AN ADAPTIVE SCAFFOLDING FRAMEWORK FOR SELF-REGULATED LEARNING IN AN OPEN-ENDED LEARNING ENVIRONMENT

By

Anabil Munshi

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

COMPUTER SCIENCE

May 12, 2023

Nashville, Tennessee

Approved:

Gautam Biswas, Ph.D.

Douglas Fisher, Ph.D.

Maithilee Kunda, Ph.D.

Noel Eneyedy, Ph.D.

Ryan Baker, Ph.D.

Tyler Derr, Ph.D.

ACKNOWLEDGMENTS

First, I express my sincere gratitude to my advisor Professor Gautam Biswas for his guidance and mentorship over the last six years, leading to the completion of this dissertation. I also want to thank Professors Ryan Baker, Maithilee Kunda, Douglas Fisher, Noel Enyedy, and Tyler Derr, for serving on my dissertation committee and providing insights that have greatly improved the quality of this work.

I joined the OELE Lab / Teachable Agents Group at Vanderbilt in 2016. In the years since, I have had the opportunity to collaborate with and learn from some amazing researchers. This list includes, among others, Ryan Baker and Jaclyn Ocumpaugh at UPenn, Luc Paquette at UIUC, Jennifer Quinde at VUMC, Haeseon Yun at HU-Berlin, and my colleagues and friends at the OELE lab - Ram, Naveed, Marian, Ningyu, Nicole, Caitlyn, Ben, Yi, Shitanshu, Allison, Mona, Gayathri, Olivia, Yiran, Caleb, Eduardo, Clayton, Joyce, and Celestine. A special thanks to Marian for being such a supportive teammate on the Betty's Brain project for over four years! Thanks to the teachers - Mr J.P., Ms J.S., and others - who let me run studies in their classrooms, and the students who participated in these studies. I am also grateful to the National Science Foundation for supporting the foundational research towards this dissertation through Award 1561676.

Moving to the other side of the world and navigating the journey through grad school were challenging at times, and I am grateful for the support of friends who have often been a home away from home. Thanks to Neelanjana, Purboday, Deepayan, Avisek, Souhrid, Tonuka, Rudra da, Soumita di, Gourab da, and others, for the Friday adda, the food, and everything else. Thanks to Shyamali Mukherjee for making me feel like family since my first day in the US. Thanks to my friends in India and extended family in India / US who kept checking in from time to time. Thanks also to former teachers like Prof Kaushik Dasgupta who encouraged my interest in research from a young age.

Music has always been a big part of my life and helped me relax in the most stressful of times. Thanks to Aman for jamming with me for over ten years now, from the rehearsals and performances in India to the virtual sessions since I moved to Nashville. Thanks to JJ, Paul, and the others, for the weekend jam sessions at Blair that were a very welcome source of relaxation in the dissertation writing period.

And finally, I am very thankful to my family for their unconditional love, support and understanding. Thanks to my father Professor Asit Munshi, for inspiring a career in research. Thanks to my brother Satyapan, for being the annoying younger brother since I can remember, and for growing up into the brilliant and slightly less annoying younger brother I am so proud of. And most of all, thanks to my mother Kakali Munshi, for being my first teacher, for always encouraging my questions about anything and everything, and for showing me the impact of a good educator on a young mind, thereby inspiring my interest in doing research in the education and learning space.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Introduction

I.1 Motivation for this Dissertation Research

An important goal of computer-based learning environments (CBLEs) is to help students develop self-regulated learning (SRL) behaviors and strategies to help them become effective life-long learners (Bransford et al., 2000; Zimmerman and Martinez-Pons, 1990). Self-regulated learning (SRL) focuses on learners' abilitiesy to understand and control their learning behaviors, which helps them to accomplish their learning and problem-solving goals in an effective manner (Panadero, 2017). This process emphasizes students' autonomy, self-monitoring, strategy use, and self-reflection during problem-solving.

Open-ended learning environments (OELEs) are constructivist learning environments that are designed to support SRL development by providing students with (1) *targeted learning goals* in the form of problem-solving tasks (e.g., to construct a model of a scientific process); (2) *a set of tools* to facilitate the learning and problem-solving processes; and (3) *an open-ended approach* that supports student agency i.e., provides choice in how students combine these tools to achieve their learning goals (Biswas et al., 2016). OELEs developed in our lab use model-building tasks to help students improve their strategic thinking skills when building process models in scientific domains (Kinnebrew et al., 2017; Basu et al., 2017; Hutchins et al., 2020).

More advanced learners who typically have well-organized domain knowledge structures are able to utilize the agency and exploration opportunities provided by OELEs to their advantage, engaging in strategic SRL behaviors at different phases of their learning process. These self-regulated learners develop the ability to set their own goals and sub-goals, devise a plan towards achieving their goals, execute the plan by invoking relevant cognitive strategies, engage in metacognitive monitoring behaviors to assess the effectiveness of their approach, and self-reflect to decide whether they need to change their plans or apply a different set of strategies to achieve their learning goals (Winne and Hadwin, 2008; Panadero, 2017).

Novice learners, on the other hand, often face more challenges in these open-ended problem-solving spaces (Segedy et al., 2013; Metcalfe and Finn, 2013; Basu et al., 2017). These students may find it difficult to use the learning tools in the OELE in an efficient manner, and may also lack the experience, prior knowledge, and understanding necessary for effective self-regulation (Zimmerman, 2002b). When they encounter obstacles during their learning process, these students often struggle to engage in successful monitoring and self-reflection behaviors to think critically about the problem. As a result, they continue to engage in sub-optimal behaviors (Devolder et al., 2012; Schwartz et al., 2009; Winne, 2010), and fail to resolve their difficulties. In turn, their lack of success leads them to become frustrated, and then bored and disengaged from their learning tasks (D'Mello and Graesser, 2012). These students require timely guidance via targeted *adaptive scaffolds* to help them develop more effective problem-solving behaviors and adopt strategies that address their learning difficulties and improve learning outcomes.

Adaptive scaffolding, which detects and responds to learners' difficulties as they work in OELEs, can be an effective way to support these students to develop their self-regulated learning (SRL) process, so they can overcome their current obstacles and also become more independent and strategic in preparation for future learning (Lajoie and Derry, 1993; Bransford and Schwartz, 1999).

However, designing adaptive scaffolds to support students' SRL skills and strategies in an OELE comes with its own set of challenges. First, the scaffold design process requires an in-depth understanding of the factors involved in students' self-regulated learning processes. Current frameworks for studying SRL recognize it to be a dynamic process made up of interacting cognitive, affective, metacognitive and motivational components, often abbreviated as 'CAMM' processes (Azevedo et al., 2012). An OELE that scaffolds one or more of these CAMM processes during learning can empower students to become more strategic in their learning and problem-solving process and ultimately help them to successfully complete their learning tasks (Azevedo et al., 2017; Taub et al., 2020).

Implementing this form of scaffolding in an OELE requires *online adaptation*, where the system can track and model students' behaviors and performance as they learn, and use this information to adapt and generate appropriate feedback (Dabbagh and Kitsantas, 2012; Moreno and Mayer, 2000). Therefore, to design and implement such adaptive scaffolds for SRL processes in an OELE, we need to understand (a) what constitutes an effective versus ineffective SRL (CAMM) strategy in the context of the specific problem-solving tasks and goals in a learning environment, (b) how to track and model the use of such strategies from learners' online activities, (c) how to use this information to identify key moments which signal that a learner is having difficulties in invoking appropriate cognitive or metacognitive strategies or in regulating their motivation or affect, and is therefore in need of external guidance, and, last, (d) how to design scaffolds that respond to the learner's needs at such moments via strategic guidance. For an adaptive scaffold to be meaningful and beneficial to the learner, it needs to be contextualized to the learner's current task, learning artifacts, and activities (Segedy et al., 2013). As Vygotsky (1978) suggestsed, the scaffold should attempt to bridge the gap between the student's *current* (what they can do by themselves) and *potential* (what they can do with the help of others) planes of development, by helping them recognize and fix ineffective learning behaviors through cognitive-metacognitive or emotion regulation.

To this end, this dissertation presents an adaptive scaffolding framework to support some of the CAMM components of K-12 students' SRL process in Betty's Brain (Leelawong and Biswas, 2008; Biswas et al., 2016), an agent-based open-ended learning environment (OELE). The framework is designed to primarily model students' cognitive and metacognitive strategy use and effectiveness as they interact with the learning environment, and to detect moments that suggest students are having difficulties with their self-regulation behaviors. System scaffolds are provided at such moments, in the form of conversational feedback initiated by a mentor agent, Mr Davis, present in the Betty's Brain environment, e.g., (Segedy et al., 2013; Munshi et al., 2022b). The conversational mode of the feedback aims to engage students in the type of authentic social interactions that support learning and critical thinking (Vygotsky, 1978). The scaffolding framework ensures that the feedback is contextualized to a student's current task and recent activities, and targets the development of task-oriented SRL strategies needed at that point to resolve their learning difficulties and bring them closer to their learning goals in the environment. These scaffolds primarily take the form of strategic hints (often with varying levels of contextualization) intended to fill in knowledge gaps and provide actionable information to help students develop more effective cognitive strategies and engage in the associated metacognitive monitoring and self-reflection behaviors. Depending on the triggering condition and the interconnected nature of CAMM processes, the feedback may also target improving affect or motivational components of students' SRL process.

The adaptivity framework developed for this dissertation also advances prior research on learner modeling and adaptive scaffolding in OELEs. Students' activities and temporal behaviors in Betty's Brain are interpreted using a modified version of Kinnebrew et al. (2017)'s OELE task model to determine the use of cognitive and metacognitive strategies. A strategy detection approach, which builds upon Basu et al. (2017)'s learner modeling scheme for the CTSiM environment, facilitates the diagnosis of deficiencies in learners' cognitive strategy use and the inference of underlying sub-optimal metacognitive monitoring and self-reflection behaviors. The scaffold triggering conditions and the feedback content are informed by findings from prior scaffold design and evaluation cycles (Munshi et al., 2022b,a). In the current version of the adaptive scaffolding framework discussed in this dissertation, scaffolds are adapted to an understanding of students' cognitive-metacognitive or cognitive-affective self-regulation deficiencies in different task contexts within the learning environment, and are offered to support the development of more effective cognitive and metacognitive, (and in few cases, affect) regulation strategies for their knowledge construction, scientific model-building and model-debugging tasks in Betty's Brain.

I.2 Approach for the Development and Evaluation of Adaptive Scaffolds

There are two primary components to this dissertation research: (1) Building a framework for the design and implementation of adaptive learner scaffolds in Betty's Brain; (2) Conducting a classroom study and doing data analysis to evaluate the impact of the delivered adaptive scaffolds on helping students to develop cognitive and metacognitive strategies or improve their emotions while engaged with model-building tasks in the learning environment.

For the *scaffold design and implementation* component of this work, six major theoretical models of SRL were first reviewed to understand the connections between cognitive, affective, and metacognitive processes during learning (Chapter II). Empirical evidence from prior research on learner behaviors in Betty's Brain (Chapter III) was used to further interpret students' cognitive-metacognitive strategy use and cognitive-affective behaviors in the context of their causal model-building tasks and activities within the learning environment. This formed the *conceptual framework* for scaffold design. The design was then operationalized in Betty's Brain using (a) a *strategy detection* framework that identified patterns of ineffective strategy use by keeping track of changes in learner activities and performance, and (b) a set of *conversational scaffold trees* that delivered *in-the-moment actionable feedback* to the learner, first to foster an awareness and understanding of the current state of their models, and then to support the development of more effective cognitive strategies, metacognitive behaviors, and positive emotions as they learn (Chapter IV).

For *scaffold evaluation* (Chapter V), we ran a classroom study with 55 middle school students who worked on Betty's Brain equipped with the adaptive scaffolding framework, to build causal models of a climate change process. The data collected from this study was first explored to find clusters among students based on differences in their learning and model-building behaviors. This exploratory analysis facilitated the formulation of more targeted research questions for scaffold evaluation. The impact of adaptive scaffolds on students in each cluster/group was then assessed by tracking the temporal changes in their cognitive-metacognitive behaviors, model-building performance, (and their affective states, based on data availability) after scaffolding. Results from this analysis (Chapter VI) show how learners in the four groups differed in their responsiveness and strategic use of different scaffolds provided to support their self-regulation behaviors for tasks like knowledge refinement, model debugging, or model assessment in Betty's Brain.

I.3 Primary Contributions

The primary contributions of this dissertation can again be summarized in two directions.

1. *Design and Development Contributions:* This dissertation presents a novel framework for the design and development of adaptive scaffolds to support the use of SRL strategies in an OELE. As outlined in Section I.2, this design builds on and extends the prior research in this field, is driven by a conceptual framework, and is operationalized by strategy pattern detectors and timely agent-based conversational feedback. Additionally, a number of design decisions for the current adaptive scaffolding framework, including the optimal scaffold triggering conditions and the content of scaffolds delivered at these

conditions, have been based on the results of a design-based research (DBR) process that spanned five years of research and included iterative cycles of scaffold design and evaluation (Sections III.2 and IV.4.3).

2. *Research Contributions:* The findings in Chapter VI of this dissertation demonstrate the effectiveness of the current iteration of adaptive scaffolding in Betty's Brain. These results show the differences in responsiveness and strategic use of scaffolds by four different groups of learners derived using a clustering approach (disengaged students, inefficient information generators, strategic map builders, and experimenters or tinkerers), suggesting how the different adaptive scaffolds may have helped some of these groups of students invoke and apply the intended self-regulation strategies while having less impact on others. We also discuss how these results present an opportunity to further improve scaffold adaptivity and contextualization to address the nuanced behavioral differences observed among these groups.

   Additionally, the path towards developing this dissertation research has involved multiple phases of generative and evaluative research, with each phase resulting in significant contributions to the state-of-the-art, in the form of peer-reviewed publications that expand our understanding of the SRL processes in OELEs and how adaptive scaffolds should be developed to support these processes. Some of the more specific topics explored in these research papers include: understanding learners' cognitive-affective interactions in OELEs (Munshi et al., 2018c,b), modeling their temporal behaviors and performance (Rajendran et al., 2018b; Munshi et al., 2022b), mapping their achievement and basic emotion states during learning (Munshi et al., 2020), as well as a gradually improved understanding of the impact of adaptive scaffolds on cognitive and metacognitive behaviors (Munshi et al., 2022b,a). (A more complete list of co-authored publications relevant to the content of this dissertation is provided in Appendix A).

I.4 Organization of the Dissertation

The rest of this dissertation document is organized as follows. Chapter II presents a literature review on SRL, OELEs, learner modeling, and adaptive scaffolding; all major aspects of our theoretical framework to inform scaffold design and development. Chapter III presents an overview of the Betty's Brain OELE and discusses the research studies using Betty's Brain that have contributed to the current scaffold design. Chapter IV presents the design and implementation procedure for adaptive scaffolding in Betty's Brain, including the theoretical framework, and scaffold triggering and delivery approaches. Chapter V outlines the approach for scaffold evaluation, including the study design, data collection, data analysis methodology, and research questions. Chapter VI reports and discusses the findings from the data analysis for scaffold evaluation, and Chapter VII summarizes the contributions of this research. Appendix A contains a list of relevant publications, while Appendix B and Appendix C contain supplementary tables, figures and other material used in this dissertation.

CHAPTER II

Literature Review

This chapter reviews the major components of the theoretical framework that have informed the design and development of the adaptive scaffolding framework to help students develop self-regulated learning (SRL) behaviors and strategies in Betty's Brain. The first step to detecting and modeling students' SRL behaviors is to understand the SRL process. **Section II.1** reviews (a) how the understanding of SRL as a learning construct has evolved with time, (b) the contributions of major theoretical models towards identifying the CAMM components of SRL, and (c) the relationships of CAMM components and their implications on modeling and scaffolding SRL. Next, **Section II.2** studies open-ended learning environments (OELEs) to better understand the problem-solving space where we detect, model, and provide scaffolds for SRL processes. **Section II.3** reviews the learner modeling literature to frame the design of our learner modeling approach in the Betty's Brain OELE. This section also discusses methods to diagnose learners' CAMM processes to construct a learner model that supports SRL. Finally, the scaffolding literature is reviewed in **Section II.4** to (a) identify the factors involved in the design of successful learner scaffolds and (b) study how adaptive scaffolds can be delivered to support SRL processes in OELEs. The literature reviewed in this chapter informs the design of our adaptive scaffolding approach in Betty's Brain, while also establishing the scope and contributions of our approach beyond the current state-of-the-art.

II.1 Self-Regulated Learning

Self-Regulated Learning (SRL) refers to learners' ability to understand and control their learning behaviors and their environment to accomplish their learning and problem-solving goals. Panadero describes SRL as a conceptual framework, which forms *"an extraordinary umbrella under which a considerable number of variables that influence learning (e.g., self-efficacy, volition, cognitive strategies) are studied within a comprehensive and holistic approach"* (Panadero, 2017). This view succinctly explains the importance of SRL in students' learning process.

SRL emphasizes autonomy, self-monitoring, control, reflection, and intrinsic motivation (Panadero, 2017). Zimmerman's 1994 model of SRL defined the term as involving "*goal-directed activities that students instigate, modify, and sustain*" (Zimmerman, 1994). In 2000, Borkowski et al. defined self-regulated learners as the students who are "*metacognitively, motivationally, and behaviorally active in their own learning*" (Borkowski et al., 2000). More recently, the importance of SRL as an important construct for effective learning has been emphasized by multiple researchers (Dinsmore et al., 2008; Plass et al., 2015; Verpoorten et al., 2009) as well as by education policy-makers who have asserted it as *"a critical skill for staying relevant and advancing in a rapidly changing world"* (U.S. Department of Education, 2016).

II.1.1 SRL: From *trait* to *process*

Early models of SRL defined the construct as a static *"trait"* that can be assessed using self-report instruments such as the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich et al., 1993). Panadero's review of SRL models describes this as the *first wave* in SRL measurement Panadero (2017).

But the end of the 1990s and the publication of the SRL handbook in 2000 (Boekaerts et al., 2000) saw the beginning of a *second wave*, with the research consensus on SRL shifting from a static *trait* to a dynamic *"process"*, or a sequence of events that interact and evolve as students learn (Panadero et al., 2016). This switch in the conceptualization of the term implied that self-report measures could no longer suffice as an approach for measuring SRL. Prominent SRL researchers updated their earlier models to reflect this newer understanding of self-regulated learning as an entity that changes with time, as students progressively complete their tasks and approach their learning goals by planning, monitoring, and reflecting upon their learning process (Zimmerman, 2001; Winne, 2001; Winne and Hadwin, 2008). Measuring SRL in learning environments now required the development of 'online' measures that can track this temporal evolution of student activities during their learning process.

II.1.2 Primary Models of SRL

Panadero (2017) discusses six major models of SRL, which have been developed by Zimmerman; Boekaerts; Winne and Hadwin; Pintrich; Efklides; Hadwin et al.. This section reviews these prominent SRL models to derive a better understanding of the major factors involved in this complex learning process.

In 1989, **Zimmerman**, one of the pioneers in the field of SRL measurement, proposed his earliest model of SRL (Zimmerman, 1989). Known as the Triadic Analysis model, it explored SRL from a socio-cognitive perspective (where knowledge acquisition is governed by social interactions), and described three forms of SRL - at the *self* (person), *behavior* and *environment* levels. Later in 2002, Zimmerman proposed his Cyclical Phases model (Zimmerman, 2002a), which delved into the interaction of metacognitive and motivational processes at the individual level. The Cyclical Phases model is the most prominent and well-validated model proposed by Zimmerman, and characterizes SRL as a repetitive cyclical process composed of three phases: (1) *forethought* - analyzing a task, setting goals, and generating plans to solve the task, (2) *performance* - executing the task by applying strategies and monitoring their progress, and then (3) *self-reflection* - assessing and reflecting on the outcomes (Zimmerman, 2002a). Zimmerman asserted that self-regulated learners who adopt this dynamic, cyclical learning process are self-aware, knowledgeable, decisive about their learning, intrinsically motivated, and capable of monitoring and controlling their learning behaviors and environment. The more recent version of Zimmerman's Cyclical Phases model (Zimmerman and Moylan, 2009) includes *metacognitive strategies* in the performance phase.

Another early SRL researcher, **Boekaerts**, proposed two major models of SRL. Her six-component model (Boekaerts, 1996) considered *cognitive* and *affective/motivational* self-regulation to be the two basic mechanisms of SRL. The Dual Processing model (Boekaerts and Corno, 2005; Boekaerts, 2011) emphasized the important role played by positive and negative emotions in SRL, and described strategies for affect regulation during learning.

Winne and Hadwin's Information Processing Theory (IPT)-based SRL model (Winne and Hadwin, 1998) explores SRL from a *metacognitive* perspective, by recognizing that self-regulated learners are active participants in their learning process; they exercise agency and manage their learning by the use of cognitive and metacognitive *strategies*. This model is especially relevant in the context of open-ended learning environments (OELE) that offer learners a greater possibility to exercise their agency during learning. Winne and Hadwin's model asserts the goal-driven nature of SRL and posits that learning occurs in four linked and recursive phases: (1) *task definition*, (2) *goal setting and planning*, (3) *applying study tactics and strategies*, and (4) *adaptations to metacognition*. Each phase is described in terms of the interactions among five "COPES" facets of tasks : (1) ***C**onditions* - available resources and task/environment constraints; ***O**perations* - cognitive processes and strategies applied to complete the task; ***P**roducts* - new knowledge generated by these operations; ***E**valuations* - generated internally by the student or provided externally through feedback, by comparing the products with respect to (5) specified ***S**tandards* or criteria.

A *strategy* in the COPES model is defined as a collection of if-then rules (also known as tactics) that form larger if-then-else rule patterns over time (Winne and Hadwin, 2008). Strategies are more complex than tactics in their structure, and they also have a larger scope that yields more information that can be used as feedback during learning. In the recursive loop defined in the COPES model, a learner first activates the memory of previous strategy use in the *task definition* phase, and then the strategy is linked to specific learning goals in the *goal setting and planning* phase. Following that, in the *enactment* phase, the learner uses linked strategies to address the learning goals. Finally, in the *adaptations to metacognition* phase, the learner evaluates the effect of the strategy used, and then tunes or restructures the strategy to make it more effective for the goals and plans of the learning task (Winne and Hadwin, 2008).

Winne and Hadwin's COPES framework explains the cognitive processing of information during different phases of learning, viz., planning, strategizing, metacognitive monitoring, and reflection. While Winne's earlier models do not allude to emotions during self-regulation, later versions of Winne's model (Winne and Hadwin, 2008), as well as a review by Greene and Azevedo (2007) recognize affect regulation as an important component of SRL processes.

Pintrich's SRL model (Pintrich, 2000) considers four areas for regulation of learning: (1) Cognition; (2) Motivation/Affect; (3) Behavior; (4) Context. An important focus for this model is the regulation of cognition, which can be achieved by setting target goals, activating prior knowledge, metacognitive awareness, and monitoring of cognition (through *judgment of learning (JOL)* and *feeling of knowing (FOK)*) and the selection and adaptation of cognitive strategies for learning. Motivation or affect regulation is dependent on the activation of task value, the selection and adaptation of strategies for managing affect, and the affective reactions and attributions during self-reflection.

