiple measures of learners’ CAM processes (e.g., eye tracking and log files for cognition and metacognition, videos of facial expressions for emotions) to identify CAM processes indicative of effective SRL. By using these approaches, we can understand specific sequences and signatures present in these data channels to identify how they occur during learning. Ultimately, we can use these experimental approaches to expand
current theoretical frameworks, investigate how CAM SRL processes unfold and interact, and design ITSs capable of adaptively scaffolding and responding to these processes during learning.

**Designing CAM-sensitive ITSs.** In addition to augmenting current theories and models, the chapters presented in this dissertation outline how indices of learners’ CAM processes derived from multimodal, multichannel data can be used to develop future ITSs and make these systems more adaptive. For example, Mudrick et al. (2017, 2018) identified specific instances of learner-system interactions that provide evidence for designing future systems. Specifically, in Mudrick et al., (2018, Chapter 3), the authors analyzed log files to identify the significant relationship between accurate EOL judgments and learners’ multiple-choice performance during learning. Because these EOL judgments significantly predicted performance and did not require significant effort from learners to respond (as indicated by the mean duration of these judgments), the authors suggest the design of future ITSs should include EOL-type features because they provide the system with important information about learners’ accurate metacognitively monitoring and may provide information on their subsequent CAM SRL processes. The system can use these judgments to tailor scaffolding and feedback to learners’ EOL judgment accuracy and/or the resulting CAM SRL processes that occur based on these judgments. For example, based on learners’ inaccurate and over-confident EOL judgments, an embedded IVH tutor agent could prompt learners to pay more attention to the to-be-learned content because it may be more difficult to learn than they had originally judged. Alternatively, an embedded agent could model an effective cognitive learning strategy like coordinating informational sources (COIS) to help learners integrate the multimedia content (Greene & Azevedo, 2009). For example, the agent could scaffold learners’ understanding by engaging in an eye-movement modeling example (EMME) (e.g., van Marlen, van Wermeskerken, Jarodzka,
& van Gog, 2016). The agent could overlay their own eye-movement sequences on the multimedia content to cue learners to relevant information and explain why the content is important to learn and how to engage in this strategy in the future. Furthermore, if learners become confused by the difficult content, an agent could intervene and prompt learners to engage in an emotion regulation strategy, such as cognitive reappraisal (e.g., a process by which individuals modify appraisals of their current situation; Gross, 2015) to prevent learners’ confusion from transitioning into frustration, boredom, and task disengagement (e.g., D’Mello & Graesser, 2012). Specifically, the embedded virtual human tutor could prompt learners to re-evaluate the source of their confusion from perceiving the content to be too difficult to learn, to perceiving the difficult material as an opportunity to improve their content understanding and inform their subsequent CAM SRL strategies by stating, “This content is not too difficult to learn, it is a learning experience you can draw from to build your knowledge and improve your self-regulatory strategies”. Although this study analyzed log-file data and assessed learners’ metacognitive monitoring, the findings can contribute to the development of adaptive scaffolds to address all of learners’ CAM SRL processes during learning with ITSs.

In Mudrick et al. (2017, Chapter 4), the authors analyzed log-file and facial expression data and found that learners’ performance was highest when the embedded tutor agent’s facial expressions were congruent with the relevancy of the content. Furthermore, results revealed learners experienced more confusion during the embedded tutor agent’s neutral facial expression (neither congruent nor incongruent with the content). Based on these findings, the authors suggest that the contextual congruency (e.g., with the content, the emotions of the learners themselves, the agent’s own understanding, etc.) of an agent’s emotional expressions should be considered when designing these agents to embed in future systems. More specifically, when an
agent facially expresses certain emotions to scaffold learners’ behaviors, the system should be aware of the content learners are interacting with, as well as the specific CAM SRL processes that learners are engaging in (e.g., taking notes, accurate JOLs, facial expressions of confusion, etc.). By taking these contextual considerations into account, the system can foster increased learning outcomes and mitigate potential disruptions to learners’ SRL behaviors with ill-timed or inappropriate agent-based facial expressions. For example, when learners take notes on content not relevant to their current learning goal, an agent could facially express confusion to prompt learners to engage in a CE to evaluate the relevancy of the content. However, if the agent provides a neutral facial expression following an accurate CE, it may dysregulate learners and negatively impact their learning by causing them to experience confusion and continue to interact with irrelevant content. These findings emphasize the utility of converging multiple data channels to ensure that the scaffolds provided by the system do not dysregulate learners’ CAM processes and disrupt their SRL.