Efklides' Metacognitive and Affective Model of Self-Regulated Learning (Efklides, 2011), similar to Winne's model, is also backed by metacognition research and explores how emotion and motivation interact with metacognition in SRL. This model describes (1) a "Person" level of self-regulation that is composed of *cognition, motivation, self-concept, affect, volition, and metacognition (as metacognitive knowledge and metacognitive skills).* The "interactions of the person's competencies, self-concept in the task domain, motivation, and affect, vis-à-vis the perception of the task and its demands" determine the effort they invest in their cognitive processing. While the Person level represents the general trait-oriented features of SRL, similar to Zimmerman's model, the Task × Person level in Efklides' model describes more person-oriented and less conscious actions that are more similar to Winne's model (Panadero, 2017; Winne, 2011). Efklides identifies four functions at the Task x Person level: (a) cognition, (b) metacognition, (c) affect, and (d) regulation of affect and effort.

The sixth major model of SRL, proposed by **Hadwin, Järvelä, and Miller**, explores the social aspects of the regulation of learning (Hadwin et al., 2011), for example, in computer-supported collaborative learning situations. This model contrasts self-regulation (SRL) to other regulation modes that exist in collaborative settings, viz., co-regulation (CoRL) and socially shared regulation (SSRL).

A critical analysis the SRL models discussed above suggests that *most of these models allude to four primary areas* for self-regulation of learning:

1. **Cognition**: The use of prior knowledge, skills, and strategies to develop solutions for the learning task (Entwistle and Ramsden, 2015; Pressley et al., 1992). The role of cognition in the SRL process is discussed in the SRL models of Zimmerman; Winne and Hadwin; Boekaerts; Pintrich; Efklides;
2. **Affect**: The ability to identify and regulate one's emotional reactions and become an effective learner. Affect as a component of SRL is discussed in the SRL models of Boekaerts; Pintrich; Efklides and later models of Winne and Hadwin;
3. **Metacognition**: The ability to decompose a complex task into sub-tasks (i.e., sub-goals), apply strategies to develop solutions, monitor progress toward completing the sub-tasks, and periodically reflect on

how to improve one's performance (Pintrich, 2002; Schraw et al., 2006). This is an important aspect of Winne and Hadwin's model and also discussed by Pintrich; Efklides; Zimmerman and Moylan;

4. **Motivation**: The perceived value of the learning task and the subject matter being learned (task value), as well as the self-perceived ability to accomplish the task (self-efficacy) and one's personal goals (intrinsic versus extrinsic) for doing the task (Pintrich, 1999; Schunk and Zimmerman, 2012). This is discussed in the SRL models of Zimmerman; Boekaerts; Winne and Hadwin; Pintrich; Efklides.

Other researchers (Aleven and Koedinger, 2002) also consider the regulation of cognition, affect, metacognition, and motivation (collectively referred to as "CAMM" processes) to be integral to the process of self-regulated learning. Azevedo et al. (2015) discuss emerging empirical evidence that also suggests links between CAMM processes and SRL, especially when learning using advanced learning technologies (viz., intelligent tutoring systems, hypermedia environments, etc). This helps us characterize SRL as a learning process made up of a dynamic sequence of interacting CAMM events (Azevedo et al., 2015, 2017; Bannert et al., 2017).

II.1.3 CAMM Relationships in SRL and their Implications

SRL models suggest that learners' CAMM regulation processes are interrelated and these relations may influence their overall learning regulation and performance. Therefore, we study CAMM relationships and their implications on modeling and supporting CAMM processes in learning environments.

For instance, when we look at cognition and metacognition, it is clear that these two CAMM processes are intertwined. Winne (1995) characterizes *cognition* as dealing with the knowledge of "objects" or skills and operations of objects, whereas *metacognition* is the corresponding meta-level that monitors and evaluates the use of cognitive processes and modifies them if necessary. Since cognitive processes are more situation-specific, they are easier to detect from observable information in a learning environment. Metacognitive monitoring processes are more internal and reflected by learners' changes to their cognitive processes and strategies. Therefore, to model and support learners' cognitive-metacognitive processes in our framework, we will have to track the changes in their cognitive processes during learning and infer underlying internal metacognitive monitoring and self-reflection processes.

Learners' motivation and affect regulation processes are also strongly related to their cognitive and metacognitive processes during learning. Pintrich (2000)'s SRL model suggests that learners' regulation of motivation or affect is dependent on their perceived value of the learning task, the selection, and adaptation of strategies to manage affect, and the affective reactions and attributions during metacognitive self-reflection processes. Therefore, to help students manage a sudden decrease in motivation or the onset of a possibly harmful emotion such as frustration or boredom (D'Mello and Graesser, 2012), we need to track changes in their motivational/affect states and also monitor the (cognitive-metacognitive) attributions for a change in motivation or an affect appraisal. For example, D'Mello and Graesser (2012) suggest that the affect state of *confusion* signals the detection of an impasse by the learner, which may be attributed to an external cause such as a difficulty encountered while interacting with the learning material, or an internal cause such as the lack of sufficient prior knowledge and effort. These attributions are likely to impact how a confused learner regulates their confusion. Sometimes, learners are unable to resolve the learning difficulties that cause an emotion like confusion. This unresolved confusion can then transition into a state of frustration which may be more difficult for the student to regulate. Therefore, a learner modeling framework to support the affective component of self-regulated learning should (1) detect potentially harmful affect appraisals such as frustration (signifying an impasse or obstacle in learning) or boredom (signaling disengagement with the task), and (2) evaluate these emotions in the context of their (cognitive-metacognitive) behaviors and task performance to determine the affective attributions and thereby trigger appropriate scaffolds that can help resolve deficiencies in strategy use.

Motivation, much like metacognition, is an internal process that is difficult to detect from observable information in a learning environment. But learners' affect states may provide insights into motivational factors such as their perceived value of the learning task. The observation of an affective state like boredom often suggests hopelessness and disengagement from the task (D'Mello and Graesser, 2012), further suggesting that the learner is demotivated and maybe attaching a low value to the task. Therefore, feedback to help such students regulate their affect can also improve the motivational aspect of their SRL process, for instance if the feedback includes positive reinforcement or encouragement that specifically intends to improve task motivation.

The relations between learners' CAMM regulation processes (viz., cognition and metacognition, affect and cognition/metacognition, affect and motivation) strengthen our understanding of these processes, and how they can be modeled and supported by our adaptivity framework to support students' overall self-regulated learning processes in the learning environment. The next section studies open-ended learning environments or OELEs, which form the setting where we design and develop our learner modeling and adaptive scaffolding approach to support SRL.

II.2 Open-ended Learning Environments

Open-ended learning environments (OELEs) are a class of computer-based learning environments (CBLEs) that use the constructivist theory of learning to support the acquisition of knowledge and skills (Land, 2000). OELEs provide learners with opportunities to practice their problem-solving skills in real-world contexts (Wang and Hannafin, 2005). A learning environment is termed to be *"open-ended"* if the learner has the freedom to choose "the learning goal, the means to support learning, or both" (Hannafin et al., 2014). Therefore, OELEs are specifically designed to facilitate student agency. Furthermore, OELEs may provide tools and resources that engage learners in activities like generating hypotheses by knowledge acquisition, constructing solutions, using tests to verify the hypotheses, and revising hypotheses in different phases of learning (Land, 2000). Some prominent OELEs have been reviewed below, including *Ecolab* (Luckin and du Boulay, 2016), *MetaTutor* (Azevedo et al., 2010), *nStudy* (Winne and Hadwin, 2013) and *Betty's Brain* (Leelawong and Biswas, 2008; Biswas et al., 2016).

Ecolab is a family of constructivist learning environments for learning ecology, that focus on middle school science topics like food chains and food webs (Luckin and du Boulay, 2016). Learners using Ecolab are provided with a simulated environment where they are free to select different organisms on the food chain and explore relationships between these organisms without having to deal with the complexity of the entire food web. A variation of the environment named M-Ecolab adapts to the learner's goal orientation to determine the form of scaffolding provided to the student (Zhang et al., 2021a)). The Ecolab environment includes scaffolds that support metacognitive monitoring and task selection processes in learners with low prior knowledge or metacognitive skills (Luckin and Hammerton, 2002). Some versions of Ecolab attempt to operationalize Vygotsky (1978)'s zone of proximal development (ZPD) framework by providing flexivle and tailored assistance (Luckin and du Boulay, 2016). **MetaTutor** is an OELE where students can learn about complex topics in biology such as the human circulatory system. Learners have access to resource pages to acquire information on the topic and are given a fixed period of time to generate a summary based on the acquired information. The OELE supports help-seeking behaviors from pedagogical agents present in the system. There are four types of agents in MetaTutor: Gavin the guide, Pam the planner, Mary the monitor, Sam the strategizer. Each agent is assigned to scaffold the development and use of a specific SRL process, such as planning, monitoring, and active strategy use (summarizing, note-taking, inferring). (Azevedo et al., 2010). MetaTutor provides opportunities to the student to explicitly state their learning task, and uses this choice of task as the context to adapt the feedback given to the student. **nStudy** is a web-based application that leverages Winne and Hadwin's COPES framework of SRL (Winne and Hadwin, 2008) and offers a toolkit for learners to practice their self-regulated learning skills while studying in a digital environment (Winne and Hadwin, 2013). Learners are provided the agency to define their own learning strategies and link them to their learning artifacts, such as bookmarks and notes. They can also evaluate their strategies (Zhang et al., 2021a). Learner behaviors during reading or annotating in different phases of learning reflect cognitive and metacognitive events during learning (Beaudoin and Winne, 2009).

The **Betty's Brain** OELE adopts a *learning-by-teaching* paradigm to help middle school students learn science by building *causal models* of scientific processes to teach a virtual pedagogical agent, generically named Betty (Biswas et al., 2005; Leelawong and Biswas, 2008). The system provides students with a number of resources and tools to learn, build, and check their models. Students are free to combine these tools in different ways and form their own learning strategies to teach a causal model to Betty. (Since our adaptivity framework is built in Betty's Brain, we discuss the learning tools and problem-solving process in this learning environment in more detail later in Chapter III).

The review of learning tools and features of these OELEs illustrate how open-ended learning environments provide an exploratory problem-solving space, where learners can accomplish their learning tasks using their own learning strategies. In the next section, we study how learner modeling can help us diagnose and model learner proficiency in the use of their learning skills and strategies, and thereby detect deficiencies in strategy use in OELEs.

II.3 Learner Modeling

Learner modeling is a component of computer-based learning environments (Brusilovski et al., 2007; Basu et al., 2017; Mitrovic, 2012; Azevedo et al., 2005) that allow researchers and educators to capture learners' domain knowledge, cognitive skills, strategies, and interests in a systematic manner to tailor their experience in the learning environment. Learner modeling approaches were initially conceived for intelligent tutoring systems (ITS). Self (1998) discuss a tripartite architecture for an ITS that is composed of: (1) a *domain knowledge* module that contains knowledge of the domain under consideration (e.g., ecology in case of Ecolab or climate change in Betty's Brain); (2) a *student model* that contains knowledge about the student who is interacting with the system; and (3) a *tutoring strategy* or pedagogical module that contains knowledge about how the system should interact with the student. Self emphasizes the importance of learner modeling by suggesting that learning environments "which care about what the student knows, misunderstands, wants to do, etc." need to incorporate a student modeling structure (Self, 1998).

However, despite the emphasis on learner modeling to understand and support students in learning environments, building learner models that capture students' misconceptions or flaws, also known as *perturbation based models* (Chrysafiadi and Virvou, 2013), is a challenging task. This type of learner modeling requires an ongoing diagnosis of students' activities, behaviors, and performance to build and update structures of student knowledge and proficiency in their tasks. Elsom-Cook (1993) discuss 'learner-based methods' of learner modeling that do not intend to simply turn a learner into a copy of an expert but instead place an emphasis on why and how the learner acquires a piece of knowledge, how the relationship between their pieces of knowledge evolves over time, and how they may benefit from receiving an intervention. Such a nuanced approach to student modeling is necessary to model learner behaviors in open-ended problem-solving spaces such as OELEs.

II.3.1 Learner Modeling Approach in OELEs

Learner modeling schemes have been developed in OELEs such as CTSiM (Basu et al., 2017). The learner model in CTSiM adopts a task-oriented approach combined with a strategy modeling framework to capture students' cognitive and metacognitive processes during learning. Basu et al. (2017) show how this learner modeling scheme facilitates the development of adaptive scaffolds to support learner behaviors in CTSiM.

To construct an efficient learner modeling framework to support SRL (CAMM) processes in Betty's Brain, we build on Basu et al. (2017)'s learner modeling scheme to detect students' cognitive-metacognitive strategies during learning and adapt this scheme to account for task contexts specific to the Betty's Brain learning environment. We also use affect detector models trained on Betty's Brain data (Jiang et al., 2018) to predict states like frustration or boredom during learning, and diagnose the potential cognitive attributions of these emotion appraisals, considering the type of affective-cognitive/metacognitive relationships discussed in Section II.1.3.

The major objective of this process is to diagnose difficulties in students' cognitive-metacognitive strategy use and affect regulation processes and respond with appropriate scaffolds to improve their SRL process. However, this requires the online diagnosis of CAMM regulation difficulties during learning, which comes with its own set of challenges. In the next section, we explore these challenges and methods that have been developed to address them.

II.3.2 Online Detection of CAMM Processes for Learner Modeling

The online detection of CAMM processes *as students learn* to support a learner modeling framework in computer-based learning environments is often a difficult task (Azevedo et al., 2017; Taub et al., 2014), and measuring each type of CAMM process in an OELE comes with its own unique challenges. Tracking students' internal processes in an online setting is difficult. Wittgenstein (1968) said, *"An inner process stands in need of outward criteria"*.

Affect states exhibited by a learner, such as boredom, can sometimes help us determine their (lack of) internal motivation in the task and support the provision of scaffolds to improve affect and engage learners, as discussed in Section II.1.3. But while affect states are externally observable through students' emotional reactions, the online detection of affect states in learning environments is still a difficult task. This is because: (1) retrospective self-report measures of affect rely heavily on the accuracy of student memory (Boekaerts and Corno, 2005); (2) reflection methods, such as learning diaries may annoy students by interfering with their learning activities; and (3) logged activity traces, which capture students' cognitive and strategic behaviors, do not directly reveal the learner's affective processes.

To address this problem, Jiang et al. (2018) developed fine-grained detectors of learners' achievement emotion states, like frustration, confusion, engagement, delight or boredom, from their interactions with the Betty's Brain OELE. We use these affect detectors to predict students' emotions and inform the affect modeling component of our learner model. However, we concede that developing these affect detector models requires features that are specific to a learning environment, so the models are not easily transferable to other systems. Additionally, these models predict *emotion likelihood* based on learner activities, so they may not always be as accurate as human-coded affect labels and may need to be validated against other affect sources or a detection of associated cognitive states.

Facial affect detection approaches have also been developed, which use features built from a person's facial action units to build more robust affect detector models; however, since most commercially available and well-trained facial affect detection models predict universally applicable *basic emotions* (Ekman and Friesen, 1978), viz., joy, sadness, anger, etc., instead of academically relevant *achievement emotions*, the use of such models in academic settings is limited. In Munshi et al. (2020), we mapped achievement emotions predicted by Jiang et al. (2018)'s affect detectors to basic emotions predicted by commercial face detection software (McDuff et al., 2016), to support the use of robust face detection models for predicting achievement emotions; however, employing face detection mechanisms in online settings still remains a challenge. Therefore, we use the affect detector models by Jiang et al. (2018) to predict learners' emotions in our design framework. We track the temporal sequences of emotions during learning in Betty's Brain to identify shifts towards the more negative emotion states like frustration or boredom (D'Mello and Graesser, 2012) and detect associated cognitive behaviors and task performance (Sections II.1.3;II.3.1) to develop a more complete understanding of the student's current cognitive-affective state while working on the learning task.

The detection of cognitive and metacognitive strategies and processes is facilitated by the availability of activity logs in learning environments. Researchers have developed model-driven (Kinnebrew et al., 2017) and data-driven (e.g., differential sequence mining Kinnebrew et al. (2013a)) approaches to derive and interpret patterns of students' strategic behaviors by distilling contextualized information from raw activity traces in OELEs. These methods have been used in a number of experimental studies to analyze learners' productive and unproductive strategy use in OELEs like Betty's Brain and CTSiM (Biswas et al., 2016; Munshi et al., 2018b; Zhang et al., 2021a). EEG/HR bio-sensor devices (Yun et al., 2019) and eye-gaze tracking devices (Rajendran et al., 2018a) can be used to derive further measures that augment our understanding of students' cognitive and metacognitive processes in learning environments. In the learner modeling framework presented in this dissertation, we adopt a strategy detection approach to detect sequences of students' logged activity traces and analyze these activity sequences using Kinnebrew et al. (2017)'s task model to infer the use of task-oriented cognitive and metacognitive strategies.

As discussed earlier in this section, our objective of modeling learners' CAMM processes during learning is to detect self-regulation difficulties and trigger appropriate scaffolds to respond to these difficulties. Scaffolding can be an effective way to respond to learners' SRL difficulties in an OELE. In the next section, we study the parameters for designing successful learner scaffolds.

II.4 Parameters for Successful Scaffold Design

Scaffolds have been described as "*tools, strategies, and guides used during learning to enable the development of understanding(s) beyond one's immediate grasp*". These include pedagogical support that is calibrated to a learner's current level of understanding and helps them accomplish tasks that they could not accomplish alone (Wood et al., 1976). This difference between what a learner can accomplish by themselves versus with assistance from external sources is typically characterized by the 'zone of proximal development' (ZPD) (Vygotsky, 1978). The ZPD informs how students' range of understanding and problem-solving can be extended in a way that provides sufficient challenge, while preventing frustration or boredom. So, learners should ideally be scaffolded in their ZPD, to close the gap between their current and potential planes of development (Vygotsky, 1978).

Researchers like Elsom-Cook (1993); Puntambekar and Hubscher (2005) further point out certain key features of effective learner scaffolds. These include *inter-subjectivity* (i.e., a shared understanding of the activity and goals), *ongoing diagnosis* (i.e., knowledge of the task, and subtask routines, as well as the student's current level of understanding), *tailored assistance* (i.e., providing help when needed), and *fading* (i.e., gradually removing the assistance).

Self (1988) breaks down the primary applications of student scaffolds into six categories: (1) *corrective* - helping students correct their learning/problem-solving errors (2) *elaborative* - helping students gain the knowledge that they lack, (3) *strategic* - helping students invoke a known procedure or piece of knowledge that they are unable to apply appropriately, (4) *diagnostic* - using inference procedures to understand where students may lack abilities, (5) *predictive* - predicting how a student is likely to respond in a specific learning situation, and using this prediction to inform the hints/feedback given, and (6) *evaluative* - providing a comprehensive assessment of the level of knowledge and achievement of the student.

In view of the above considerations for effective scaffold design, we further frame the current scaffolding approach in Betty's Brain, to design scaffolds that can use an online diagnosis of students' learning process (via modeling their use of cognitive-metacognitive strategies and affective experiences) to provide tailored assistance, as suggested by Elsom-Cook (1993); Puntambekar and Hubscher (2005)). The scaffolds need to be flexible, meaning that the level of adaptivity may be varied based on the learner's proficiency in their tasks, to support the learner in their zone of proximal development (Vygotsky, 1978). The content of the scaffolds in our case are primarily strategic, but depending on the triggering condition may also include one or more of the diagnostic, predictive or elaborative features suggested by Self (1988).

II.5 Adaptive Scaffolds to Support SRL in OELEs

There is evidence of the positive effect of SRL strategies on students' academic achievement (Wolters and Hussain, 2015; Zimmerman and Pons, 1986), especially when learning complex topics in OELEs (Azevedo et al., 2017; Winne, 2017). So, researchers have developed methods to support learners' SRL process in OELEs (Aleven et al., 2003; Azevedo et al., 2010; Biswas et al., 2016; Narciss et al., 2007; Sabourin et al., 2013; Winne et al., 2010). Adaptive scaffolds have been found to be an especially effective method for supporting SRL processes in OELEs (Belland et al., 2011; Segedy et al., 2013; Basu et al., 2017; Taub et al., 2014; Munshi et al., 2022b).

While the study of successful scaffolding practices in Section II.4 helped us to frame the major characteristics of our adaptive scaffolding approach to improve SRL processes in Betty's Brain, we needed to determine a method to deliver these scaffolds in a way that is meaningful and engaging to the learner receiving the feedback.

Biswas et al. (2016) discuss how a multi-agent architecture in Betty's Brain equipped with a listener interface can allow explicit communication between multiple actors, viz., the student, the teachable agent (Betty), and the mentor agent (Mr Davis) in the learning environment. Segedy et al. (2013) use this communication framework to deliver conversational student scaffolds that are embedded as agent-initiated dialog using a conversation tree representation (Adams, 2010). Furthermore, the feedback is contextualized to Betty's current causal map and the student's recent interactions with the system (Biswas et al., 2016), making it particularly useful and relevant for the student. Munshi et al. (2022b) and Munshi et al. (2022a) also use the conversation tree representation to deliver contextualized conversational scaffolds in Betty's Brain (Figures **??**; B.2), and the evaluation of these scaffolds further support the scaffold design approach presented in this dissertation, as outlined later in Section IV.4.3.

In this dissertation, we are extending Segedy et al. (2013)'s previous work on conversation trees by mapping scaffold triggering conditions detected by our learner modeling and strategy detection framework to generate conversations with the learner. The content of the trees would reflect our understanding of the most appropriate and task-specific guidance that can be provided to help the learner overcome deficiencies in their cognitive-metacognitive strategies or affect regulation processes at specific moments during learning.

II.6 Critical Summary and Motivation for this Dissertation

To summarize, the literature reviewed in this chapter helped us understand self- regulated learning (SRL) as a dynamic process composed of interacting cognitive, affective, metacognitive, and motivational events during learning. To support the different dimensions of SRL as learners build and debug their causal models in an open-ended problem-solving environment like Betty's Brain, we need to build a learner modeling approach that can store (and update) the evolving self-regulation profiles of students by performing an online detection of the cognitive processes and strategies that students apply to their tasks, infer underlying metacognitive conditions, and also attempt to detect their emotional states. Students' self-regulation difficulties, diagnosed from their learner model as *ineffective* cognitive-metacognitive strategy use (that lead to a drop in task performance) or potentially harmful cognitive-affective states, would then trigger scaffolds that respond to the difficulties by adapting to the student's current level of understanding, and by delivering conversational feedback that is strategic, flexible and engaging to the learner. We present a complete description of our adaptive scaffolding design framework in Chapter IV. Prior to that, Chapter III discusses the Betty's Brain OELE where this design is implemented, and prior research studies with Betty's Brain that have led to the current design framework.

CHAPTER III

The Betty's Brain Open-Ended Learning Environment

III.1 System Overview and Features

Betty's Brain (Biswas et al., 2005, 2016) is an open-ended learning environment (OELE) that helps students acquire knowledge and understanding of scientific phenomena, such as climate change and human body thermoregulation. To learn about a science topic in Betty's Brain, students construct *causal* models that depict the cause-and-effect relationships between different concepts involved in the topic.