In Mudrick et al. (in press, Chapter 5), the authors outlined how data signatures of learners’ CAM processes identified by expert human tutors can be used to develop adaptive interventions that can be designed back into the system to scaffold and support learners’ self-regulation in real-time. Additionally, the system described in this chapter also offers researchers the opportunity to empirically test many of the suggestions outlined above. For example, if the expert human tutor identifies that learners are confused based on learners’ facial expressions, they could prompt learners to engage in cognitive reappraisal to assess its effectiveness in assisting learners in regulating their emotions. Alternatively, if the expert human tutor judges the content learners are currently viewing as irrelevant to their current learning goal (based on the learners’ eye-movement and log-file data), they could prompt the IVH to facially express
confusion to signal to learners the content is irrelevant and assess whether or not the agent’s facial expression is enough to scaffold effective SRL behaviors. Ultimately, this approach offers researchers to compare and contrast the behavioral signatures of learners’ CAM SRL processes identified from more constrained within-subjects experimental approaches. Additionally, this approach allows us to apply the findings from these within-subjects by empirically testing the effectiveness of the suggested design components derived from these findings.

Taken together, these chapters emphasize how learners’ multimodal, multichannel process data can be used to design interventions ITSs can use to scaffold learners’ CAM SRL processes. Furthermore, these chapters demonstrate the need to comprehensively address all of learners’ CAM SRL processes to improve their ability to self-regulate their CAM processes and achieve high learning outcomes. However, there are still several broader implications and future directions based on this dissertation that should be addressed in future research and will assist in developing fully adaptive, CAM-sensitive ITSs.

**Broader Impacts and Future Directions**

The empirical results presented in this dissertation and the issues raised in each chapter emphasize how we can use multimodal, multichannel process data to identify CAM SRL processes to augment theoretical frameworks and design adaptive systems that detect and respond to these processes. This dissertation provides insight into broader implications and future directions for research that investigates learners’ CAM SRL processes with multimodal, multichannel process data during learning with ITSs. Specifically, these implications regard methodological and measurement, theoretical and conceptual, pragmatic, and design considerations to be addressed in future research with ITSs.
**Methodological and measurement considerations.** Methodologically, broader implications and future directions based on these chapters relate to the validation of specific data channels to indicate particular CAM SRL processes. Specifically, future research will need to validate multichannel process measures to detect, understand, and explain the individual and combined impact of specific CAM SRL processes. This dissertation proposed how we can use within-subjects experimental designs to validate some of our a priori assumptions of how these processes occur and present themselves during learning. However, there are other benefits to this experimental approach can assist the validation of these signatures by using more bottom-up and data-driven methods. With within-subjects designs, we have multiple instances from within the same learner over time and can use these designs to identify if there is any stability to behavioral signatures of CAM SRL processes (i.e., if the activation of Action Unit 4 [Brow Lowerer] consistently serves as a precursor an accurate judgment of learning after inspecting diagrams), or if some processes reliably and consistently impact others (i.e., concise and accurate summaries prompting learners to facially express joy during learning with multimedia STEM content). Another benefit to within-subjects design allows us to compare behavioral signatures across participants, but in the same learning context (i.e., inspecting multimedia STEM content) that may not be possible with other types of learning environments with less structured learning tasks (e.g., traditional ITSs with multiple pages of content, or a game-based learning environment where learners have full agency over where they go in the game, when, and for how long, etc.). As such, employing these types of experimental designs can offer researchers the opportunity to systematically examine the impact of CAM processes on learners’ SRL with ITSs.