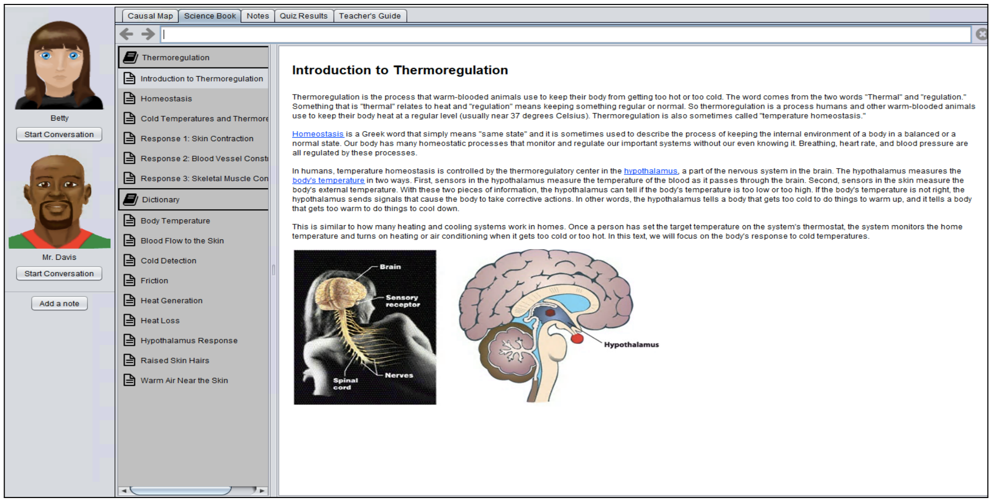
As discussed briefly in Chapter II, the system adopts a *learning-by-teaching* paradigm (Leelawong and Biswas, 2008), where students construct the causal map to teach a virtual *teachable agent*, generically named Betty. As she is being taught a particular topic, for example, the causes and effects of climate change, Betty can answer queries from the student, such as, ”If deforestation increases, what will happen to the amount of heat trapped by the earth?”. To answer the question, Betty uses the current causal map she has been taught (by the student) to follow a succession of causal links and derive her answer to the question.

A series of experimental studies in middle-school classrooms with the Betty's Brain software have demonstrated that (a) students achieved significant pre- to post-test learning gains on the science content (Leelawong and Biswas, 2008; Segedy et al., 2015; Munshi et al., 2018c, 2022b), and (b) students with higher learning gains and model scores (i.e., who had more correct links in their model) used more effective learning strategies (Kinnebrew et al., 2014; Munshi et al., 2018b, 2022b).

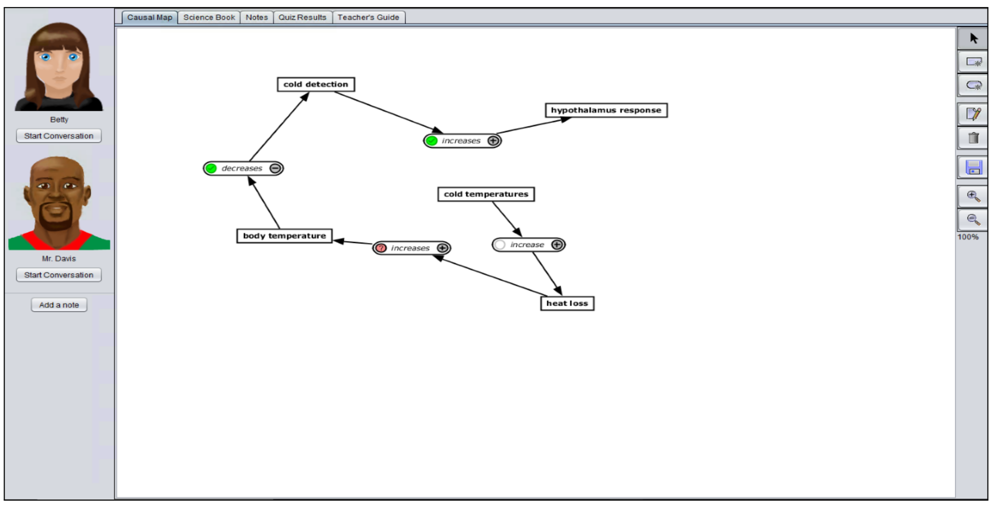
The primary learning tools in the Betty's Brain system are described below, and the user interfaces to these tools are illustrated in Figure III.1.

A **science book**, which is a set of hypermedia resource pages embedded within the system, helps students access the knowledge they need to learn and build causal models in the science domain. Students have to identify the required concepts and the causal (i.e., cause-and-effect) relations among these concepts to teach Betty by reading relevant sections of the science book. An accompanying **teacher's guide** provides students with information on procedures and strategies that they can apply to construct and check the correctness of their evolving causal maps.

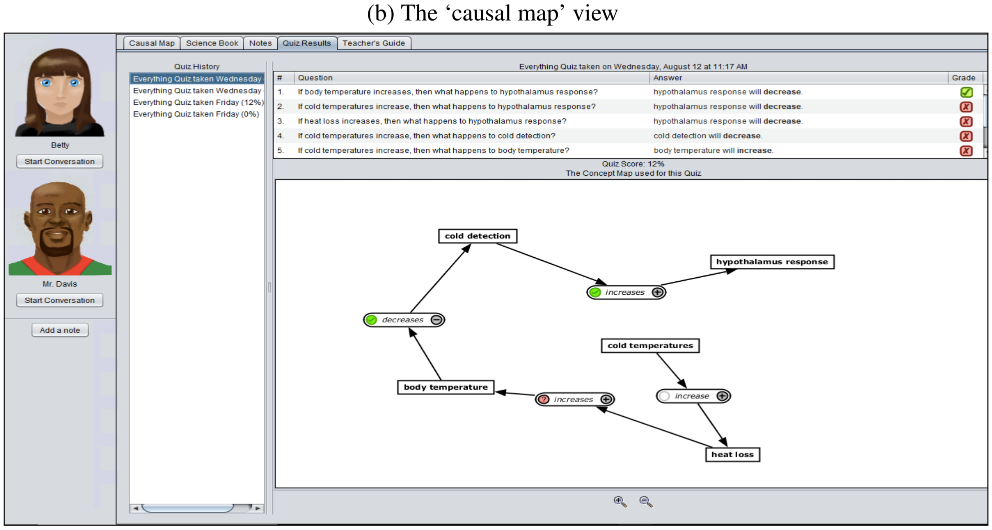
A **causal map building tool** includes a visual interface with a drag-and-drop menu to help the student build and edit the causal maps they are developing to teach Betty. The interface provides students with a visual representation of the current state of their causal map, and tools that they can use to add, delete, and modify the concepts and links that make up the causal map.



(a) The 'science book' view



(b) The 'causal map' view



(c) The 'quiz results' view

Figure III.1: System interface for the Betty's Brain thermoregulation unit

The **query and quiz tools** allow students to probe Betty's knowledge of the domain. The quiz tool requests Betty to take a quiz. Betty answers the quiz questions dynamically generated and scored by the **mentor agent**, Mr. Davis. The quiz results help students evaluate the correctness of the current causal map and they can use this information to make corrections to their map or go back to the resources to read further and gain more knowledge of the science topic. After the quiz results are displayed, students can also click on individual questions, and get Betty to explain her answer to the question. She does this by highlighting the links used to answer that question. This provides students with more details on the set of links used to answer that question, and ways in which they may further assess these links. Overall, the quizzes help students track Betty's progress in learning the domain, and by implication their own knowledge of the science concepts and relations needed to build the domain model. In addition to administering and grading the quizzes, Mr. Davis provides strategic feedback in the form of adaptive conversational scaffolds. Betty and Mr Davis also provide encouragement feedback intended to help the students regulate their affect and engagement during learning.

Overall, Betty's Brain is a socio-constructivist learning-by-modeling environment (Hickey, 1997). It offers tools to facilitate exploration, strategic thinking, and developing monitoring skills, as learners seek information using the "science book", build their cause-and-effect models in the "causal map", and check their models using "quiz" or "query" tools (Biswas et al., 2016) to help Betty to learn the science content. The mentor agent, Mr. Davis, is present to provide relevant timely feedback when students have difficulties in building and checking their maps. But due to the complex and differential nature of student learning in this open-ended and exploration-friendly environment, the system needs to have a very good understanding of learners' task progression for any scaffold from the mentor agent to be successful.

III.2 Betty's Brain Research Studies Leading to the Current Design of the Adaptive Scaffolding Framework

A set of classroom studies with Betty's Brain have been conducted over the last five years following a design-based research (DBR) approach. This research has culminated in the current design of the adaptive scaffolding framework presented in Chapter IV. Therefore, the findings from these studies are discussed briefly in this section and the implications of this DBR process on specific aspects of the scaffold design are further elaborated in Section IV.4.3.

The March 2017 Study: In March 2017, an experimental study was conducted with 93 students from two sixth-grade science classrooms of a public middle school in Nashville, TN. The objectives of this study were two-fold: (a) To collect Betty's Brain log data that could help us derive better measures of learners' cognitive and metacognitive behaviors, and (b) To collect labels of learners' affective states *(viz., delight, engaged concentration, confusion, frustration, boredom)* using *'BROMP'* (Ocumpaugh et al., 2015), a momentary time-sampling technique for trained observers to record affect labels of learners working on a computer-based learning environment in a classroom. The BROMP affect labels collected from the March 2017 study were aligned *post-hoc* to learners' activity sequences extracted from Betty's Brain logs. This data was then used for a number of research and development purposes, also outlined in this section. We first briefly discuss some of the main data analyses and research findings below.

Learners showed significant pre-to-post learning gains during the study. We determined behavioral patterns by analyzing frequent activity sequences using a differential sequence mining approach (Kinnebrew et al., 2013a) and studied these patterns and their relation to task performance to infer the use of *cognitive strategies* during learning. We also found that learners' affective states in Betty's Brain were linked to their cognitive strategies and causal modeling performance in the system. This provided empirical evidence from student interactions with Betty's Brain on the relations between cognitive and affective states discussed in Chapter II. The different findings from this study are reported in detail in **Munshi et al. (2018b)** and **Munshi et al. (2018c)**.

While the sequence mining method helped us derive fine-grained measures of learners' behaviors, a more broader picture of their temporal cognitive-metacognitive processes was obtained by applying a *process mining* approach, reported in **Rajendran et al. (2018b)**. Another major finding from our analyses using this data set was that learners' cognitive-affective behaviors in Betty's Brain were mediated by the feedback they received from the mentor agent *and* their causal modeling performance in the system. The findings, reported in Munshi et al. (2018c), helped us derive a better understanding of how to design more strategic agent-initiated feedback.

Beyond the research findings from data analysis, the data collected in the March 2017 study was also **used for developing tools** to better track and understand students' SRL processes in Betty's Brain. The first of these was the development of a set of **affect detector models** (Jiang et al., 2018) - binary classifiers trained on BROMP affect labels aligned to learners' activity sequences, which were then embedded into the Betty's Brain system. These *activity-based* BROMP-trained affect detector models, for use in future studies, generated likelihood values of five achievement emotion states *(delight, engaged concentration, confusion, frustration, boredom)* at a 20-second interval, using a sliding window of students' cognitive activity sequences within the Betty's Brain environment.

A **cognitive-metacognitive pattern detection module** was also developed in Betty's Brain using the data from the March 2017 study. This module used *regex-pattern matching* to detect *online*, the type frequent behavioral patterns identified from differential sequence mining in Munshi et al. (2018b)). The pattern detection module looked for online *inflection points*, or changes in learners' cognitive-metacognitive strategic behaviors, as they interacted with the Betty's Brain system.

The affective and cognitive-metacognitive detectors described above were built to students' affect states and changes or inflections in their cognitive-metacognitive processes in real time as they worked in Betty's Brain. We further developed a **message communication framework** to communicate key inflection points detected from student laptops to a Betty's Brain Ruby-on-Rails server. This server then processed and passed this information to a **Quick Red Fox** (QRF), a mobile app (Baker et al., 2021), which was used by a classroom researcher in the next study (Dec 2018, discussed below) to interview students *in-the-moment* and gain more insights on their *metacognitive processes and strategy use* at different key inflection points during learning. An initial design of a framework to scaffold SRL CAMM processes in Betty's Brain was formalized first in **Munshi and Biswas (2019)**.

The Dec 2018 & Feb 2019 studies: Next, the newly developed detectors and the Betty's Brain-QRF communication framework were deployed in two successive classroom studies, conducted in Dec 2018 and Feb 2019, with 99 sixth-grade students from a Nashville public school. The same students participated in both of these studies. In the Dec 2018 study, students worked on the "climate change" unit of Betty's Brain. In the Feb 2019 study, they worked on the "human thermoregulation" unit of the system. Additionally, the version of Betty's Brain used in the Feb 2019 study also included the **first design iteration of our adaptive scaffolding framework** (which included a set of *strategic hints* and *encouragements*, listed in Table B.1) that targeted an addressal of students' SRL difficulties at some key cognitive-metacognitive inflection points. Results from evaluating this design iteration prompted the next refined design of our scaffolding framework. Some of the primary research findings from the Dec 2018 and Feb 2019 studies are discussed below

In both studies, we collected Betty's Brain activity logs, affect likelihoods from the newly developed affect detector models), student responses to in-the-moment interviews conducted at inflection points using the QRF app, facial expression videos using laptop webcams, responses to pre-post and science anxiety surveys, and (for some students) eye-gaze coordinates using eye-tracker devices.

Learners showed significant pre-to-post learning gains in both studies. We tried to gain a better understanding of learners' affect states by mapping *basic emotions* extracted from the webcam facial expression videos (in the Dec 2018 study) to the *achievement emotions* detected by the Betty's Brain affect detector models in the same study. This helped us map the complex emotional states observed in 'achievement scenarios' like open-ended learning environments to the fundamental or 'basic' human emotions observable across different settings. The complete findings from this analysis are reported in **Munshi et al. (2020)** and discussed further, in the context of the analyses for this dissertation, in Section VI.1.2. The in-the-moment QRF-triggered interviews from this study which were conducted at affective inflection points were hand coded after the study and analyzed to better understand students' affective experiences like *frustration*, as reported in **Baker et al. (2021)**. This analysis showed that students who went from experiencing *engaged concentration→frustration* often reported both *experiencing difficulty* and *using strategic behavior to resolve it*. This paper suggests that ”if a student goes from engaged to frustrated when encountering difficulty, but does not adopt a strategic behavior, it may be an appropriate time for the learning system to offer recommendations of learning strategies” (Baker et al., 2021).

In a further analysis **(Hutt et al., 2021)**, students' activities were studied in the context of Winne & Hadwin's COPES framework (Winne and Hadwin, 2008) to derive SRL constructs and understand how *science anxiety* (determined from survey responses) related to these constructs, e.g., *science anxiety is positively associated with searching behaviors but negatively associated with monitoring behaviors*, suggesting that anxious students may avoid solution-evaluation and instead opt for information-seeking. In **Munshi et al. (2022b)**, we presented **the design and evaluation of the first iteration of our adaptivity framework** which was deployed in the Feb 2019 study. The scaffold design proposed in this dissertation has been influenced by the findings from Munshi et al. (2022b), as discussed in more detail in Section IV.4.3.

The Sept 2021 study: A second iteration of scaffold design was performed after studying the findings from the Feb 2019 study, and this design iteration was evaluated in a small pilot study with six undergraduate students (median age: 20 years; 2 male, 4 female) from Vanderbilt University. A classroom study with middle schoolers could not be conducted since schools were closed due to COVID-19 restrictions, so we wanted to assess the scaffolds with a more mature group in this pilot study and then use the insights to further refine the scaffold triggers and contents for K-12 students. At the beginning of the pilot study, participants were provided an overview of the system. Then they worked on an introductory Betty's Brain unit for 15 minutes, to get hands-on experience on using the learning tools available in the system. The participants then switched to building causal models in the Climate Change unit. An experimenter observed each student as they completed their tasks.

Students' interactions with the system were logged as activities with timestamps. The system interface visible on students' laptop screens was recorded as video files using Open Broadcaster Software (OBS) Studio. Experimenters conducted interviews with students at different points in their learning session, in particular after they received adaptive scaffolds from Mr Davis, to learn students' perspectives about the feedback they just received and how they now planned to use it. Interviews were recorded using OBS and were transcribed post hoc in two phases, first using Otter.ai, an automated transcription software, and then manually correcting for errors in transcribed files by verifying against the raw audio. We performed a qualitative analysis of the collected multimodal data to evaluate scaffolds delivered to students. Results from scaffold assessment in the pilot study are reported in **Munshi et al. (2022a)**, and the implications of these findings on the current scaffold design framework is discussed in more detail in Section IV.4.3.

The implications of adaptive scaffolds deployed in both Feb 2019 and Sept 2021 studies on the current design approach is outlined in more detail in the next chapter, in Section IV.4.3. Overall, the (Chapter IV) presents different components of our current approach to designing and implementing adaptive scaffolds to support SRL strategies in Betty's Brain.

CHAPTER IV

Design and Implementation of the Adaptive Scaffolding Framework in Betty's Brain

This chapter presents our approach towards designing and implementing the adaptive scaffolding framework for self-regulated learning (SRL) in the Betty's Brain OELE. This includes (1) a **learner modeling framework** that used a *strategy detection approach* to determine the optimal trigger conditions for scaffolding, by performing an online diagnosis of students' self-regulation deficiencies from their cognitive-metacognitive strategy use and predictions of certain affect states during model-building and debugging tasks in Betty's Brain; (2) a set of **triggering conditions**, determined by the learner modeling approach that would drive the adaptive scaffolding process; and (3) a set of **conversational scaffolds** that were adapted to students' learning difficulties at the trigger conditions and help them to improve the deficient aspects of their self-regulated learning strategies.

IV.1 Theoretical Framework

The learner modeling and adaptive scaffolding approach developed in this dissertation builds on the study of major SRL models (Winne and Hadwin, 2008; Zimmerman, 2002a; Boekaerts, 1996; Efklides, 2011; Pintrich, 2000) (Chapter II) and empirical evidence (Azevedo et al., 2015; Munshi et al., 2018c), which suggest that SRL is a dynamic process made up of related cognitive, affective, metacognitive and motivational (CAMM) components. To scaffold for learners' deficiencies in CAMM regulation in an OELE, the learner modeling aspect of our framework first needed to detect and interpret (1) their cognitive-metacognitive behaviors and strategy use, and (2) their affect regulation behaviors, in different task contexts (Segedy et al., 2013) within the Betty's Brain learning environment.

Cognitive and metacognitive strategies are closely related to each other and are associated with orchestrating cognitive resources and skills. *Cognitive strategies* are goal-directed, situation-specific, and not universally applicable (Weinstein and Meyer, 1994). *Metacognitive strategies* involve more generally applicable processes like planning, monitoring, and reflecting (Donker et al., 2014; Zhang et al., 2021a). While cognitive strategies operate on the knowledge of "objects" or skills (Winne, 1995), *metacognition* is concerned with deliberating on the use of particular cognitive processes and combining them to accomplish larger tasks (Winne and Hadwin, 2008). *Metacognitive monitoring* bridges the gap between cognition and metacognition, since it involves observing and evaluating one's own execution of cognitive processes to exercise control and improve cognition (Kinnebrew et al., 2017). Overall, the use of these complex monitoring processes lead to the explicit development of learning strategies - "conscious and controllable sequences of actions students perform to facilitate and enhance their task performance” (Zhang et al., 2021a) - that can help learners achieve their learning objectives. Since novice learners often have difficulties in applying, monitoring, and reflecting on their use of learning strategies, we would model their abilities, in terms of their cognitive and metacognitive strategy use, in the context of their current learning tasks. This would help us infer the effectiveness of their strategy use and determine moments of ineffective strategy application. Diagnosis of a pattern of ineffective strategy use by the system, suggestive of self-regulation deficiencies, would trigger an appropriate scaffold to make the student aware of more effective strategies that could potentially improve their process of acquisition, construction, and reasoning with knowledge.

A secondary goal of our learner modeling framework was to build pedagogical scaffolds that help students regulate affect states like frustration or boredom that are known to be detrimental to learning (D'Mello and Graesser, 2012). Since affective appraisals during learning may be attributed to the use of effective or ineffective cognitive processes and cognitive-metacognitive strategies (Pintrich, 2000), scaffolding for ineffective affect regulation process in an OELE also needs to consider such attributions to be meaningful. To accomplish this, our design framework would obtain a measure of the change in learners' affect states during learning (for instance, a transition from a positive or neutral state to a negative state of frustration or boredom) and interpret such states as indicators of ineffective affect regulation. D'Mello and Graesser suggests the cognitive/metacognitive significance of specific affective states. For example, unresolved and persistent confusion, which D'Mello and Graesser calls ”hopeless confusion” can lead to a state like frustration. And while confusion may be productive or unproductive depending on the learner's ability to resolve the confusion, frustration is generally considered to be unproductive and harmful for learning, and therefore, providing scaffolding may help the student overcome the problems they may have with their blocked goals and plans. Providing strategic scaffolds for regulation of frustration is further supported by the findings from Baker et al. (2021), as discussed earlier in Section III.2. The forced-effort theories of boredom (Larson and Richards, 1991) and empirical research by D'Mello and Graesser (2012) also suggest links between the affective state of boredom and disengagement with the task.

Therefore, to scaffold for students' *frustration* or *boredom* appraisals during learning, our framework needed to determine the learners' activities and behaviors when they started showing the frustration or boredom, to determine the root of the problem they are stuck in, and use this understanding to respond with an appropriate scaffold that would teach them learning strategies for resolution of this problem.

IV.2 Overview of the Learner Modeling and Adaptive Scaffolding Approach

In Figure IV.1, we present an overview of the approach we took for designing a learner modeling and adaptive scaffolding framework in Betty's Brain, that uses the theoretical framework discussed in the previous section.

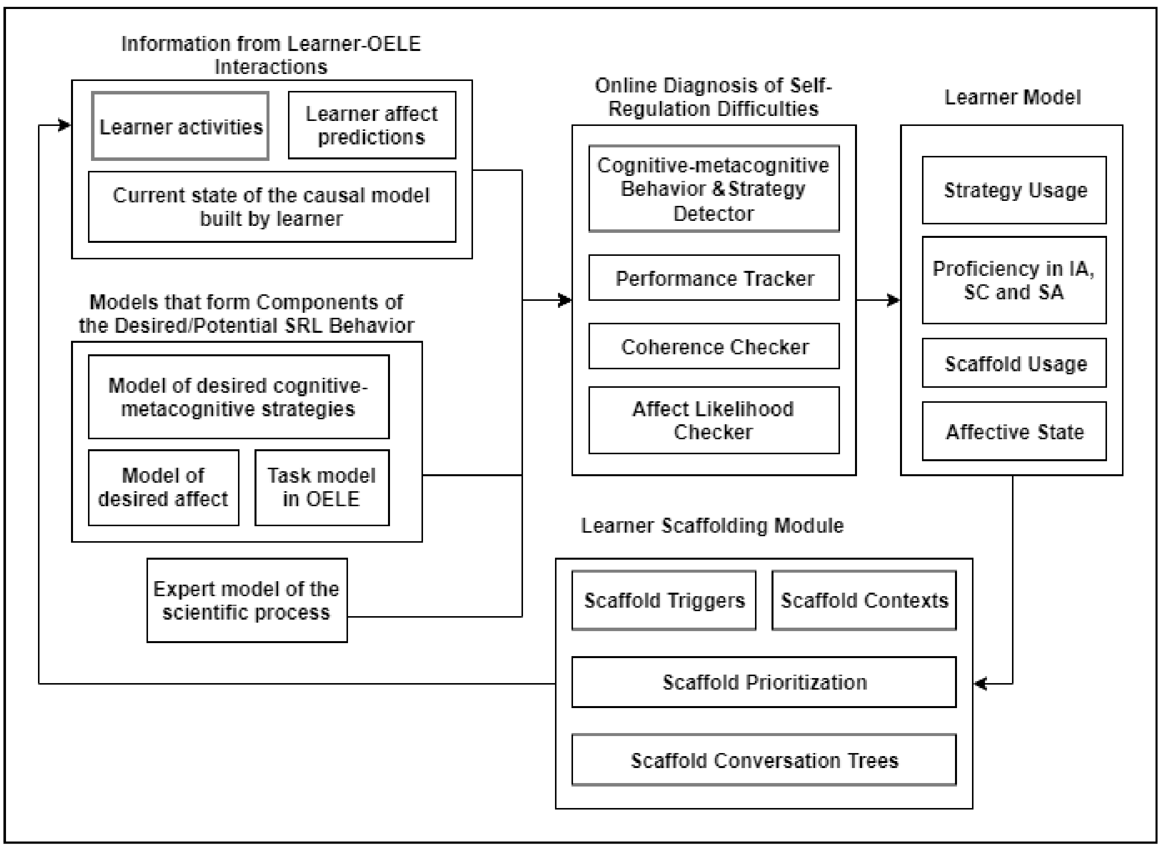


Figure IV.1: The Learner Modeling and Adaptive Scaffolding Framework

There are several components to this design framework.

The learner modeling aspect of the framework tracks observable information from learner-OELE interactions (viz., logged activities, affect predictions generated by affect detector models (Jiang et al., 2018), and the current state of the causal map) and uses a ***strategy detection approach*** to detect temporal activity sequences and interpret them as cognitive-metacognitive strategies using the task model (Kinnebrew et al., 2017). A *coherence checker* module infers whether the learner has been coherent in their strategy use so far (by using Segedy et al. (2015)'s coherence analysis approach) and a *performance checker* tracks the change in the learner's causal modeling performance after applying the strategy to determine strategy effectiveness. An *affect likelihood checker* tracks temporal changes in learner affect, looking for transition into states where emotions like frustration or boredom are likely to be present. Such states are then evaluated in the context of cognitive information obtained from recent activities and performance to determine the optimal scaffolds to deliver to the learner.

Both strategy use and affect regulation form important components of the learner modeling framework, which generates an understanding of the learner's proficiency in their cognitive processes while engaging in information acquisition (IA), solution construction (SC) and solution assessment (SA) tasks in Betty's Brain. The learner modeling approach is used to trigger scaffolds that respond to patterns of ineffective strategy use or a cognitive-affective state that is likely to suggest ineffective affect regulation. The model also tracks how the learner has responded to scaffolds already received about a strategy or affect transition, to determine the 'level' (cf., Section IV.4.2) to be used for similar scaffolding in future.

When the learner model determines that a student is ready to receive a specific type of scaffold to help with applying their learning strategies or regulating their affect, the trigger condition for that scaffold is inserted into a priority queue. The queue uses a priority assignment algorithm (see Appendix C) to assign priorities to detected trigger conditions based on an understanding of the current state of the student's causal map (e.g., whether the map is sparse, dense, contains relatively high number of correct versus incorrect links, or vice-versa) to determine which type of strategic scaffold is most likely to be beneficial to the student's model-building and debugging tasks at that point in their learning process. The learner scaffolding module in Betty's Brain (1) listens to the priority queue and keeps track of the triggering condition with the highest priority in the queue, (2) determines the level of contextualization at which the student should receive feedback for this trigger, (3) finds the conversation tree for this trigger condition (and at the appropriate level, as applicable) (4) adds additional context information as required in the tree, and (5) uses the virtual mentor agent to deliver the feedback to the learner.

In the next section, we outline the tasks completed to design, develop and evaluate this adaptivity framework to achieve the goals of this dissertation.

IV.3 Scaffold Design and Evaluation Tasks Completed in this Dissertation

This dissertation presents a principled approach to designing and implementing an adaptive scaffolding framework in the Betty's Brain OELE (Section IV.4) by accomplishing the following set of tasks:

1. Designing a learner modeling architecture (Figure IV.3) to capture learners' task proficiency and self-regulated learning behaviors to achieve their learning goals in Betty's Brain;

2. Designing an approach to monitor changes in the learner model and detect patterns that suggest difficulties in cognitive, metacognitive or emotional regulation processes, to identify *trigger conditions* for scaffolding.

3. Designing and implementing scaffolds that adapt to learner difficulties at the trigger conditions and provide contextualized meaningful feedback in an interactive conversational manner.

Upon completion of the design and implementation phase described in the current chapter, this dissertation also presents an approach to evaluate the designed adaptivity framework (Chapter V) by completing the following tasks:

1. Conducting an experimental study to collect data as K-12 learners interact with the Betty's Brain environment;

2. Analyzing the collected data to assess the impact of different adaptive scaffolds delivered by our framework on the self-regulated learning strategies of learners in Betty's Brain. (Findings from the evaluation study are reported in Chapter VI.)

IV.4 Design and Implementation of the Adaptivity Framework in Betty's Brain

There are two primary components to the design and implementation of our adaptive scaffolding framework (see Figure IV.1) in Betty's Brain: **(1) A learner modeling architecture to determine and monitor the trigger conditions** for scaffolding; and **(2) A Conversation tree format to deliver adaptive feedback** when a trigger condition is satisfied. We discuss these two components in more detail in Sections IV.4.1 and IV.4.2 respectively.

IV.4.1 A Learner Modeling and Strategy Detection Approach to Trigger Scaffolds

In this section, we discuss the primary design components of the learner modeling architecture and outline the procedure for their implementation in Betty's Brain.

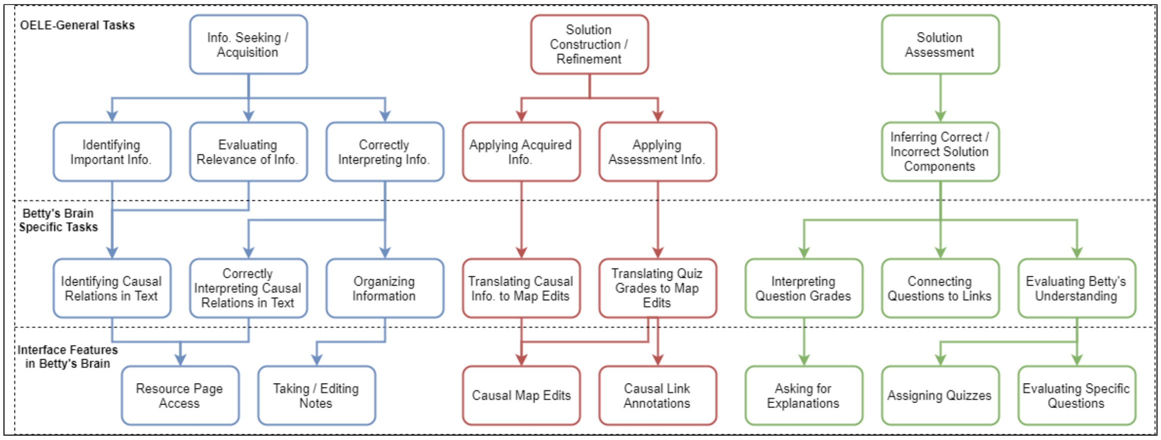


Figure IV.2: The Task Hierarchy Model in Betty's Brain; Modified from Kinnebrew et al. (2017)

IV.4.1.1 Detection of Task-Oriented Cognitive-Metacognitive Strategies

To be successful in an OELE like Betty's Brain, learners must be able to efficiently decompose their overall goal (of building a correct causal map to teach Betty) into specific sub-goals or *tasks*, and be strategic in the activities they perform, while monitoring their progress towards completing their tasks (Winne, 2014). Adopting effective SRL strategies should help students identify, interpret, and resolve any difficulties they may encounter while operating in different task contexts within the learning environment. Therefore, any scaffolding framework that intends to help learners engage successfully in such environments should incorporate a learner model, which captures an understanding of:

1. *the learner's current task context*. This can be derived from the subtask(s) a learner is currently working on. This can include acquiring information needed to build the causal map, constructing and refining the causal map, and assessing the correctness of the causal map;
2. *the context and effectiveness of their recent activities*. Context can be derived from their recent activities, such as observing whether students read a number of resource pages sequentially, or how they combine their reading, map building, and map assessment activities. Causal-modeling effectiveness or performance is measured by comparing their recent map-building activities with a correct or 'expert' map (cf., Figure B.3) of the scientific process;
3. *the specific difficulties learners have in relation to their current task and activities*. Examples may include the inability to find the science book pages that provide information they need to construct causal links, the inability to convert the information read into correct causal links, and the inability to analyze quiz results to infer correct versus incorrect links in their map.

Grounding scaffolds in the explicit context of the students' current tasks and the effectiveness of their model-building efforts provides concrete reference points on which to base any strategic feedback (Segedy et al., 2013). To understand and track learners' strategic cognitive and metacognitive behaviors in context, we adopt the hierarchical **OELE task model** developed by Kinnebrew et al. (2017). The task model helps us map learners' higher-level cognitive processes to their activities within the learning environment, thereby supporting our focus on helping students to develop task-oriented cognitive-metacognitive strategies in the Betty's Brain environment. The task model uses cognitive task analysis methods (Schraagen et al., 2000) to break up the overall task into sub-tasks, linking students' activities (such as reading resource pages, taking notes, editing and annotating their causal maps, asking Betty to take a quiz to evaluate the state of the current map, asking Mr Davis for explanations to specific causal links on their map) to more generic goal-oriented task structures that represent cognitive processes at the top levels of the task model: (1) Information Acquisition (IA), (2) Solution Construction (SC), and (3) Solution Assessment (SA) (Kinnebrew et al., 2017). Students working in Betty's Brain need to employ these three types of cognitive processes to build and analyze their causal models to teach Betty.

Figure IV.2 presents our task model in Betty's Brain, adapted from Kinnebrew et al. (2017). (This figure includes note-taking to organize information as an additional component of the information acquisition process in Betty's Brain, that is not present in Kinnebrew et al. (2017)'s task model.) Figure IV.2 shows the three sub tasks mapped on to cognitive processes expressed at different levels of detail. At the lowest level, students operationalize the information acquisition (IA) process by reading the hypertext resource pages, and by taking and organizing notes. Solution construction (SC) involves map building and map refinement tasks that students perform in the *causal map* interface. Solution assessment (SA) involves quiz-taking activities, analyzing the quiz results (checking the correct and incorrect answers to the quiz), and seeking explanations by clicking on specific questions.

The combination of multiple tasks and sub-tasks in the task model relates to Winne and Hadwin's (2008) SRL model, and illustrates the coordination and enactment of different learning and problem-solving activities in the form of cognitive processes and strategies that are a vital component of metacognitive regulation in conjunction with monitoring processes (Schwartz et al., 2009). Therefore, the task model can be used as a reference framework to track students' frequently occurring action sequences, and interpret them as cognitive-metacognitive *strategy* constructs (Kinnebrew et al., 2017). As discussed in Section IV.1, *strategies* in this context represent learners' conscious and controllable sequences of actions that facilitate and enhance task performance (Zhang et al., 2021a). While cognitive processes primarily relate to activities that suggest the use of IA, SC and SA processes, **cognitive strategies** result from observing meaningful combinations of the IA, SC and SA activities. As students work on their learning and problem-solving tasks, they may switch between IA, SC, and SA sub-tasks and combine them in different ways to accomplish their goals. Basu et al. (2017) express learners' temporal activity sequences as binary relations - for example, IA (Read)

→ SC (Build) or SA (Quiz) →SC (Build). Segedy et al. (2015)'s coherence analysis approach helps us further interpret relations involving Build actions as "effective" and "coherent". An **effective** Build implies the addition, deletion or modification of causal links that lead to an improvement in causal modeling performance, observed from an increase in *map scores* (discussed in Section V.2.1). So, reading a Science Book page and then adding a correct causal link from this page on the map would be characterized as the use of a Read→Build-Effective strategy (e.g., Read → Adding a Correct Causal Link). A **coherent** Build is a link edit which *supports* prior Read or Quiz actions (e.g., reading a page → Adding a link that is relevant to the page just read). An incoherent Read→Build, where a student adds or modifies causal links that are not supported by their prior information acquisition or solution assessment, would suggest a less strategic and possibly more experimental approach to model building.

This dissertation presents a **strategy detection approach** that tracks key binary relations from sequences of students' observable activity traces in Betty's Brain, and uses the *task model as a reference framework* to interpret such relations as as temporal a sequence of cognitive processes (IA, SC and SA tasks). Observing the (lack of) *effectiveness* or *coherence* of model-building actions allows us to further characterize these activity sequences as *deficiencies in cognitive strategy use* by the learner. The successful use of cognitive strategies also reflects the use of *deeper metacognitive monitoring and self-reflection behaviors* by the student. For example, a Read→LinkDelete-Effective-Coherent cognitive strategy, which involves a transition from information acquisition (Read) to an effective and coherent Build (where the student deletes an incorrect link from their causal map) suggests that the learner is able to search within the Science Book and identify the section containing potentially important causal information given the current state of their map, and is then also able to extract the correct causal relation from this section, compare it with the already added incorrect link on the map, and determine that the incorrect link needs to be removed to improve the quality of the causal map. The evidence of such a strategy use suggests efficient metacognitive monitoring and self-reflection behaviors by the learner. Similarly, the detection of a Read→LinkDelete-Inffective strategy would suggest deficiencies in cognitive processing and the associated metacognitive behaviors, which may be supported by appropriate adaptive scaffolding.

The different types of binary relations detected by our current strategy detection framework (each of which suggest a deficiency in the use of specific cognitive-metacognitive strategies while completing the causal modeling task in Betty's Brain) is discussed in Section **??**, along with the adaptive scaffolds provided to support the development of more effective strategies in each case. This list is determined based on findings from previous classroom studies (Section III.2) and scaffold design/ evaluation cycles (Section IV.3).

The strategy detection framework looks for 'patterns' (multiple occurrences) of an ineffective strategy use, by keeping a count of each type of binary relations that suggest strategy deficiency. If this count exceeds

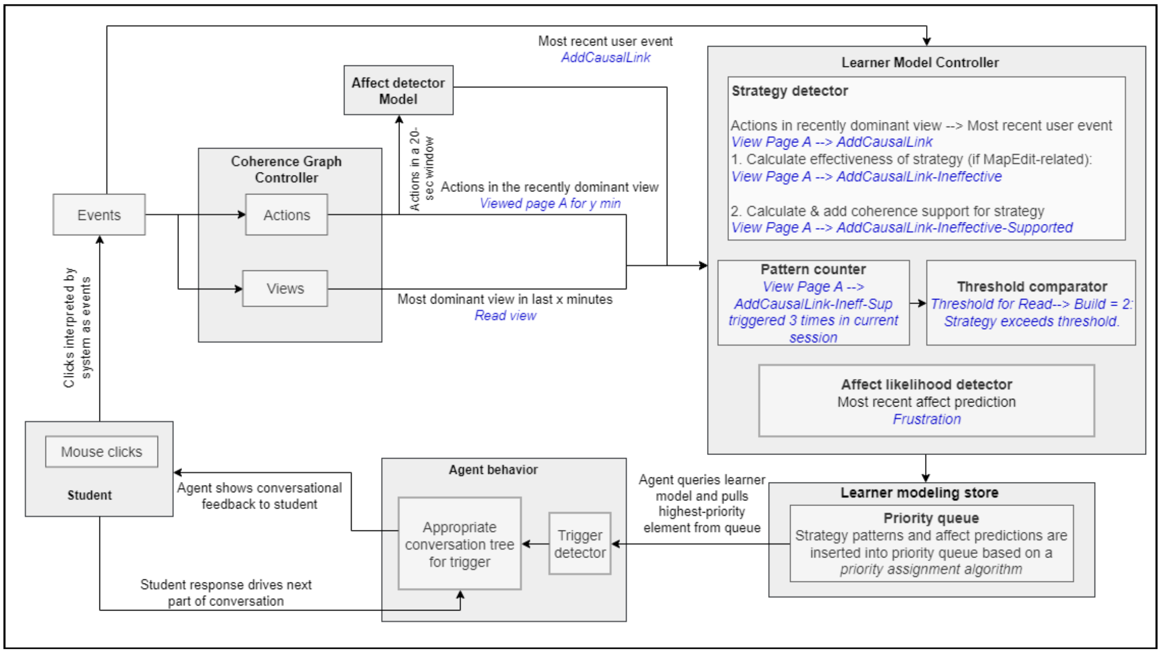


Figure IV.3: The Learner Modeling and Adaptive Scaffolding Architecture in Betty's Brain

a certain preset *threshold* (say, 2), it is considered to be a *pattern* exhibited by the student. Our objective with such patetrn detection is further clarified in the "Strategy pattern detection" subsection later in this section.

Changes in the effectiveness and coherence of cognitive-metacognitive strategy patterns employed by a student are reflected as *updates to their learner model* (Figure IV.1), which uses this information to build an understanding of how strategic and proficient the student has been in the application of different strategies related to information acquisition, solution construction and solution assessment processes while learning in Betty's Brain. This understanding of student proficiency helps to trigger appropriate scaffolds to support the student to develop the effective cognitive-metacognitive behaviors and strategies they currently lack.

Figure IV.3 shows the **architecture** that operationalizes this learner modeling and scaffold triggering approach by interpreting online learner data in Betty's Brain. We discuss the detection of the cognitive-metacognitive conditions for triggering scaffolds using this architecture below. (The affective conditions for triggering scaffolds are discussed in Section IV.4.1.2.)

Strategy pattern detection: Students' mouse clicks are interpreted as events by the system, and further processed by a 'Coherence Graph Controller' module into an $< action, view >$ representation that records their current action (e.g., addition of a "deforestation increases carbon dioxide" link on their map) and the current visible system interface or 'view' (e.g., the "causal map" view). The 'Learner Model Controller' in Betty's Brain extracts the most recent event, the recently dominant view (say, in the last 3 minutes) and the actions performed in this view. This information is then used to construct the binary relations (illustrated *in blue* in figure IV.3) that are suggestive of student's strategic behaviors. Since we want to intervene using scaffolds only when a student has difficulties in applying suitable learning strategies, the learner modeling framework specifically looks for behaviors that suggest unproductive or inefficient strategy use by the student (e.g., reading a science book page→ adding an incorrect link from that page onto the causal map). A 'Pattern Counter' keeps track of the number of times a student exhibits a certain type of strategy, and a 'Threshold Comparator' informs the learner model when the count exceeds a set threshold, thereby signaling that a strategy pattern has been detected.

Priority assignment: In the case that multiple types of strategy patterns are detected from a students' recent activities within a time interval, a pattern prioritization approach helps to identify the pattern reflective of the student's most crucial strategic difficulty at the current time. This is achieved by applying a *priority assignment algorithm* (Appendix C) to insert detected patterns into a priority queue. The priority assignment process studies each pattern in the context of their impact on the student's causal modeling progress in Betty's Brain. To illustrate this process, let us take the example of a student for whom three types of strategy patterns have been detected: IA(Read)→SC(Build); SA(Quiz)→SC(Build); and SA(Quiz)→IA(Read). Here, the algorithm has to decide whether the student has been having more difficulties in applying strategies related to IA (Read), SC (Build) or SA (Quiz). Based on an assessment of the students' recent model-building progress, if it is found that the student has been adding a disproportionately high number of incorrect links to their map, this suggests difficulties in applying model-building strategies. So, the algorithm assigns a higher priority to patterns related to model-building activities (Build), such as Read→AddIncorrectink or Quiz→AddIncorrectLink) versus non-Build patterns like Quiz→Read. The pattern prioritization process ensures that the next feedback the student receives (triggered based on the highest priority pattern) is the one they need the most, in this case to help them debug their causal model by engaging in strategic reading or quizzing behaviors.

Additional checks to trigger relevant scaffolds: Beyond the detection of strategy patterns and prioritization of the patterns reflective of the most pressing learning difficulties, two additional factors considered in our scaffolding framework ensure that the students receive scaffolds that are timely, relevant and meaningful:

1. Each pattern inserted into the priority queue is associated with a 'time-to-live' (TTL) paramater. If the time-to-live for a pattern exceeds a set time interval (say, 5 minutes), this suggests that the pattern is no longer a 'recent' or relevant one, and is dropped from the queue. This process ensures that students only receive scaffolds on their recent behaviors and not on behaviors they exhibited in the not-so-recent past.
2. To trigger a scaffold on a detected strategy pattern, the mentor agent in Betty's Brain, Mr Davis, has to request the priority queue for the highest-priority pattern. An 'inter-feedback interval' is included as a

component of the 'Agent Behavior' module (Figure IV.3), which tracks the time that has passed since a student received their last feedback and only requests the priority queue for another pattern when the time has exceeded a certain interval (say, 5 minutes). This inter-feedback interval checking ensures that the student does not feel interrupted by too frequent interventions from the agents.

IV.4.1.2 Detection of Affect Likelihood Scores

As discussed earlier in this chapter, a secondary component of our learner modeling and adaptive scaffolding process to support SRL is to detect and interpret learners' affective states during learning and respond to unproductive emotion-cognition regulation using appropriate feedback.

D'Mello and Graesser (2012) discuss a model of affect dynamics that illustrate how learners' affect states dynamically arise and evolve as they interact with their tasks in the learning environment. D'Mello and Graesser's model considers achievement emotion states such as engagement/flow, confusion, frustration and boredom and shows how learners transition between these tasks and what the emotion transitions signify. For example, one of the major emotion transitions involves a shift from *Engagement* (an equilibrium state) $\rightarrow$ *Confusion* (suggestive of cognitive disequilibrium) $\rightarrow$ *Frustration* (when the confusion has not been resolved and the student has got stuck) $\rightarrow$ *Boredom* (persistent failure leading to disengagement from the task). The set of affect transitions above suggest an ineffective emotion regulation process. But while D'Mello and Graesser some of the specific affect transitions observed by the D'Mello and Graesser (2012) model explains cognitive attributions to specific affect states like frustration or boredom, some of the affect transitions observed in this model were not replicated in other empirical studies, as investigated by Karumbaiah et al. (2018), the appraisals of affect states like frustration and boredom during learning, and the potential of addressing such states by providing strategic scaffolding to learners, is explored by further research, including for the Betty's Brain environment by Baker et al. (2021) .

In view of these findings, the goal of the affect regulation component of our learner modeling framework is to detect affect transitions (using the BROMP-based detector models discussed earlier) that lead from a positive or neutral affective state to a state of dominant negative affect (frustration or boredom), as students work in Betty's Brain. These affect appraisals would then be evaluated in the context of students' recent activities and behaviors. *For example*, if the affect detector starts predicting dominant frustration for a student and the recent activities suggest that the student has been reading a Science Book page for some time, with the strategic component of the learner model further suggesting an ineffective *Read*$\rightarrow$*Build* strategy use, we can combine this affective and strategic information to infer that the student may be struggling with the extraction of correct causal links from the page they are reading, and this continued struggle over a period of time has led to their shift towards frustration. This would trigger appropriate feedback from the mentor agent to ensure more effective cognitive processing and strategy use and possibly a decrease in frustration (see the Scaffold 9 case in Section IV.4.4).

To operationalize the detection of affect states in Betty's Brain (see Figure IV.3), our framework uses the affect detector models developed by Jiang et al. (2018) that predict learners' emotions from their activities in Betty's Brain. These detectors are trained and validated on affect labels hand-coded by classroom researchers using BROMP (Ocumpaugh et al., 2015) in an experimental Betty's Brain study (Munshi et al., 2018c), as discussed earlier in Section III.2.