Furthermore, the last chapter in this dissertation (i.e., Mudrick et al., in press) proposed a different and novel method to validate the behavioral signatures of learners’ CAM SRL.
identified from the proposed within-subjects experimental approach. Identifying these CAM processes more in situ will allow for researchers to identify how these processes occur in the absence of rigid experimental constraints. We can also expand these types of investigations where researchers and expert human tutors identify CAM SRL processes by having individual learners themselves validate their own behavioral signatures of their own CAM process data. Specifically, we can use the Wizard of Oz-type approach with researchers and expert human tutors prompting learners to confirm their own judgments of learners’ CAM SRL processes. For example, if researchers are observing learners’ eye movements on the screen and they see that the learners are fixating on the table of contents and then transitioning back to the learning goal and vice versa, researchers could potentially infer that learners are engaging in accurate CEs. Then, the researchers could prompt learners with a simple question (e.g., “Based on your eye movements, it looks like your evaluating the relevancy of the content to your current goal, do you agree?”) to confirm their evaluations. Alternatively, we can set a priori detection thresholds using off-the-shelf software to also identify learners’ other CAM processes. For example, iMotions’ FACET facial expression recognition software classifies the presence of emotions as probabilities, representing the log odds of facial expressions that represent specific emotions as coded by an expert human coder (iMotions, 2018). We can set specific thresholds for both numerical evidence values or the duration of the facial expression to then fire a prompt to learners to confirm that they are experiencing the emotion the software is detecting. For example, when learners are reading multimedia content and the evidence values for confusion exceed the predefined threshold of 2 (indicating that 100 expert human coders would classify the emotion as being present), the system could fire a question asking, “Are you confused by this content?” to confirm the detection of the emotion. Furthermore, we can administer retrospective think-aloud
protocols to learners following their learning sessions by showing the learners visualizations of their own multimodal, multichannel CAM process data and then asking them to explain what they were doing during learning. By engaging in these approaches, we can converge both learner and researcher understandings of how these processes occur during learning, contributing to the validity and reliability of multimodal, multichannel signatures of specific CAM SRL processes.

Theoretical and conceptual considerations. There are several open conceptual and theoretical issues regarding the methodological and measurement components to use these data to detect and understand CAM SRL processes during learning with ITSs. For example, if we use the previously discussed within-subjects experimental approach to help validate multimodal, multichannel process measures to detect CAM SRL processes, there may be limitations based on generalizability of the findings. Specifically, there may be instances of context-specific occurrences of particular CAM SRL processes that are dependent on the learning content format (e.g., multimedia vs. animation), task (e.g., ill-structured vs. well-structured task), domain (e.g., STEM vs. history, English, etc.), individual differences in learners’ CAM SRL profiles (e.g., emotion regulation flexibility, internal conditions, accuracy of monitoring judgments, etc.). More broadly, we have yet to determine if data signatures generalize across individual learners, constructs (e.g., emotions, cognitive strategies, metacognitive judgments, etc.), and context (e.g., learning with an ITS, game-based learning environment, MOOCs, etc.). As such, the behavioral signatures identified from one ITS may not be applicable to identify these processes in another ITS, while those identified from one learner may not generalize to all learners. Therefore, future research will need to examine how these behavioral signatures transfer from one learning context (e.g., a multi-agent hypermedia-based ITS for human biology) to another (a triologue-based ITS for learning physics), and if they do, how do they differ and why?
Other theoretical and conceptual questions regarding using multimodal, multichannel process measures relate to the stability and qualitative improvements of the constructs being measured over the course of the learning session. Specifically, contemporary theoretical frameworks do not address the stability of the constructs (e.g., metacognitive judgment accuracies, capacity to regulate negative emotions, ability to summarize multimedia content, etc.) explained in these models during learning. For example, are there qualitative and quantitative differences in the effectiveness of learners’ abilities to self-regulate their CAM processes as they learn with ITSs (e.g., do learners’ internal standards shift over time, can their emotion regulation strategies decrease in effectiveness, does working memory capacity vary based on specific learning content, etc.)? Additionally, since one of the main foci of ITSs is to scaffold and foster learners’ SRL, how do we account for the impact of the agents embedded in these systems to facilitate more effective SRL behaviors and does this vary by type of CAM process? Future research will need to examine how learners’ CAM SRL processes change over time and identify how these behaviors improve or worsen during learning. Furthermore, future research will need to account for the system influence on learners’ abilities to regulate their CAM processes to fully understand the specific impact that ITSs can have on learners’ SRL.