The affect detector models are embedded within Betty's Brain, and use a traditional feature engineering approach to distill features from a sliding window of learners' recent activity traces. The features are used as input to a subsequent classification algorithm that generates the likelihood of five types of achievement emotions: *engaged concentration, boredom, delight, confusion, and frustration* (Jiang et al., 2018). The detectors execute every 20sec to generate updated likelihood values of each emotion state. The emotion state with the highest likelihood value for a 20-sec interval is considered the affect prediction for that interval.

Our learner modeling approach tracks the changes in learners' affect predictions determined using the affect detector models. For this purpose, we construct binary relations (similar to the approach for strategy detection in Section IV.4.1.1) from successive affect predictions generated by the affect detector models. These relations infer affect transitions to interpret transitions towards frustration or boredom that, if not regulated, may be detrimental to the learning process. Keeping in mind the relations between cognitive-metacognitive and affective processes during learning (Section II.1.3), the detected affect transitions are analyzed in conjunction to detected cognitive-metacognitive learning difficulties (possible attributions of affect) to trigger the agent scaffolds.

In prior attempts at scaffolding learners in Betty's Brain (Munshi et al., 2022b), we have observed that a non-uniform triggering criteria across different types of scaffolds leads to some scaffolds being triggered too frequently and some others being triggered very infrequently. Therefore, we include the concept of priority assignment to affect-based trigger conditions, so that they are inserted into the same priority queue as the strategy patterns (cf., Section IV.4.1.1). Additionally, affect-based triggers are checked for their relevance using the 'time to live' and 'inter-feedback interval' parameters discussed in the case of the strategy-based triggers in Section IV.4.1.1.

Upon the determination of trigger conditions for scaffolding and the insertion of the triggers to priority queue, the next step is to **monitor the trigger conditions** to deliver appropriate scaffold for the learner's most pressing self-regulation difficulty at the current moment, as determined from the highest-priority trigger condition in the priority queue. For this purpose, the learner scaffolding module in Betty's Brain (Figure IV.1) requests the priority queue for trigger conditions in a periodic manner, pull the trigger condition with the highest priority at the time of request, and deliver the appropriate scaffold for the trigger using conversation trees. We discuss this process of scaffold delivery in the next section in more detail.

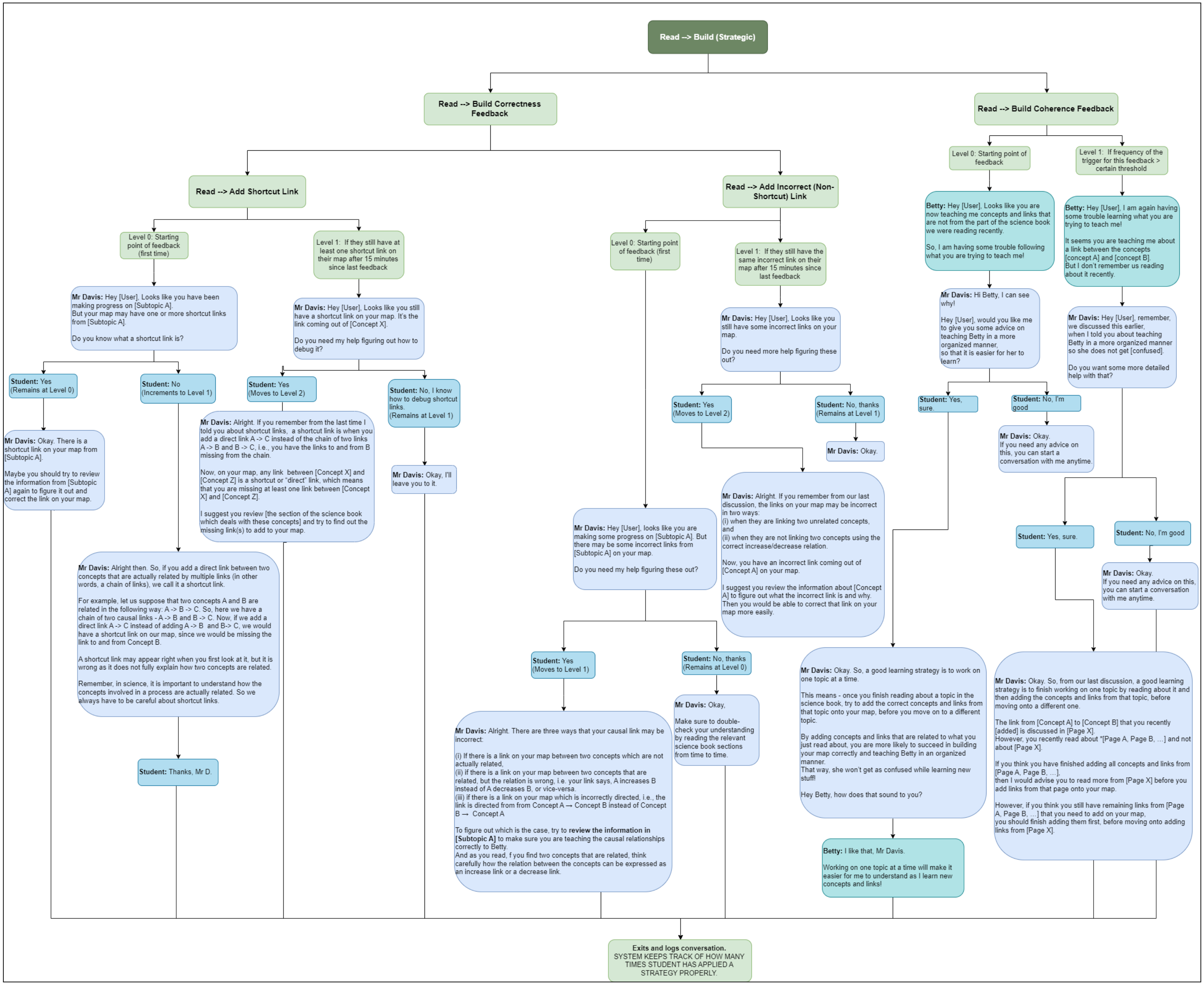
IV.4.2 Delivering Adaptive Scaffolds for different Trigger Conditions

Adaptive scaffolds from our framework are delivered by the virtual mentor agent Mr Davis present in the Betty's Brain environment. One aspect of the agent behaviors defined in the learner scaffolding module in our framework is to query the priority queue for the highest priority trigger (strategy pattern or affect transition). When the agent gets the highest priority element from the queue, this element forms the triggering condition for scaffolding. The agent then looks into a <triggering condition, conversation tree> map to extract the conversation tree corresponding to the triggering condition.

IV.4.2.1 Conversation Tree Representation for Scaffold Delivery

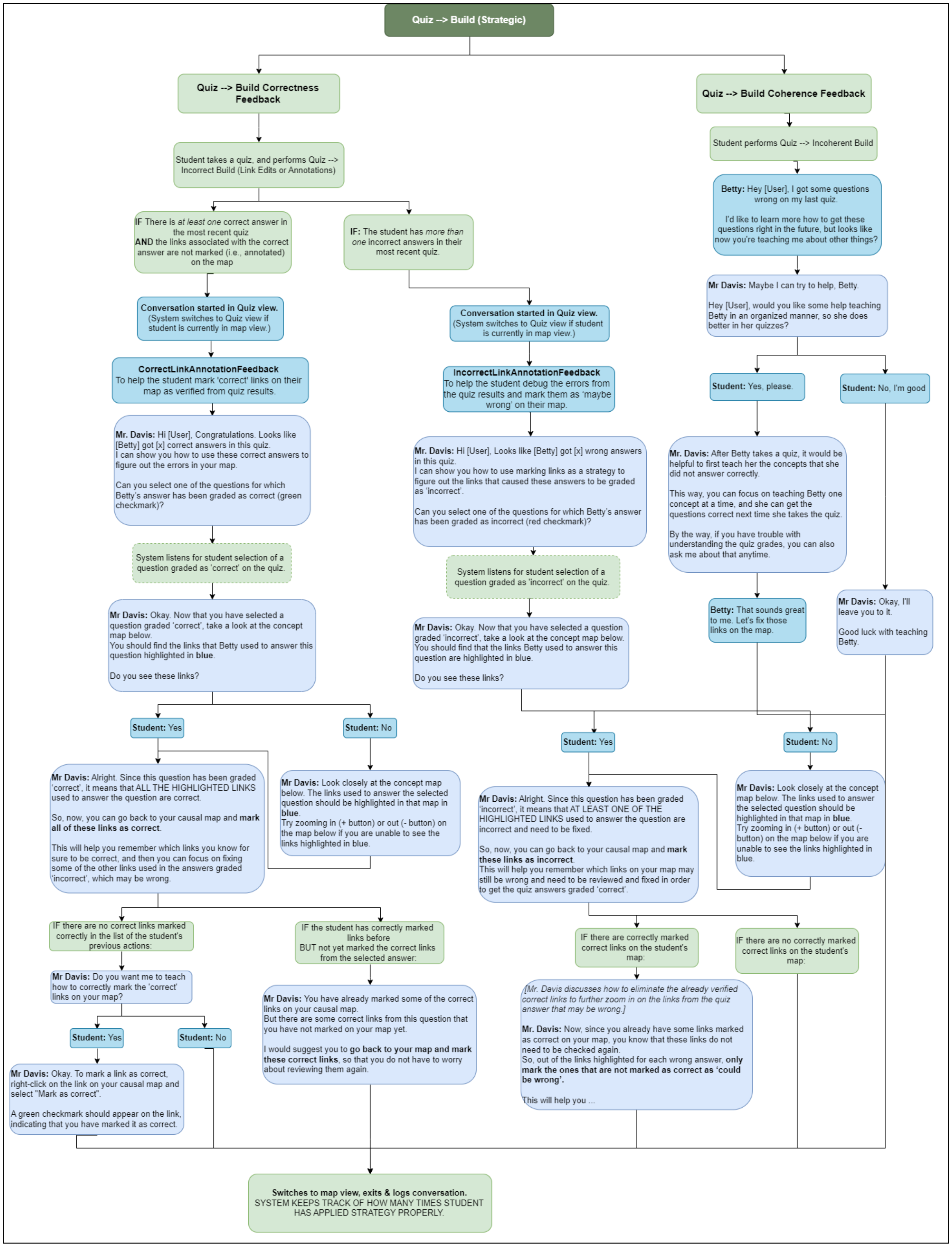
Conversation trees have been used to deliver contextualized and conversational learner scaffolds in past experimental research using Betty's Brain (Segedy et al., 2013; Munshi et al., 2022b). They facilitate back-and-forth conversation between the learner and the student that seeks to engage learners in authentic social interactions (Vygotsky, 1978). The set of 9 conversation trees (3 Read→Build, 3 Quiz→Build, 1 Quiz→Read and 2 affective/strategic) developed in the final design of the adaptive scaffolding framework for this dissertation are presented in Figure IV.4. Conversation trees developed for previous iterations of the scaffolding framework are present in Appendix B: this includes Figure B.1, used for Munshi et al. (2022b)), and Figure B.2, used for Munshi et al. (2022a)). In all of these figures, we note that the conversation tree nodes contain the skeleton of the feedback to be offered to the learner for a particular trigger condition and at a specific scaffolding level (discussed later in this section). However, the content of the tree nodes contain places where additional task information, as determined from the learner's current task and recent activities, is filled in to offer more relevant and contextualized feedback. This is illustrated in Figure B.1 where we can compare actual contextualized scaffolds received by a student to the conversation tree structure that facilitated the delivery of the scaffold.

Conversations following the tree structure are initiated by one of the virtual agents. At the end of each piece of feedback within a tree node, the agent asks the learner if they require further guidance. The student can indicate their answer to this question by selecting the appropriate choice from a drop-down list. If they select the option to continue the conversation further, the agent delivers the next piece of conversation, which includes more detailed feedback on the issue at hand; else, the agent stops the conversation and lets the student continue working, while monitoring their strategic and affective processes in the background (for providing future scaffolds). This process allows the learner to have some decision-making agency with respect to the



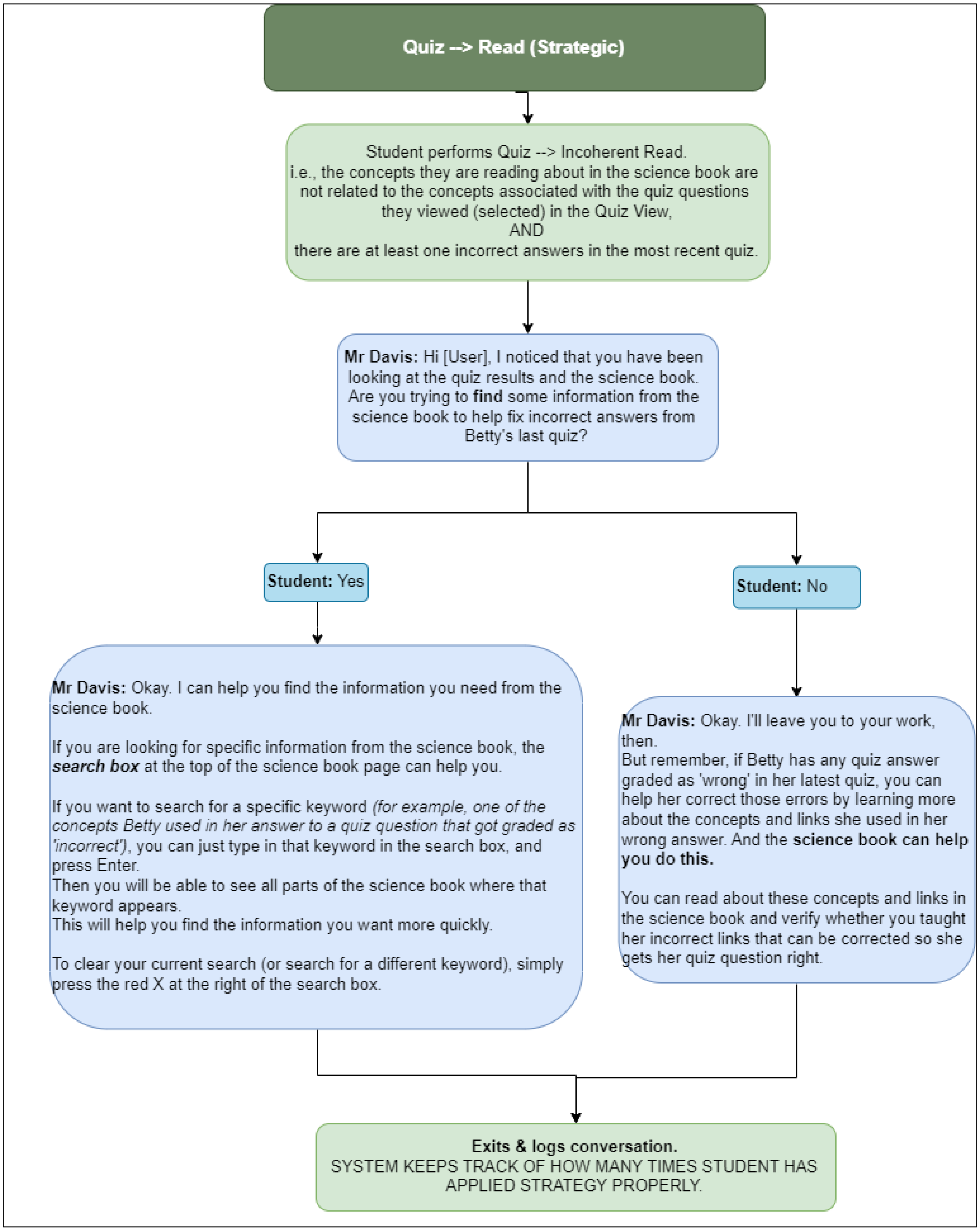
(a)

Figure IV.4: Conversation trees for different Read→Build strategic scaffolds



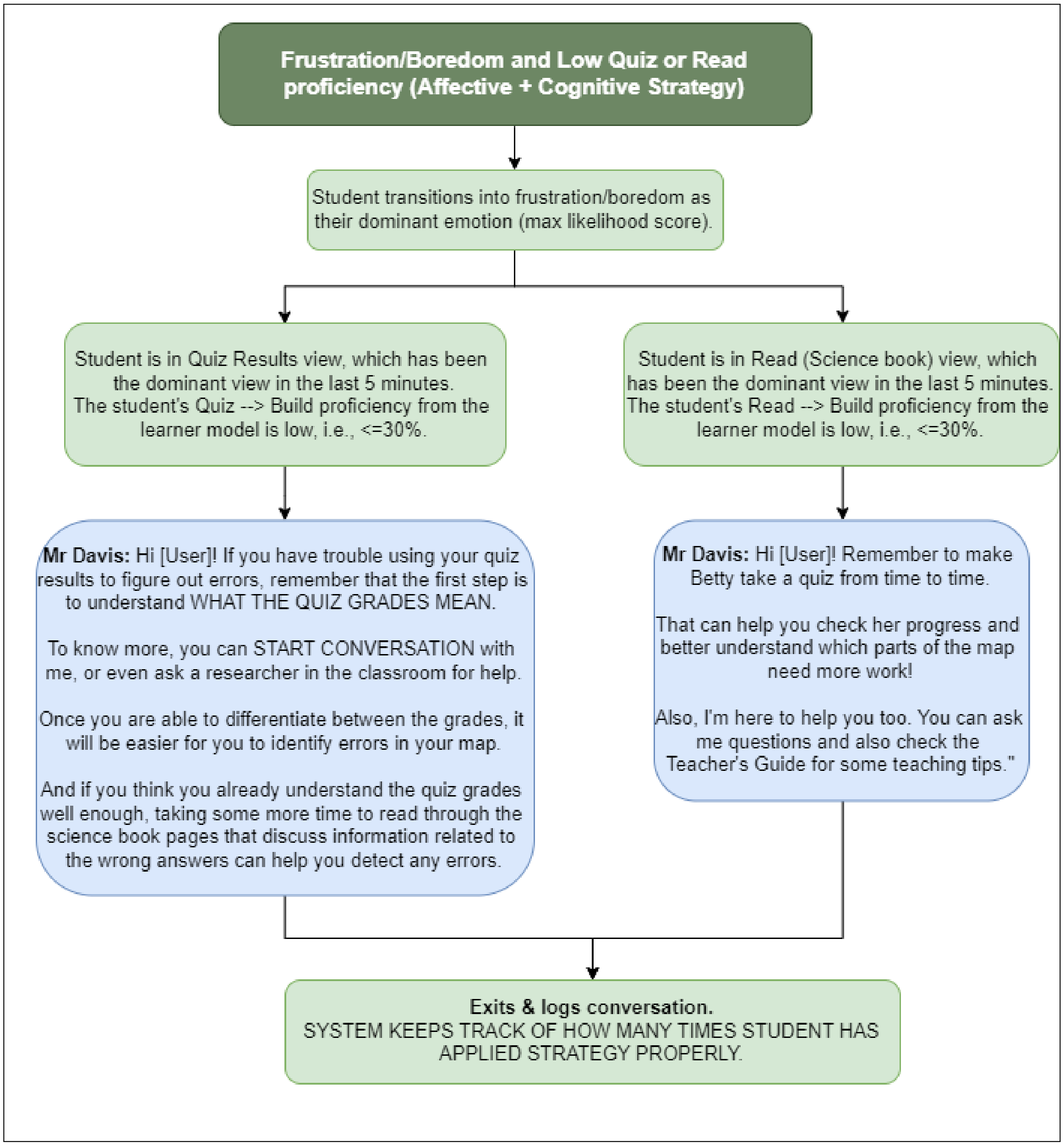
(b)

Figure IV.4: Conversation trees for different Quiz→Build strategic scaffolds



(c)

Figure IV.4: Conversation trees for different Quiz→Read strategic scaffolds



(d)

Figure IV.4: Conversation trees for affective-cognitive scaffolds amount of feedback they receive at a time and consider to be sufficient to meed their current needs.

IV.4.2.2 Scaffold Level

Our scaffolding framework includes an additional **scaffold level** feature to further increase the adaptivity of student support through the conversation trees.

1. Learners are initially offered a *Level-0* feedback, especially the first time they receive a scaffold related to a particular ineffective strategy, e.g., an ineffective Read→Build pattern. The Level-0 feedback offers general hints to help the learners become more strategic in their work. For example, in the earlier Read→Build scenario, depending on the specific Build issue, the learner would be told how they can read their resource pages more strategically to extract the correct causal links and fix their errors.
2. If the same strategy pattern is observed again during the student's learning process, suggesting that they have been unable to develop an effective strategy to counter their errors, the next time the student would be scaffolded with a *Level-1* feedback. The Level-1 feedback includes more specific hints, contextualized to the student's current learning difficulty, to help them better understand and apply the suggested feedback. In the ineffective Read→Build scenario, this may be informing the student about concepts related to erroneous causal links on their map and then helping the student locate specific pages in their science book (or specific sections within pages) that they can review to fix the errors.

Table B.2 illustrates the concept of 'scaffold level' with examples from a prior version of adaptive scaffolding (Munshi et al., 2022a), where we can observe how the scaffold level determines the level of contextualization of conversational feedback for the same type of triggering conditions. The inclusion of scaffold levels in the scaffolding framework helps to further tailor the level of support to learners' needs, by allowing them to develop their strategic processes more independently in initial stages and providing more corrective hints if they are unable to develop and apply effective strategies even after the initial support. However, even in the higher-level (more contextualized scaffolds), our design of conversations ensures that the learner only receives supportive hints and not bottom-out hints e.g., by being told the exact causal links they should build to get a particular quiz answer graded as correct. This is done to discourage ineffective behaviors like gaming the system (Baker et al., 2004, 2008).

IV.4.3 Findings from Design-Based Research Studies that Informed the Current Scaffold Design

We applied a design-based research (DBR) approach (Wang and Hannafin, 2005) to gradually refine the design of the triggering conditions and the conversation trees. Lessons from evaluating the scaffolds developed in each design cycle led to the next design. Our purpose was to use empirical evidence from the evaluation stage of each DBR cycle to improve the adaptive scaffolding efficiency. (More details on specific studies in this DBR process are provided in Section III.2.)

1. **Findings from DBR Cycle 1:** The first scaffold design iteration leading to this dissertation included a set of conversational scaffolds provided at task-specific cognitive-metacognitive trigger conditions in Betty's Brain that were termed as 'inflection points'. A complete list of the inflection points and scaffold content from this design iteration is provided in Table B.1 in Appendix B. Examples of conversation trees from this study are presented in Figure B.1, also in Appendix B. We performed quantitative analyses from data collected in a classroom study with 98 students to evaluate these scaffolds. The findings from evaluating this set of scaffolds are reported in Munshi et al. (2022b).

   These results suggest that some of the scaffolds developed in this design iteration were effective and some others were not. *Specifically*, since we used only cognitive-metacognitive triggers to direct scaffolds that targeted both strategic and affective regulation, most scaffolds targeting affect were largely ineffective, at least in terms of the emotion predictions obtained from BROMP-based detectors. This has led us to include affective trigger conditions in the current design framework.

   In addition, the scaffolds to regulate affect in Munshi et al. (2022b) only provided encouragement or reassurance prompts and did not target the resolution of the cognitive-metacognitive attributions of the observed emotions. Therefore, in the final design, we not only track learners' affect likelihood scores in Betty's Brain but we also track their cognitive behaviors to determine the potential cognitive attribution of these emotions, and use this information to direct scaffolds for affective-cognitive regulation. As discussed in earlier sections in this chapter, this decision also stems from theoretical and empirical cognitive-affective relations and the potential for adaptive scaffolding to use this relations and help affect regulation through strategy feedback (Baker et al., 2021).

   Other relevant lessons from this design iteration included the fact that certain scaffolds were triggered very infrequently with respect to others. We inferred that this was due to assigning non-uniform triggering thresholds to different types of strategy patterns. Therefore, we use a uniform triggering threshold for the detection of different types of patterns so as not to include any additional bias towards specific adaptive scaffolds in our scaffold triggering process.

2. **Findings from DBR Cycle 2:** The second scaffold design iteration was motivated by findings from Munshi et al. (2022b). We redesigned the scaffolds that were ineffective or less effective, and evaluated them in a lab (pilot) study in Sept 2021. While this design iteration included only strategic scaffolds, the specific scaffolds included in this design were significantly more complex and formalized compared to the prior design iteration. This can be realized from looking at the conversation trees from this study, presented in Figure B.2. This design of scaffolds also included the 'scaffold level' construct discussed in Section IV.4.2.2. Table B.2 (Appendix B) shows excerpts from the conversation trees for different

scaffold levels and targeting different types of task-oriented learning strategies in Betty's Brain. Since there were only six participants in the pilot study, we took a qualitative approach to track learners' temporal progress during learning in Betty's Brain and evaluate the impact of scaffolds they received throughout this process. Results from this analysis are reported in Munshi et al. (2022a).

These findings helped us further refine our scaffold design framework. For example, one of our observations from this analysis was that, some students ignored the mentor agent's suggestions in case of certain scaffolds repeatedly and engaged in their own learning strategies. One of these students was strategic in using quizzes to debug errors in their causal map but was not effective in using reading as a map debugging strategy. So, whenever they were suggested to perform strategic reading, the student ignored the feedback and proceeded with a quiz-based debugging strategy. This finding helped us provide further elaboration in the conversation tree of the specific trigger condition to emphasize to the learner how a a Read→Build scaffold would be more useful to address the current errors in their map for that specific task context.

Such lessons learnt from evaluating the prior iterations at scaffold design have helped us decide on the final set of triggering conditions and also allowed us to build more meaningful conversation trees and to provide more agency to the learner, with the conversation tree first *explaining* the purpose and context of the feedback (the strategy deficiency detected by the mentor), then asking whether they need any help with their strategy use (*diagnostic*), and only then delivering more *elaborative* feedback to fill in knowledge gaps (as applicable), and more *strategic* feedback to provide actionable recommendations on effective strategy use for problem resolution (e.g., note the conversation tree for Scaffold 1, described in the next section).

IV.4.4 Final Set of Triggering Conditions and Conversation Trees

The complete list of the adaptive scaffolds we have implemented in the most recent iteration of our scaffold design process, along with their triggering conditions, are presented in Figure IV.5 and discussed below:

1. *Read→Build* scaffolds (count = 3) for conceptual construction. Two of these are *correctness* scaffolds, with an objective to increase the students' awareness of map errors (viz., the presence of shortcut links), the reasons why such errors may have occurred (e.g., why a link may be wrong), and to suggest more effective Read→Build strategies to debug and fix errors. The third Read→Build scaffold is a a *coherence* feedback to help students perform Build (map edit) actions that are connected to the sections of the Science Book that they recently viewed.

   (a) **Read→Build Shortcut Link Feedback** *(Scaffold 1)*: This scaffold is triggered by a pattern, i.e., multiple occurrences of a *Read→Add Shortcut Link* behavior. (In Betty's Brain, a "shortcut link"

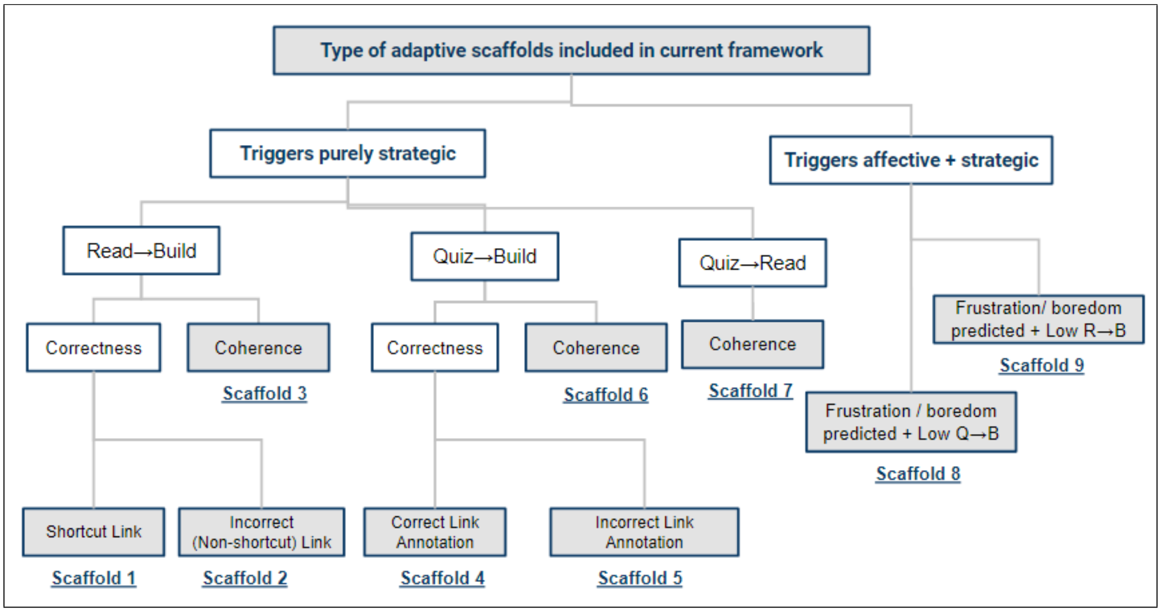


Figure IV.5: The Types of Adaptive Scaffolds Included in the Current Design and Implementation Framework

implies that the student has added a direct link between two concepts, e.g., A→C which should actually be related as A→B→C. In other words, the student has not added the complete chain of causal relations that explain a phenomena, i.e., the student is missing certain causal link(s) in their model.)

In the feedback provided to the student, the mentor agent Mr Davis first asks *(diagnostic)* if the student is aware of the notion of shortcut links. Depending on the student's response, the mentor explains the idea of shortcut links, i.e., a chain of links that provides the right answer to a query, but is missing links that fully explain the phenomena being modeled *(filling in knowledge gaps)*. To explain further, the mentor provides illustrative examples *(elaborative)*, and discusses why it is important to know and debug shortcut links in the causal model *(to increase task value and stimulate metacognitive monitoring and self-reflection behaviors)*. He confirms that the student's current causal map has one or more shortcut links, which have not yet been fixed *(increasing awareness)*. At the end, the mentor provides the student with actionable information *(strategic)* about pages or sections of the science book they can now read, and why this is a useful strategy to adopt (*to emphasize its value and encourage future strategy use)*.

(b) **Read→Build Incorrect Link Feedback** *(Scaffold 2)*: This scaffold is triggered by a *Read→Incorrect Link* pattern, where the student adds incorrect links to their map after reading the Science Book. This scaffold considers any case of incorrect link addition which is not a shortcut link, since the shortcut link scenario is handled by the previous scaffold.

In the feedback provided to the student, the mentor agent explains to the student the different reasons for a link to be incorrect *(to fill in potential knowledge gaps)*. These reasons include adding links between concepts that are not actually related, adding an incorrectly directed link (B→A instead of A→B), or adding the incorrect sign ("A decreases B" instead of "A increases B") for a link. He confirms that the student's current causal map has one or more incorrect links (from a specific section or starting from a specific source concept) which need to be fixed *(increasing awareness)* and provides actionable *(strategic)* information about pages or sections of the science book they should review. He also tells the learner to think through the reasons for incorrectness as they are reading the page and consider if any of them may apply in the current case (*encouraging metacognitive monitoring and self-reflection behaviors)*.

(c) **Read→Build Coherence Feedback** *(Scaffold 3)*: This scaffold is triggered by a *Read→Add Incoherent Link* pattern. Here, Betty (the teachable agent) initiates the conversation (to engage the student in more social interactions) and mentions that she has trouble understanding the recently added links because they do not appear to be related to the Science Book pages they (the student and Betty) were recently reading. Mr Davis then responds to Betty and tells the student that the reason for Betty's confusion is that the incoherent Read→Build behavior. He explains to the student *(strategic)* that they should add concepts and links related to what they recently read, and how such organized read-build behavior can help avoid Betty's confusion when teaching her new information.

The content of the feedback provided for each of the Read→Build scaffolds is further contextualized using the scaffold level (IV.4.2.2, as illustrated in Figure IV.4.

2. *Quiz→Build* scaffolds (count = 3) for model debugging. Again, two of these are *correctness* scaffolds, while the third one is a *coherence* scaffold. The correctness scaffolds aid in model debugging by teaching link-annotation practices that can help the student understand and use their recent quiz results more strategically to debug and correct the errors in their causal model. The coherence scaffold helps the student become more coherent in their model edits after a quiz.

(a) **Quiz→Build Correct Link Annotation Feedback** *(Scaffold 4)*: This scaffold is triggered by a pattern that involves the student taking a quiz (with at least one quiz answer graded correct, implying the presence of correct links on the map) and then not annotating (marking) the correct links on the map using the "Mark as correct" feature available in Betty's Brain. In this feedback, Mr. Davis takes the student through a short guided training session on marking links, by making them click on one of the correct answers from the recent quiz, highlighting the links associated with this answer, and explaining how to mark these correct links on their map, and why such annotation can be useful to confirm links that are correct and thereby make it easier to spot the links that may be incorrect. He then asks the student to mark the correct links currently present on their causal map.

(b) **Quiz→Build Incorrect Link Annotation Feedback** *(Scaffold 5)*: This feedback is similar to the previous one but it is triggered when the student has at least one wrong answer in their map, and they have used the ”Mark as correct” feature to confirm the correct links but not yet used the ”Mark as maybe wrong” feature to check the links that are possibly wrong and need to be reviewed. Mr. Davis teaches how to use the “Mark as maybe wrong” feature to further zone in on links that need to be reviewed. At the end, Mr Davis asks the student to ”Start conversation” if they need help with understanding very specific quiz grades (like the ”?” grade), since previous studies (Munshi et al., 2022a) have shown that students often get stuck with debugging model errors related to this specific quiz grade. (By starting a conversation with the mentor anytime, the student can get cognitive skill-level information to interpret the quiz grades which can be helpful to better understand and use the suggested cognitive-metacognitive strategy).

(c) **Quiz→Build Coherence Feedback** *(Scaffold 6)*: This feedback is triggered when the student takes a quiz, looks at quiz results, but then edits links that are not related to the quiz answers. Mr Davis explains that after taking a quiz, it would be helpful to first teach Betty the concepts she did not answer correctly. In this scaffold (like the last one), he suggests that the student start a conversation with him if they need help about understanding quiz grades.

3. **Quiz→Read Coherence Feedback** *(Scaffold 7)* (count = 1) for knowledge refinement: This scaffold is triggered when the student follows their Quiz views with viewing different Science Book pages that are incoherent (not the ones relevant to) the quiz answers they viewed. The idea behind this scaffold is that a student may be struggling to identify the sections of the Science Book relevant to their recent quiz views, so Mr Davis verifies this by first asking *(diagnostic/validational)* the student if they are trying to find something from the science book. If the student says yes, he then suggests them *(strategic)* to use the ”search box” feature in the Science Book to look for keywords related to the quiz answers, which would lead them to all sections of the book where these keywords were discussed.

4. *Affective/strategic* scaffolds (count = 2) to help the students regulate the impact of potentially negative affect states. These scaffolds are triggered by tracking students’ affect likelihood scores from the affect detector models in Betty’s Brain (Jiang et al., 2018) and upon identification of an affect transition that

ends in frustration or boredom (negative affect states). To provide actionable and useful feedback at these states (and also account for the possibility of incorrect affect detection by the detector models), the feedback provided at these moments does not include any mention of the detected emotion but includes cognitive strategies designed to improve current students' cognitive-affective states.

(a) **Negative affect + Quiz Feedback** *(Scaffold 8)*: This feedback is triggered when (a) the student shows the high likelihood of frustration or boredom in the Quiz Results view , (b) this view is dominant in the last 5 minutes, and (c) the student also shows a low (¡30%) Quiz→Build proficiency in the last 5 minutes. The premise is that the negative affect, if actually present, is due to the student having trouble understanding or using the quiz results. Mr. Davis explains that the first step for effective Quiz→Build is to understand what the grades mean. So, if the student does not know that, they should "Start conversation" with Mr Davis or ask an instructor present in the classroom. He further explains that once the student is able to differentiate between the grades, it would be easier for them to identify map errors. If the student is confident that they understand the quiz grades, Mr Davis suggests that they take some more time to read through the Science Book pages that discuss information related to the wrong answers and try to detect any errors.

(b) **Negative affect + Read Feedback** *(Scaffold 9)*: This feedback is triggered when (a) the student shows the high likelihood of frustration or boredom in the Read (Science book) view , (b) this view is dominant in the last 5 minutes, and (c) the student also shows a low (¡30%) Read→Build proficiency in the last 5 minutes. The premise is that the negative affect here is due to the student having trouble using their science book reading in an effective manner. Previous studies (Munshi et al., 2022b) have shown that low performers in Betty's Brain often get "stuck" in a reading loop and get disengaged. Mr. Davis therefore tries to break any possible Read→Read or Read→Build-Ineffective loop by reminding the student to take a quiz from time to time, and explaining why that can be helpful to check their progress and better understand which parts of the map need more work. He also assures the student that he is there to help so the student can ask him questions anytime or also go through the "Teacher's guide" in Betty's Brain for some tips on how to teach Betty.

In the following chapter, we discuss our approach for evaluating the current design of the adaptive scaffolding framework in Betty's Brain.

CHAPTER V

Evaluation of the Adaptive Scaffolding Framework

Our scaffold evaluation framework, to assess the impact of the adaptive scaffolds presented in Chapter IV, consisted of: (1) an experimental study to collect data from middle school students using Betty's Brain, and (2) data analysis to evaluate students' responsiveness and strategic usage of adaptive scaffolds they received during the study. The classroom study design and data collection procedure are described in Section V.1, and the data analysis methodology is discussed in Section V.2.

V.1 Classroom Study for Data Collection

We conducted an exploratory study with middle school students in late March-early April of 2022. The students were participants in a *Day of Discovery* program run by the School for Science and Math at Vanderbilt University. A total of $n = 55$ students participated for the duration of the study and consented to data collection. (This number accounts for factors like student absence on one or more days of the study, technical issues leading to data loss, and the subset of students who did not provide written consent for data collection). The 55 participants included 26 female students, 28 male students, and 1 non-binary student (46 white, 3 black, 1 Hispanic, 4 mixed race, and 1 with no race information). They were students from seventh ($n = 29$) and eighth ($n = 26$) grades came from three urban public schools in Nashville, TN. The study was conducted in the students' DoD classroom in the presence of their two instructors. No significant difference in prior knowledge (as determined from pre-test scores) was observed between the participants of the two grades. Similarly, there were no differences by gender or the students' schools. During this study, all students worked on the climate change unit of Betty's Brain, which involved learning (and teaching Betty) about the human causes and environmental effects of climate change by constructing a causal (cause-and-effect) model of this scientific process in the *causal map* in Betty's Brain.

V.1.1 Study Design

The study was conducted over a period of three days.

On Day 1, students worked on a paper pre-test (see Appendix C) for 30 minutes. The pre-test consisted of two parts:

(a) A test on climate change, which included a combination of multiple-choice (MC) and short-answer (SA) questions to evaluate students' domain knowledge and causal reasoning skills prior to interacting with Betty's Brain.

(b) A five-point Likert-scale questionnaire to assess student motivation (self-efficacy and task value), adapted from Tuan et al. (2005).

After completing the pre-test on Day 1, students worked on an introductory *training unit* of Betty's Brain for the remaining time (15 minutes). In this training session, Mr Davis, the mentor agent in Betty's Brain, provided students with a guided introduction to the resources and tools available in the learning environment.

On Day 2, students began working on the Climate Change unit of Betty's Brain. Each learner worked individually on their own laptop to build a causal model of climate change.

On Day 3, students continued building their climate change models in Betty's Brain for about 30 minutes. Then they were asked to respond to a post-test (Appendix C), which again included two sections: one on climate change (identical to the pre-test) and a second one on motivation. The post-test also contained an additional questionnaire on science anxiety.

V.1.2 Data Collection

The following data was collected from all 55 students during their interactions with the Betty's Brain system:

1. Responses to the pre- and post-tests;
2. Logged trace data in Betty's Brain on student activities in the Climate Change unit. This log data consisted of a sequence of timestamped activities that reflected students' interactions with the system. The log data was processed into an $< action, view >$ representation (Figure IV.3) to facilitate the interpretation of student behaviors *in context* (Kinnebrew et al., 2017). Map scores (discussed in Section V.2.1) derived from the logs helped us track students' performance in their causal modeling task in Betty's Brain;
3. Timestamped affective state predictions provided by the BROMP-based detector models embedded in Betty's Brain, which generated likelihood values of engaged concentration, boredom, delight, confusion and frustration at 20-sec intervals
4. Information on adaptive scaffolding provided by the system in terms of the timestamp at which the scaffolding was initiated, the type of scaffold that was provided, the trigger condition, the scaffold level, and the number of times this scaffold was provided.

Additionally, facial webcam videos were also collected using OBS Studio during Days 2 and 3 for a subset of students ($n = 18$). (Due to technical issues with certain laptops, webcam data could not be collected for all 55 participants.) The purpose of collecting face videos was to obtain learners' facial emotions by processing these videos offline using the iMotions AffDex API (McDuff et al., 2016). This was intended to serve the dual purpose of (a) providing a secondary emotion source to validate the emotion likelihoods generated by the affect detector models, and (b) being the primary source of affect data in case the reliability of predictions from the BROMP-based affect detector outputs could not be determined. (Being trained on a larger set of data points, the facial affect detector models are generally considered to be more robust than the BROMP-based action-driven affect detectors.)

Also, at the end of Day 3, audio interviews were conducted with some randomly selected students ($n = 13$), where they were asked about their experience working with the Betty's Brain system, and especially their thoughts on the usefulness of the adaptive feedback they received from the mentor agent while learning.

V.2 Data Analysis Procedure

Our overarching research question for scaffold evaluation was: **RQ** *What is the impact of adaptive scaffolds triggered by the framework on learners' self-regulated learning (SRL) processes?*

We started the data analysis towards answering this question by first defining our learning outcome metrics in Betty's Brain (Section V.2.1) and conducting an exploratory data analysis (Section V.2.2) by studying students' overall learning outcomes, behavioral and affect indicators from the classroom intervention. The data-driven insights obtained from this exploratory analysis allowed us to formulate a more targeted research question (Section V.2.3) and a corresponding set of targeted analyses for scaffold evaluation (Section V.2.4).

V.2.1 Learning Outcome Metrics

We developed two metrics, one summative and one formative, to measure students' learning outcomes from the Betty's Brain intervention study.

1. *Summative Assesment: Normalized Pre-to-post test learning gains*, calculated as $\frac{Post\,score - Pre\,score}{Max\,score - Pre\,score}$: This measure, derived from grading students' pre- and post-test responses using a pre-defined rubric, helps us to evaluate how well the intervention helps students learn their science content;
2. *Formative: Map scores*, calculated as *# number of correct minus incorrect causal links in a student's causal map at any point in time during the intervention*. This measure helps us track the correctness of students' causal models built in Betty's Brain over time, and provides us with a more direct measure of their causal modeling abilities in the system. We compute the *final map scores* achieved by each student at the end of the intervention as a second summative measure of their overall performance during the intervention.

V.2.2 Exploratory Data Analysis

An exploration into students' overall learning outcomes from the Betty's Brain intervention was performed using the metrics outlined in Section V.2.1. The findings, reported in Section VI.1, revealed that the study participants ($n = 55$) as a whole did not exhibit significant pre-to-post learning gains. However, large variances in both pre-to-post gains and final map scores were observed, which prompted us to apply an unsupervised learning approach to determine if students clustered into groups based on their behavioral differences and if this explained the wide variations in performance. In addition, behavior differences could also help us determine if the adaptive feedback provided were differentially used by the different groups. The results, reported in Section VI.1, revealed four clusters: *Cluster 1* ($n = 6$) consisting of disengaged learners, *Cluster 2* ($n = 19$) consisting of inefficient information appliers, *Cluster 3* ($n = 22$) consisting of strategizers, and *Cluster 4* ($n = 8$) consisting of experimenters or tinkerers who exhibited trial-and-error behaviors while working on Betty's Brain. We discuss the clustering approach and the labeling of each of these clusters in the next chapter.

V.2.3 Research Questions for Scaffold Evaluation

The clusters observed from the exploratory data analysis were used to frame a more targeted research question to study and compare the impact of adaptive scaffolds from our scaffolding framework on the four groups (clusters) of students.

RQ: *How did students from the four different clusters respond to receiving the different types of adaptive scaffolds (listed in Section IV.4.4?* More specifically,

(a) *Were students in each cluster* ***responsive*** *to the scaffold?* In other words, did they appear to follow the actionable recommendation provided in the scaffold by then performing the activities suggested by the mentor?

(b) *Additionally, did the subsequent behaviors of students in each cluster convey a* ***strategic use*** *of the scaffolds?* For instance, if the objective of a scaffold was to teach a specific cognitive-metacognitive regulation strategy, did the students show a change in their relevant model-building activities, behaviors and performance after they received the scaffold? If the scaffold also included an affective component, was it also possible to detect a change in learner emotions?

V.2.4 Temporal Analysis for Scaffold Evaluation

To answer the above research questions, we performed temporal analysis for student clusters (findings reported in Section VI.3) that tracked the change in their use of suggested strategies, their causal modeling performance, (and emotions, as applicable) as they received conversational feedback in the form of adaptive scaffolds from the agents. For this purpose, we created sequences of *scaffold-triggered* ***'before' and 'after' intervals*** across a student's learning timeline in Betty's Brain, so that we could compare student behaviors before receiving scaffolds to their behaviors after receiving scaffolds.

Here, the *after-interval* for an adaptive scaffold started when the scaffold was given to the student and continued up to the time the student received their next scaffold from the system. Similarly, the *before-interval* for an adaptive scaffold considered the time starting from when the last (previous) scaffold was given to the time when they received the current scaffold (Munshi et al., 2022b). For example, consider a student who received two adaptive scaffolds during the course of their learning session - Scaffold 1 at time t_i and Scaffold 2 at time t_j. For Scaffold 1, the student's *before* interval would be $[0, t_i]$ and *after* interval will be $[t_i, t_j]$, where the time 0 represents the start of the current session. Similarly, for the second delivery of a Scaffold 2, the *before* interval would be $[t_i, t_j]$ and *after* interval will be $[t_j, end]$, where *end* represents the end time of the session.