There are additional considerations regarding the conceptual alignment of contemporary theoretical frameworks with current multimodal, multichannel process measures. Many contemporary models vary in terms of the grain-size (e.g., macro- and micro-level processes) of the constructs they describe, and many multichannel process measures also vary in terms of the grain-size the sensors measure. Log files sample learner-system interactions at the millisecond level, eye trackers capture learners’ eye movements at 30hz – 1000hz, while facial expression evidence values are calculated a rate of 30Hz, and EDA sensors typically measure learners’
physiological arousal and heart rates at 128Hz. On the other hand, many theoretical frameworks and models of learners’ CAM SRL processes conceptualize these constructs at varying levels (e.g., monitoring and controlling vs. JOLs and summaries), which leads to questions related to the alignment of these grain-sizes specified in these models with the actual capability of process measures. For example, are there optimal grain-sizes and units of analysis that are needed to identify specific CAM processes? Do macro-level processes (e.g., monitoring, planning, etc.) require macro-level measurements (e.g., total amount of time spent in a specific area of interest, total learning session duration, etc.)? Alternatively, how do we align micro-level processes with micro-level measurements (e.g., a 250ms fixation with an accurate EOL judgment)? Should we emphasize some data channels more importantly than others based on the grain-size of the construct measured? Do we use different measures for different grain sizes (e.g., EDA for global affect, facial expressions for specific emotions)? Can all grain-sizes be measured in real-time, or do some processes require post-processing and analyses to interpret? And if they require post-processing, how can we design ITSs capable of detecting and responding to these processes in real-time? As such, answering these questions will assist not only in augmenting our impoverished theoretical frameworks, but also in the development of CAM-sensitive ITSs capable of responding to learners’ real-time and unfolding self-regulatory CAM processes.

**Pragmatic considerations.** In addition to the theoretical, methodological, and conceptual issues outlined above, several pragmatic components exist when using multimodal, multichannel data that should be considered. Specifically, multimodal, multichannel process measures are largely dependent on industry development. Sometimes this reliance on industry development can severely detract from conducting experimental studies. For example, if an eye tracking company is purchased by a large technological corporation, researchers who use the eye-tracking
company’s equipment and depend on the eye-tracking company to provide support and update the firmware and software are left on their own. If a problem with the eye tracker occurs and the eye tracker cannot function and collect experimental data, the researchers may need to stop their data collection and may not be able to purchase another eye tracker because of a lack of funding. Therefore, it is important that researchers are aware of the potential risks to their experiments and plan accordingly to mitigate these problems.

Additionally, there are other considerations researchers must make before conducting experimental studies where they collect multimodal, multichannel process data. Firstly, the equipment and sensors required to collect these data (e.g., eye trackers, facial recognition software, EDA bracelets and sensors) can be very expensive and challenging to use without proper training. Secondly, research assistants must go through extensive training to be able to equip learners with these devices and sensors and accurately calibrate the equipment. Furthermore, extensive time is required to accurately calibrate these devices and sensors (e.g., a 9-point calibration procedure for the eye tracker to collect accurate eye movement behavior, a 5-minute baseline is needed to accurately capture learners’ EDA, etc.). As such, future research will be needed to identify accurate and reliable indices of CAM SRL processes to overcome some of these issues which affect not only the ability to design CAM-sensitive ITSs, but also scale these systems up to other learning contexts (e.g., from the laboratory to the classroom).