To determine the *responsiveness* and *strategic use* of the adaptive scaffolds received by students in each cluster, we studied the change in their cognitive activities (as relevant to the triggering context and conversation content of the delivered scaffold), their map scores (performance), and the likelihood of their prevailing affect states (when relevant and accurately available) in the *before* and *after* intervals for each feedback they received. The specific activities and strategy use to be assessed depended on the type of cognitive or metacognitive strategy (and as relevant, affect state) that was supported by the agent's feedback. For example, in the case of the Read→Build Incorrect Link scaffold, we first checked whether the learner followed the mentor agent's suggestion in the scaffold by reading the suggested page(s). Then we also assessed strategic usage by checking if their causal modeling performance improved in the following interval, suggesting that they were able to use the strategic reading to develop more effective model construction behaviors that fixed prior errors in their causal model. We also studied the impact of scaffolds after the first, second, ..., n^{th} time it was received by students in a cluster.

Chapter VI presents the findings from data analysis conducted following the methodology outlined in this chapter. The results from an exploratory analysis of students' learning outcomes are reported in Section VI.1.1, followed by a study of the affect indicators available from the study data, especially focusing on data reliability and its implications on further analysis. This is followed by feature construction for cluster analysis using causal modeling behaviors in Betty's Brain (Section VI.1.3). The clustering procedure and its results are reported in Section VI.1.4, which shows four behavioral profiles or groups among students. Then Section VI.2 presents the statistics on adaptive scaffolds received by these groups, and discusses the implications of these numbers on future refinement of the scaffold design framework. Section VI.3 goes deeper into the temporal analysis of the impact of scaffolds received by students in each group, with discussions on group and student responsiveness / strategic use of each scaffold. Section VI.4 presents case studies exploring the behavioral evolution and scaffold use of two students from the classroom study. Finally, while Sections VI.3, VI.3, and VI.3 discuss our inferences on the effectiveness of scaffolds and the implications on future scaffold designs in Betty's Brain, Section VI.4 discusses some of the major future research directions that can extend the findings from this dissertation, while also outlining the limitations from the current evaluation study that can be explored further in future work.

CHAPTER VI

Results and Discussion

Exploratory analysis of the data collected from the classroom study was conducted to investigate students' overall learning outcomes, study their behavior and affect indicators, and then study the impact of adaptive scaffolds provided in the system.

VI.1 Exploring Learner Outcomes, Affect and Behavioral Indicators

As a first step, we analyzed students' learning outcomes and their behavioral and affect indicators.

VI.1.1 Pre to Post Learning Gains and Causal Modeling Outcomes (Map scores)

The metrics discussed earlier in Section V.2.1: *normalized pre-to-post test learning gains* and *final map scores*, were used as measures of students' learning outcomes. The pre-post scores are reported in Table VI.1 and the distribution of final map scores is presented in Figure VI.1.

A Shapiro-Wilk test showed that the distribution of the pre-post learning gains was close to normal, with $W(55) = 0.97, p = 0.24$ indicating a non-significant departure from normality. This justified the use of parametric statistical tests that are presented in Table VI.1. One-way ANOVA tests of the students' summative pre-test and post-test scores show that the overall pre-to-post learning gains of all students ($n = 55$) were not statistically significant (i.e., $F = 2.7, p = 0.105$) and the overall effect size was small (i.e., Cohen's d $== 0.18$). Additionally, Table VI.1 also suggests a high variance (mean = 0.2, sd = 0.4) in students' pre-to-post learning gain scores, which were normalized between 0 and 1.

The distribution of students' final map scores in Figure VI.1 also suggests a high variance among students. The median map score obtained at the end of the Betty's Brain intervention was 7 (out of a maximum score of 25), with $mean(sd) = 8.76(9.1)$. A moderate correlation *Pearson's* $r = 0.6(p < 0.05)$ between students' pre-post learning gains and final map scores suggests that there was a relationship between students' performance in the Betty's Brain intervention and their pre-post learning gains in science content.

Due to the variation in learning outcomes across students, we decided to further explore their learning behaviors and affect states with the objective to uncover representative behaviors among groups of students and study how they may relate to their performance. A similar variance in learning outcomes had been observed among the participants of the prior study that evaluated a previous iteration of the adaptive scaffolding framework (Munshi et al., 2022b) as well as in other studies conducted using Betty's Brain, Munshi et al. (2018b). In a more recent study Munshi et al. (2022b), we divided students into "high performer" and "low performer"

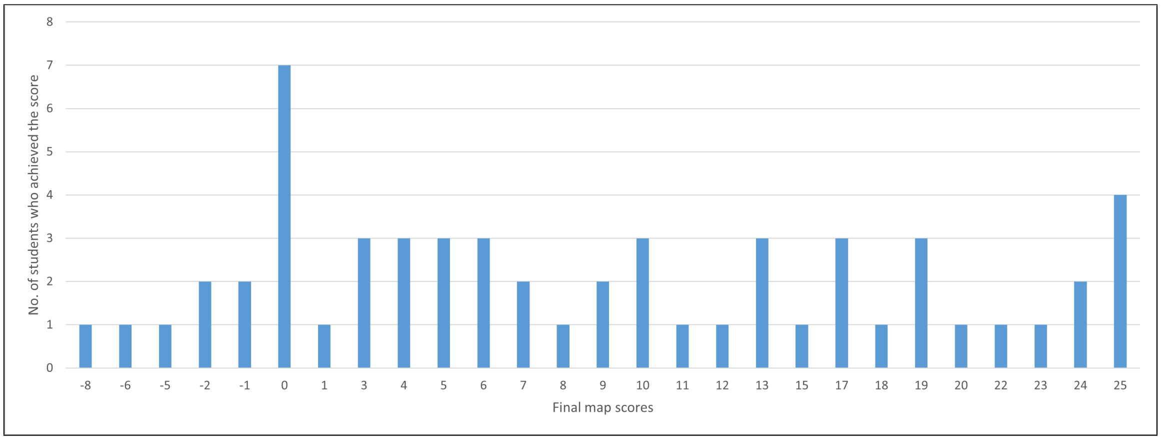


Figure VI.1: Learning outcomes from the Betty's Brain causal modeling task (final map scores) for all students (n=55)

Table VI.1: Pre-post learning outcomes of all students (n=55)

	Pre-post learning scores				
Pre-post question type	Pre-test score mean (sd)	Post-test score mean (sd)	Normalized learning gains mean (sd)	Pre to post 1-way ANOVA F-ratio (p-value)	Effect size Cohen's f
MC (Max=7)	5.2 (1.3)	5.7 (1.5)	0.23 (0.65)	3.8 (0.055)	0.18
SA (Max=16)	3.56 (1.5)	4.02 (2.7)	0.04 (0.17)	1.2 (0.28)	0.105
Overall (Max=23)	8.8 (2.3)	9.7 (3.7)	0.2 (0.4)	2.7 (0.1)	0.15

categories and assessed the impact of scaffolding on learner behaviors in these two groups, we took a more nuanced approach by deriving and analyzing students' behaviors and relate them to their performance and ability to use the feedback provided by the agents in the system.

In more detail, differences in self-regulation behaviors, including the regulation of cognitive-metacognitive as well as affective states, may have contributed to the difference in outcomes across students, therefore we studied both behavioral and affect indicators available from the log data collected in the study.

VI.1.2 Affect Indicators

The primary source of affect data collected during the Betty's Brain study, as discussed earlier in Section V.1.2, included likelihood scores of achievement emotions (engaged concentration, delight, confusion, frustration, and boredom) predicted by the BROMP-trained activity-based affect detector models (Jiang et al., 2018). This data was available as a set of affect likelihood scores every 20 seconds a student spent on Betty's Brain. A secondary source of affect data was collected from webcam videos recorded using OBS Studio for a random subset ($n = 18$) of the 55 total participants. (This data could not be collected for all 55 students due to memory issues with many of the student laptops, which slowed down and disrupted Betty's Brain activities when OBS Studio was turned on.), The facial affect data was detected by the iMotions AffDex API (McDuff et al., 2016) using action units derived from students' face videos as they worked in the Betty's Brain environment, and included likelihood scores of basic emotions (joy, anger, contempt, disgust, sadness, fear, surprise), as well as scores of confusion and emotional valence.

In previous work, we (Munshi et al., 2020) collected these two types of affect predictions (i.e., achievement emotion likelihood scores from the BROMP-based detectors and basic emotion likelihood scores from the AffDex-based detectors) during a Betty's Brain classroom study, synchronized the data by time, and used random forest classifiers to obtain a *mapping* of the affect data available from the two types of detectors. The results showed that predictions of *confusion* were mapped most closely to a likelihood of *high anger and high disgust*, while predictions of *frustration* mapped more closely to that of *high disgust and low fear*, or *low disgust, the presence of sadness, and low contempt*. Our objective with the Munshi et al. (2020) analysis was to use this mapping in future to explore the possibility of using commercial software like AffDex as a less context-sensitive and more development-friendly alternative to the BROMP-based affect detector models for predicting affect in academic situations. However, since using the AffDex models *online* to trigger adaptive scaffolds posed higher challenges, we used the BROMP-based detectors in our learner modeling approach, while collecting facial AffDex data as a secondary source of emotions.

In the current study, we performed a correlational analysis to check for mapping between the two emotion sources. The results, reported in Table VI.2, do not provide any evidence of the type of mapping observed before from Munshi et al. (2020). Additionally, the emotional valence scores obtained from the facial emotion detectors did not seem to match the expected valence for co-occurring affect state predicted by the BROMP-based detectors. In the absence of more accurate human coded emotion labels (such as ones used in Munshi et al. (2018c), it was difficult to further validate the affect predictions from the BROMP-based affect detectors. Therefore, to avoid biasing our subsequent analyses with potentially incorrect affect prediction data, we chose to focus more on behavioral (i.e., cognitive) features only to perform the cluster analysis and derive students' behavior profiles. (Even in the scaffold evaluation analysis reported in Section VI.3, we have avoided making claims about the affective implications of scaffolds like Scaffold 8 or 9, due to this lack of validated affect data. We will explore this further in future work.)

VI.1.3 Causal Modeling Behaviors

To construct features for cluster analysis that would provide an accurate insight into students' causal modeling behaviors in Betty's Brain, we first divided the *total system time* spent by a student in the Betty's Brain system into: (a) the time spent in the *Read* view, acquiring new information by reading Science Book pages, (b) the time spent in the *Map Edit* view, building the causal map by adding or modifying concepts and links, (c) the time spent in the *Quiz Results* view, assessing the state of the map by looking at quiz grades and Betty's causal explanations to quiz answers, and (d) the time intervals of at least 5 minutes where the student did

Table VI.2: Correlation matrix of the two types of affect data for 5 randomly selected students

	From BROMP-trained activity-based affect detectors					From facial video processed through AffDex									
	Boredom	Frustration	Confusion	Engaged conc.	Delight	Valence	Joy	Anger	Confusion	Engagement	Fear	Surprise	Disgust	Contempt	Sadness
Boredom	1														
Frustration	0.45	1													
Confusion	0.36	0.71	1												
Engaged conc	-0.08	-0.38	-0.18	1											
Delight	0.01	0.14	0.22	0.06	1										
Valence	0.29	0.04	0.07	0.04	0.08	1									
Joy	0.22	0.1	0.08	0	0.01	0.86	1								
Anger	-0.09	0.12	0.2	0.02	-0.02	-0.23	-0.02	1							
Confusion	-0.09	-0.03	0.04	-0.05	0.01	-0.28	-0.06	0.51	1						
Engagement	-0.02	0.11	0.05	0.03	-0.01	0.17	0.4	0.4	0.5	1					
Fear	-0.09	-0.04	-0.09	0.07	0	-0.12	-0.08	0.23	0.6	0.47	1				
Surprise	-0.07	-0.05	-0.1	0.07	0.01	0.09	0.16	0.17	0.5	0.51	0.87	1			
Disgust	-0.07	0.11	0.11	0.09	0.01	0.02	0.17	0.17	0.13	0.36	-0.01	0.25	1		
Contempt	-0.05	-0.02	-0.13	0.02	-0.08	-0.28	-0.05	0.01	0.03	0.22	-0.04	-0.05	0.23	1	
Sadness	-0.05	0.18	-0.06	-0.13	-0.16	-0.4	-0.06	0.14	0.07	0.25	0.04	0	0.11	0.77	1

not engage in any reading, quizzing or map editing activity and was likely to be in a *disengaged* or off-task state. We then looked into the time and effort spent by students on the cognitive activities within each view to derive more fine-grained behavioral indicators.

Reading the Science Book helped students acquire new knowledge about causal relations in the climate change domain, while viewing the Quiz Results allowed them to assess their progress and identify errors in their current causal model. Together, both these Read and Quiz Results views served as the sources of important information that could motivate map edits. Therefore, we denoted the time spent by a student in the Read or Quiz Results views as the **information viewing time**. Within the *information viewing time*, the periods of time, when a student read a Science Book page or looked at a quiz answer that contained potentially important information to support causal map edits and improve their map score, was further labeled as the **potential generation time**. This indicated how well a student was able to identify potentially useful information while reading or checking quiz results. But beyond looking at the sources of potentially important information, a further indication of the quality of students' read or quiz activities was their ability to then extract causal relations from these resources and translate them to link edits on the causal map. Therefore, we labeled the subset of *potential generation time*, which was actually used by the student to support future link edits, as the **potential usage time**. Within the Map Edits view, only the link edit activities had the potential to increase or decrease students' map scores. So, the **link edit frequency** was calculated to gain an insight into the number of link edits performed by a student over the time they spent in Betty's Brain. To further understand the quality of these link edits, we used the notion of *coherence* from Segedy et al. (2013) to label each link edit or link annotation activity as *supported* or *unsupported* compared to prior actions in Betty's Brain, thereby helping us to calculate the **unsupported edits percentage** and evaluate the quality of the link edit activities. The percentage of total system time when the student was off-task was calculated as the **disengaged percentage**. The final set of six behavioral features developed for our cluster analysis, some of them adapted from Segedy et al. (2013), are further described below.

1. **Information viewing percentage**: The percentage of *total system time* that formed the *information viewing time* (i.e., which was used for information viewing purposes by reading Science Book pages for $>=$ 10 seconds or by looking at a quiz result for $>=$ 2 seconds) was determined to be the *information viewing percentage*, and indicated the quantity of Read/Quiz activities performed by a student.

2. **Potential generation (from info viewing) percentage**: The percentage of the *information viewing time* that formed *potential generation time*, i.e., could support subsequent causal map edits to improve the map score, was determined to be the *potential generation percentage*, and suggested the quality of Read/Quiz activities.

3. **Potential usage percentage**: The percentage of *potential generation time* that formed *potential usage time* i.e., was actually used to support future map edits, was considered to be *potential usage percentage*. This helped further determine the quality of Read/Quiz actions.

4. **Link edit frequency**: The number of causal link edits performed by a student out of their *total system time* was determined to be the *link edit frequency*, and denoted the quantity of Build actions that could influence the map score.

5. **Unsupported edits percentage**: The percentage of link edits or annotations that were not supported by previous actions formed the *unsupported edits percentage*. This denoted the quality of Build actions.

6. **Disengagement percentage**: The percentage of *total system time* classified as *disengaged time* (i.e., no actions performed by the student for at least 5 minutes) was determined to be the *disengagement percentage*.

After feature construction, a set of three pre-processing steps were applied on the computed feature values to prepare them for clustering: (1) *Feature correlations* were used to assess feature independence. Results showed $Pearson's\ r$ in a range of $-0.4, +0.3$. (2) *Min-max normalization* was applied for data scaling, and removed variable bias while still preserving the variance in the data. (3) Finally, *coefficient of variation* ($CV = S.D./mean$) was used as a feature selection criterion to ensure that the features were informative. All 6 features had $CV > 25\%$ and were therefore used for clustering.

VI.1.4 Clustering Results Based on Behavioral Features

A hierarchical clustering approach (agglomerative, Euclidean distance metric, Ward's minimum variance method) was applied for clustering students. The code was implemented in the R programming language.

To determine the *optimal number of clusters* "*k*", the dendrogram generated was cut at several levels, to produce a number of groupings of the student data, starting from $k = 2\ to\ k = 7$. For each value of k,

Figure VI.2: Demographic information for the four groups obtained from cluster analysis

two metrics were computed: (a) *average silhouette coefficient* (Rousseeuw, 1987), (b) *Calinski-Harabasz (CH) index* (Calinski and Harabasz, 1974). To maximize the value of these two metrics (see Figure VI.3 and Figure VI.4) and obtain stable and interpretable clusters, the optimal number of clusters was selected as $k = 4$. The four clusters: **C1 (n=6), C2 (n=19), C3 (n=22), and C4 (n=8)** are shown in the dendrogram in Figure VI.5.

The demographic information for these groups obtained from cluster analysis is presented in Figure VI.2. The C1 group (n=6) included 5 female students (4 white, 1 Hispanic) and 1 male student (white). The C2 group (n=19) included 9 female students (7 white, 2 black) and 10 male students (8 white, 1 Hispanic, 1 black). The C3 group (n=22) included 11 female students (white), 10 male students (9 white, 1 with no race information), and 1 non-binary student (white). The C4 group (n=8) included 1 female student (white) and 7 male students (5 white, 1 Hispanic, 1 black). Table VI.3 further reports the demographic classification of students in the four groups, with the percentages suggesting between-group and overall differences in demographic distribution.

We analyzed the statistical differences in feature values between the four groups obtained from clustering (see Table VI.4). A 1-way ANOVA and Tukey-HSD test showed that the differences in *information viewing percentage* were statistically significant ($p < 0.05$) between C1 and C3, C2 and C4, and between C3 and C4. The differences in *potential generation percentage* were significant between C1 and C3, C2 and C3, and C3

Table VI.3: Number of students (percentage of group) in each demographic profile

Group	Female			Male				Non-binary	Total in group
	White	Black	Hispanic	White	Black	Hispanic	No race information	White	
C1	4 (67%)	0	1 (17%)	1 (17%)	0	0	0	0	6 (100%)
C2	7 (37%)	2 (10.5%)	0	8 (42%)	1 (5%)	1 (5%)	0	0	19 (100%)
C3	11 (50%)	0	0	9 (41%)	0	0	1 (4.5%)	1 (4.5%)	22 (100%)
C4	1 (12.5%)	0	0	5 (62.5%)	1 (12.5%)	1 (12.5%)	0	0	8 (100%)
All students	23 (42%)	2 (3.6%)	1 (2%)	22 (40%)	2 (3.6%)	2 (3.6%)	1 (2%)	1 (2%)	55 (100%)

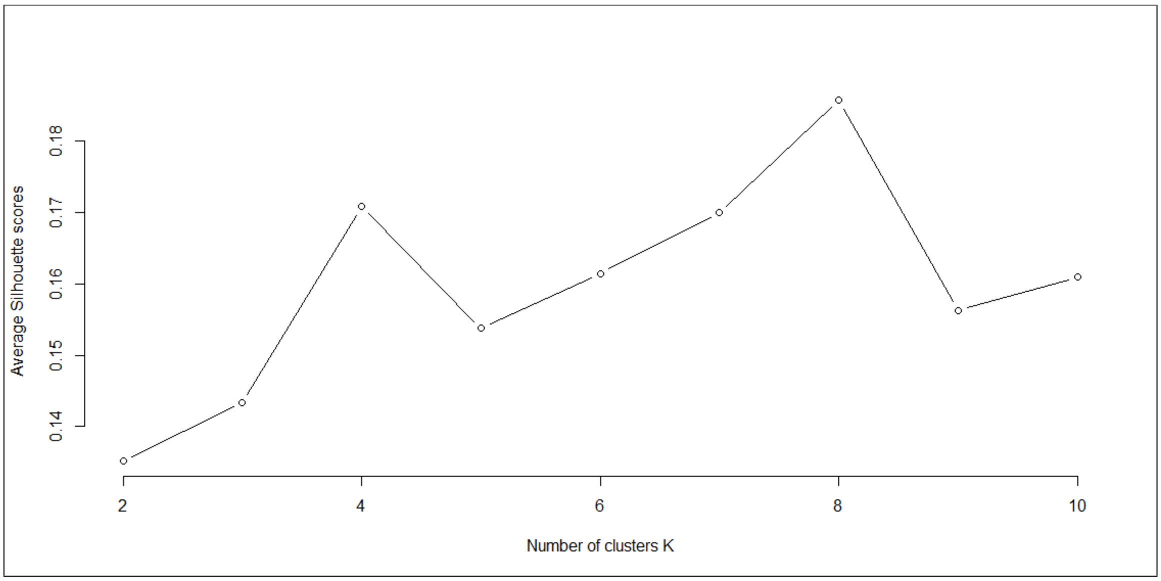


Figure VI.3: Average silhouette scores with number of clusters k

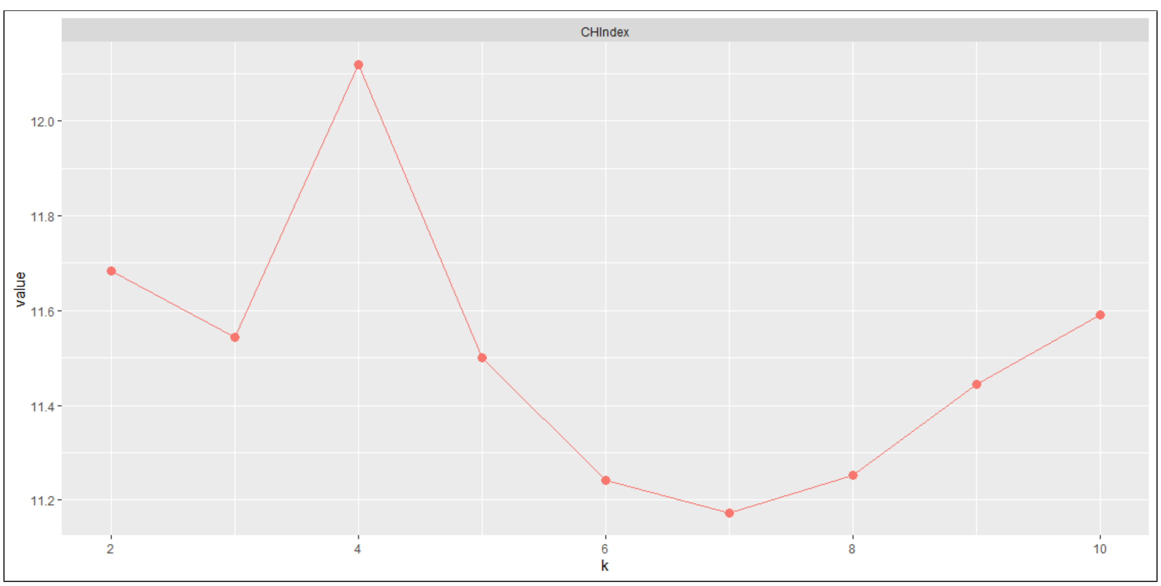


Figure VI.4: CH-Index plot with k cuts of dendrogram

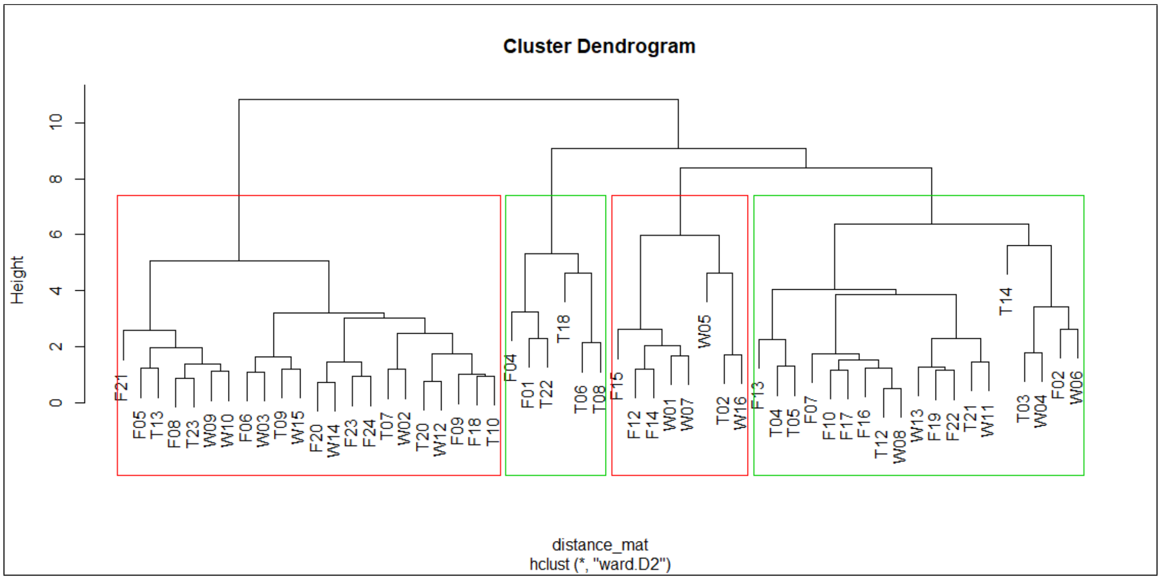


Figure VI.5: Dendrogram with k = 4 clusters

and C4. (C3 generated higher potential from reading and quizzing activities compared to all other groups.) The differences in *potential usage percentage* was significant only between C2 and C3, with C3 students using 72% of their generated potential, on average, to support map edits compared to 48% for C2 students. The differences in *link edit frequency* were significant ($p < 0.05$) between C1 and C4, C2 and C4, and C3 and C4, suggesting that C4 showed higher link edit frequency compared to all other groups. The differences in *unsupported edits percentage* were significant between C1 and C3, C2 and C3, and C3 and C4, with C3 showing a lower percentage of unsupported edits (i.e., higher model-building coherence) compared to all other groups. In terms of *disengagement percentage*, the differences were significant between C1 and C2, C1 and C3, and C1 and C4, with C1 showing high disengagement (28% of the time, on average) compared to the other groups ($<$ 0.06% on average).