**Design considerations.** Addressing these previously discussed issues and answering these questions will assist in the development of CAM-sensitive ITSs. However, other considerations need to be addressed before we can develop automatically adaptive systems capable of responding to and fostering learners’ CAM SRL processes in real-time. Designing ITSs to use these process measures requires researchers to create and refine accurate and
generalizable domain models of CAM SRL processes that outline the set of skills, knowledge, and strategies required to internalize and enact these processes during learning (Sottilaire et al. 2017). For example, what are the relevant declarative and procedural metacognitive knowledge components contributing to an accurate JOL? What are the procedural steps learners need to engage in to effectively regulate their frustration? As demonstrated by this dissertation, creating these domain models of effective SRL processes entails advancing theoretical and conceptual frameworks to identify sequences, interactions, dynamics, qualitative and quantitative changes, etc. based on behavioral signatures of effective and accurate CAM SRL behaviors abstracted from multimodal, multichannel data.

Additionally, research is still needed to identify the optimal scaffolds that ITSs can use effectively support learners’ CAM SRL processes and learning. Extensive research has been conducted with many contemporary ITSs and advanced learning technologies (e.g., MetaTutor, Betty’s Brain, AutoTutor, Crystal Island, etc.) on how to scaffold specific cognitive, affective, and metacognitive processes during learning (Azevedo et al., 2013; Biswas, Segedy, & Bunchongchit, 2016; Graesser, 2016; Rowe, Shores, Mott, & Lester, 2011). While these studies have significantly increased our understanding of the impact of these scaffolds on certain CAM SRL processes, more research is needed to identify the instructional interventions that can adaptively support learners’ collective CAM processes and foster effective SRL in real-time (e.g., Azevedo et al., 2018; Kim, Belland, & Walker, 2018). For example, should the framing of a prompt to engage in an emotional regulation strategy (e.g., cognitive reappraisal) differ from a prompt to engage in a cognitive learning strategy (e.g., taking notes)? Should ITSs deliver scaffolds to assist learners’ in regulating their confusion as soon as learners’ facially expressions indicate they are confused? Additionally, should these interventions always be delivered verbally
by an embedded tutor agent or could it be more effective when ITSs adapt the content presentation/format to regulate these processes for learners?

**Conclusion**

Successfully adapting to learners’ real-time SRL during learning with ITSs requires a comprehensive understanding of how learners’ CAM process unfolds during learning. By investigating, understanding, and identifying how these processes unfold and interact with multimodal, multichannel process data, we can augment our current theoretical frameworks and design systems capable of responding to these processes in real-time. Ultimately, by designing systems to detect, understand, and respond to learners’ real-time CAM SRL processes, we can individualize instruction, scaffold and support learning, and contribute to enjoyable and successful learning outcomes with ITSs.

The research presented in this dissertation shows that valid and reliable indicators of temporally unfolding CAM SRL processes still need to be identified to design ITSs that can respond to these processes and foster learning in real-time. Furthermore, though this dissertation outlines several experimental and methodological approaches (e.g., within-subjects experiments, retrospective think-aloud protocols) to investigate and validate these processes, alternative approaches should be considered to address the limitations of the outlined methods (e.g., generalizability of the behavioral signatures). Additionally, this dissertation focused exclusively on developing adaptive ITSs; however, using multimodal, multichannel process data to identify, understand, and act on learners’ CAM SRL processes is not limited to these systems. Therefore, using these measures to identify learners’ CAM SRL processes across learning environments (e.g., game-based learning environments, virtual and augmented reality learning experiences, etc.) will help validate these measures and design other types of systems that can use these data.
channels to adapt to learners’ real-time CAM SRL processes. Effective SRL requires learners to monitor and regulate their CAM processes during learning, and by designing systems that respond to these processes in real-time, we can assist learners, optimize their experience, and ensure they engage in effective SRL to achieve high learning outcomes.
REFERENCES


*Learning and Instruction, 22*, 145-157.