The above findings and the results from Table VI.4 allowed us to interpret and label the behavioral profiles of students in each cluster/group, as discussed below:

1. C1 (Disengaged Group): The students in this group showed higher disengagement than students in the other groups, and the differences were statistically significant. In addition, students in this group spent only 38% of the time viewing information (lower than groups C2 and C3 but about the same as group C4, the tinkerers), with potential-generating information viewed for a net 17% of the total time spent in the system. This group also showed a low edit frequency, with high proportion (64%) of unsupported edits. In terms of learning outcomes, the C1 group showed the lowest average final map scores (3.16) and pre-to-post gains (0.04), although a high variance suggested some outliers in terms of performance.

Overall, students in this group were presumably *off-task* for a large portion of the intervention with the Betty's Brain Climate Change unit.

2. C2 (Inefficient Information Generators): This group of students spent a relatively large amount (52%) of their time viewing information (only Group 3 $-$ the strategic map builders were higher) but despite that, they converted what they viewed into a relatively small number of causal links (link edit frequency 0.28 $-$ only the disengaged group was lower with 0.24; Groups 3 and 4 were relatively higher). Similarly, they spent 29% (56% of 52%) time viewing potentially important information. They converted almost half (48%) of this information to map edits. 52% of the map edits for C2 were unsupported compared to prior Read or Quiz activities. This group also showed generally low final map scores and pre-post learning gains (higher than C1 but lower than C3 and C4). While these students spent more time viewing information than C4, they were not as efficient in finding relevant information or translating the information into link edits on their causal model. This was further observable from the low pre-post gains and final map scores.

3. C3 (Strategic Map Builders): This group of students spent 59% of time on the average viewing information (similar to C2) but were significantly better than all the other groups at generating useful information to building or debugging their causal maps (75% compared to 56%, 54%, and 46% for the other three groups). This group was also very successful in translating potential generated into correct link edits on their causal map and had fewer unsupported edits than all other groups, suggesting that they were careful and strategic map builders. This group showed the highest pre-to-post learning gains and final map scores among all four groups in Betty's Brain. This was also the only group with a statistically significant increase ($p<0.05$) in learning outcomes from pre- to post-test, with $F = 4.1, p = 0.032$.

4. C4 (Experimenters and Tinkerers): This group spent less time looking up the Science Book to learn causal information or checking quiz answers for building and debugging their maps (information viewing time = 37%; the disengaged group had a similar viewing time, but the time spent by the Strategic Map Builders and Inefficient Information Generators on Read and Quiz views was higher). The C4 students also showed a high edit frequency, with a high proportion (51%) of unsupported edits. But their Unsupported Edits were much larger than the Strategic Map Builders (51% to 28%). This group was not as good as C3 at identifying information from the Science Book or Quiz Results, but made frequent map edits (higher than all other groups), and while 51% of edits were unsupported, they were sometimes able to translate information to supported map edits, which may explain the final map scores (second highest average map scores among the four groups after C3). The pre-to-post learning gains

Table VI.4: Behavioral profiles of the four groups obtained from clustering

Label	Log-based behavioral features used for clustering						Learning outcomes (task performance / pre-post)	
	Information Viewing %	Potential Generation %	Potential Usage %	Link Edit frequency	Unsupported Edits %	Disengaged %	Final Map Score	Normalized Pre to post gains
C1: (n=6) Disengaged Group	0.38 (0.11)	0.46 (0.23)	0.58 (0.18)	0.24 (0.16)	0.64 (0.37)	0.28 (0.06)	3.16 (11.17)	0.04 (0.42)
C2: (n=19) Inefficient Info Generators	0.52 (0.12)	0.56 (0.18)	0.48 (0.21)	0.28 (0.13)	0.52 (0.2)	0.05 (0.05)	4.35 (8.28)	0.11 (0.39)
C3: (n=22) Strategic Map Builders	0.59 (0.14)	0.75 (0.07)	0.72 (0.06)	0.37 (0.08)	0.28 (0.13)	0.04 (0.04)	13.52 (8.28)	0.31 (0.27)
C4: (n=8) Experimenters / Tinkerers	0.37 (0.12)	0.54 (0.19)	0.69 (0.08)	0.63 (0.05)	0.51 (0.26)	0.04 (0.05)	12.83 (5.74)	0.27 (0.38)

showed a trend ($0.1 < p < 0.05$) with $p = 0.08$. Overall, this group showed behavioral characteristics that are suggestive of experimenters or tinkerers.

Beyond the significant between-group differences for metrics from Table VI.4, we also performed 1-way ANOVA with Tukey-HSD on the between-group prior knowledge (from pre-test scores), science anxiety (from post-study questionnaire responses), and motivation (from pre- and post-study questionnaire responses) scores, but found no significant differences ($p > 0.05$) between the four groups for any of these metrics.

We also note here that some of the groups obtained from this clustering analysis are similar to clusters observed from prior studies with Betty's Brain (Segedy, 2014). The following analysis now extends such prior work by going deeper into the impact of adaptive scaffolding on students in each group.

VI.2 Scaffolding Statistics: Across All Students, and Within Groups Obtained from Cluster Analysis

The number of adaptive scaffolds of each type received by students during the study is reported in Table VI.5. The 3rd column of this table reports the number of scaffolds received across all 55 students, while column 5 reports on the number of scaffolds (with the scaffold level specified, where applicable) received by students within each group. The table also reports the range, mean, and standard deviation of the number of times each group received a particular type of adaptive scaffold (columns 6-7), and the percentage of each group who received a scaffold 'n' number of times (columns 8-11), with n ranging from *0 (never)* to *3 or more times*.

From Table VI.5, we note that the highest number of a specific type of adaptive scaffold that students received during the study was the **Strategic: Quiz→Build**, with a total of 178 scaffolds delivered across the 55 students. This suggests that *students had more difficulties in using their quiz results* to debug their causal models, thereby satisfying the trigger conditions for Quiz→Build scaffolding. Additionally, the priority assignment algorithm (Appendix C), which was a component of the scaffold triggering framework, assigned *high priorities to the Quiz→Build scaffolds when students had a high number of links* (both correct and incorrect) on their map, to help students use the link annotation strategies like Scaffold 4 and Scaffold 5 to debug their models. This prioritization implied that students with many links on their causal maps received high numbers of Quiz→Build scaffolds, especially towards the end of their learning sessions when they usually had a more dense map.

The Quiz→Build model-debugging scaffolds included Scaffold 4 (delivered 62 times), Scaffold 5 (delivered 64 times), and Scaffold 6 (delivered 52 times). Of these, **Scaffold 4** was received by a relatively higher proportion ($> 40\%$) of C3, C2, and C4 groups compared to (17% of) the disengaged group C1. The highest counts of Scaffold 4 were delivered to the C3 group, followed by C4 (see Figure VI.5). Since this scaffold taught a "correct link annotation" strategy, its trigger condition required students to have correct answers in their recent quiz results (therefore, correct causal links on their maps). This may be the reason that the two high scoring groups C3 and C4 received this feedback more than the two low scoring groups C1 and C2. However, results from the temporal analyses in Section VI.3 suggest that the strategic map builders in C3 were more effective in using this feedback compared to the tinkerers in C4 (who did not apply the link annotation strategy to review incorrect annotations they had made before receiving the scaffold). We discuss in Section VI.3 how *such findings provide an opportunity to further improve the content of Scaffold 4* to account for behavioral characteristics shown by the tinkerers (e.g., by using ineffective "correct marks" on the student's map as a component of the scaffold triggering condition. Also, since Scaffold 4 was generally useful for responsive students, we may want to ensure that more students from the disengaged group C1 also receive this scaffold and learn this useful link annotation strategy. To this end, *future design of Scaffold 4 triggers may include* a more *dynamic* 'Threshold Comparator' (see Figure IV.3), where a diagnosis of disengagement from student activities would lower the threshold value for triggering this scaffold, enabling disengaged learners to receive the feedback at the next earliest opportunity. **Scaffold 5** and **Scaffold 6** were received at least once by similar proportions ($50\% - 67\%$ for Scaffold 5 and $41\% - 50\%$ for Scaffold 6) of each group. But with time, the C1 group gradually received higher counts of Scaffold 5 (eventually leading to significantly higher total counts compared to C3). However, with each receipt of Scaffold 5, C1 students also started doing activities like moving around map elements (indicative of disengagement) after scaffolding, which suggests that they did not find this scaffold very useful. So, in contrast to Scaffold 4, this is a situation where the scaffold may need to be gradually *faded away with time*, especially upon any diagnosis of disengagement after scaffolding. For Scaffold 6, there was not a lot of between-group differences in the count of the feedback delivered to students, but students were also not very responsive to this scaffold, so its inclusion in our design framework for the next iteration may need to be reconsidered.

With 116 scaffolds triggered across 55 students, high counts of the **Affective+Strategic** scaffolds are also observed in Table VI.5 and Figure VI.5. This may be attributed to the fact that these scaffolds (Scaffold 8 and

Scaffold 9), which had both affective and strategic trigger components, were *assigned higher priorities* in the design framework compared to their purely strategic counterparts (viz., the Read→Build and Quiz→Build triggers), with the reason being to prioritize the resolution of frustration or boredom when such affect states were predicted alongside deficiencies in cognitive strategy use. **Scaffold 8** was received most by C3 and C1 but had generally low responsiveness while **Scaffold 9**, received the most by C3 and C4, had high responsiveness. Students' cognitive activities after scaffolding (Section VI.3) suggest that the scaffolds were generally useful for the responsive students (with some exceptions, viz., C2 for Scaffold 9). In terms of affect, since the validity of the affect predictions from the BROMP-based detectors in our study could not be clearly established (Section VI.1.2), we refrain from making any inferences from the current data either on the affective implications of these scaffolds or relating the trigger counts to affective differences between groups. *Future studies will* use validated affect labels to rebuild these scaffolds and then perform further tests to decide whether to apply more group-specific trigger conditions based on any differences in emotion regulation observed between groups. (For instance, we may expect to see higher boredom predictions for the disengaged group using accurate affect data, given the type of cognitive-affective relations observed in boredom scenarios in D'Mello and Graesser (2012).)

The **<u>Read→Build</u>** type of scaffolds (Scaffold 1, Scaffold 2 and Scaffold 3) had a total trigger count of 103 (26 + 42 + 35), and were the only types of scaffolds delivered at two distinct *levels* of contextualization. **Scaffold 1** (Shortcut Link Feedback) was triggered for a high percentage of C3, with higher overall counts across time compared to the other groups. C3 also showed evidence of strategically using this scaffold to eliminate shortcut links from their map. One of the reasons for Scaffold 1 to be triggered more frequently in C3 may be the fact that *Scaffold 3 (coherence feedback) was assigned higher priority in our algorithm* compared to Scaffold 1 (correctness feedback), since we wanted to first teach students to be coherent in their Read→Build process and then provide correctness feedback to students who do coherent but ineffective Read→Build. This also explains why **Scaffold 3** was triggered so frequently for C4 (the experimenters who performed a lot of unsupported i.e., incoherent edits) compared to other groups. While Scaffold 3 helped C4 to improve their coherence and likely contributed to their potential usage times (Table VI.4), future design of the Read→Build type scaffolds should also ensure that the prioritization order between these three scaffold is more dynamic, so that all other groups can also receive and use feedback like Scaffold 1 that was very useful for the strategic map builders. **Scaffold 2** was frequently received by both C3 and C4 (see Figure VI.5) but the difference in C3's strategic map building and C4's characteristic tinkering behavior was observed in their use of this scaffold.

The disproportionately low trigger count of the **<u>Quiz→Read</u>** scaffold (**Scaffold 7**) compared to all the other scaffold types suggests that students generally showed more Read→Build and Quiz→Build behavior patterns (i.e., where they tried to use the information from Science Book pages or Quiz Results to build their maps) compared to Quiz→Read. Moreover, the priority assignment algorithm (see Appendix C) assigned a lower priority to the Quiz→Read trigger condition in the situations when students were performing less Build actions on their maps, in order to prioritize other types of scaffolding (like Read→Build) that would encourage them to develop more strategic map-building behaviors. The results from Section VI.3 suggest that C2 students responded to Scaffold 7 more than the other groups and the scaffold led them to find sections of the Science Book that were important for debugging their quiz results, although they still had trouble extracting causal relations from this section. This suggests that *in future*, the trigger condition for Scaffold 7 may be assigned a higher priority when students show behavioral characteristics of C2 (inefficient information generators) and the feedback should also be followed up with additional Read→Build scaffolding that then helps the student to extract the correct causal relations from the text they are reading.

Overall, when we look at the number of scaffolds of different types received by each group from Table VI.5 and Figure VI.5, we observe that **C1 received** a high number of Scaffold 5, followed by Scaffold 8, Scaffold 3, Scaffold 9, and very low or negligible numbers of other scaffolds. This includes a mix of Quiz→Build-correctness and Read→Build-coherence scaffolds. **C2 received** relatively higher counts of Scaffold 5, Scaffold 6, Scaffold 4 (the three Quiz→Build scaffolds) and Scaffold 9 (affective + Read→Build), followed by lower counts of the other scaffolds. **C3 received** high counts of Scaffold 9, Scaffold 1, Scaffold 4, Scaffold 8, and Scaffold 3. This includes both Quiz→Build-correctness and Read→Build-correctness scaffolds. **C4 received** high counts of Scaffold 3 and Scaffold 9 (both having Read→Build-correctness triggers), followed by Scaffold 2 and Scaffold 4 (Read→Build and Quiz→Build correctness scaffolds).

However, between-group differences in the number of scaffolds received were statistically significant over the duration of the study in two cases: higher numbers of *Scaffold 3: Read→Build Coherence Feedback* received by C4 compared to C2 and C3, and higher numbers of *Scaffold 5: Quiz→Build Incorrect Link Annotation Feedback* received by C1 compared to C3. Both of these cases have been discussed above, and in more detail in the next section. (Detailed descriptions of each of the adaptive scaffolds discussed above is available in Section IV.4.4, and their conversation trees have been presented in Figure IV.4.)

3 out of the total 55 participants in the study did not receive any adaptive scaffold from the mentor agent during their time in Betty's Brain. This included 2 students from C2 and 1 student from C4. For the remaining students, the *mean inter-scaffold time*, i.e., the average time interval between two consecutive scaffolds received by a student, was 6.1 minutes (s.d 7.2). The mean (sd) of inter-scaffold times per cluster, *in minutes*, were: 5.9 (7.4) for C1, 5.8 (6.9) for C2, 6.4 (7.7) for C3, and 5.9 (5.9) for C4. The difference in inter-scaffold time between clusters was not significant, with ANOVA (0.7) = 0.5. This suggests that the *inter-feedback interval* included in the current design allowed for a more uniformly triggered scaffolds across

time, providing students in each group with similar (and sufficient) amount of time to understand and apply each received feedback.

VI.3 Temporal Analysis on the Impact of Adaptive Scaffolds Received by Students in Each Group

To further address the research questions for scaffold evaluation outlined in Section V.2.3, the temporal analysis procedure discussed earlier in Section V.2.4 was applied to students' Betty's Brain log data to divide each learner's timeline into a set of *"before"* and *"after"* phases separated by the adaptive scaffolds they received from the mentor agent. The facilitated data analysis on the temporal changes in students' cognitive-metacognitive behaviors and model-building performance in Betty's Brain *before* versus *after* receiving an adaptive scaffold from our framework (see Section IV.4.4 for scaffold descriptions and Figure IV.4 for the conversation trees used to deliver scaffolds).

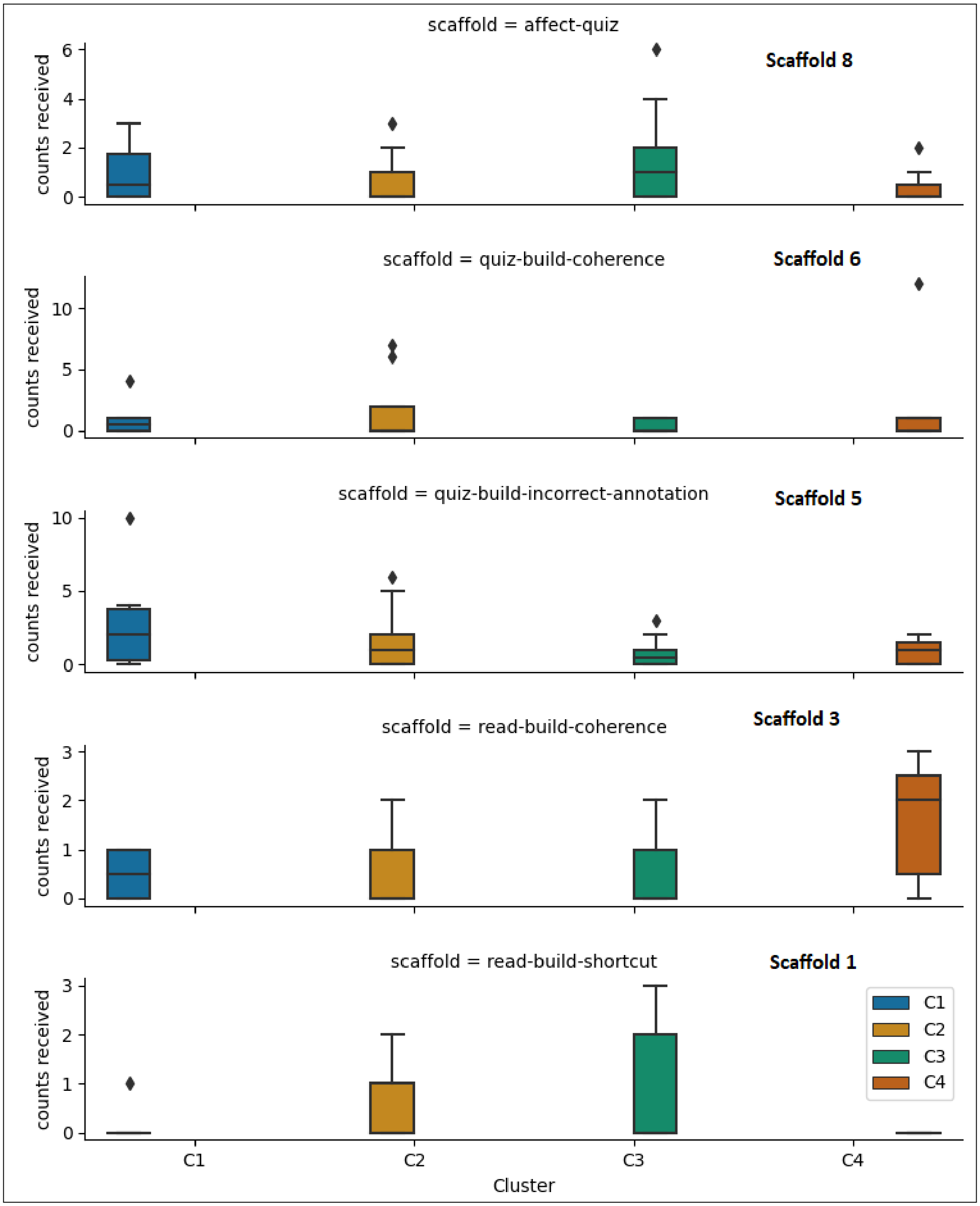
As outlined earlier in Section V.2.4, we evaluated the impact of the adaptive scaffolding along two primary directions. First, we checked for students' **responsiveness** to an adaptive scaffold they received, i.e., whether their activities and behaviors after scaffolding suggest that they followed the actionable recommendation provided by the mentor agent. Next, if a student was responsive to the scaffold, we assessed their **strategic use** of the scaffolding, by considering whether their behaviors and causal modeling performance suggested that the scaffold served its intended purpose by helping them adopt a more effective strategy for regulating their learning and model-building process. *For instance*, in case of a scaffold related to Build-correctness (e.g., Scaffolds 1-4 from Table VI.5) the strategic use of adaptive scaffolding would be determined by assessing the quality of students' link edits and annotations (whether such actions were effective, i.e., led to an increase in map score) after scaffolding. For a scaffold intended to promote Build-coherence (viz., Scaffolds 3 or 6), strategic use would be determined by the coherence of relevant Read→Build or Quiz→Build behaviors after scaffolding. For a scaffold intended to support cognitive-affective states, usefulness was determined from affect likelihood scores and cognitive activities relevant to the context of the delivered scaffold.

1. **Scaffold 1 *(Read→Build Shortcut Link Feedback)*:** Table VI.5 reports that the Shortcut Link Feedback was received by 16 students (29% of the 55 participants) a total of 26 times (16 Level-0 scaffolds and 10 Level-1 scaffolds). This scaffold was not received by any student in C4: the experimenters group, whereas 17% of C1 (n=1), 26% of C2 (n=5), and 46% of C3 (n=10) received this scaffold at least once during their time in Betty's Brain. Our inferences from these between-group differences in scaffold counts have already been discussed in Section VI.2.

   The objectives of Scaffold 1, as outlined in Section IV.4.4, were to: (a) help students understand *shortcut links* (fill in knowledge gaps) and explain why it is important to find and remove such links, (b) increase awareness of shortcut links present in the student's current causal model, and (c) explain how

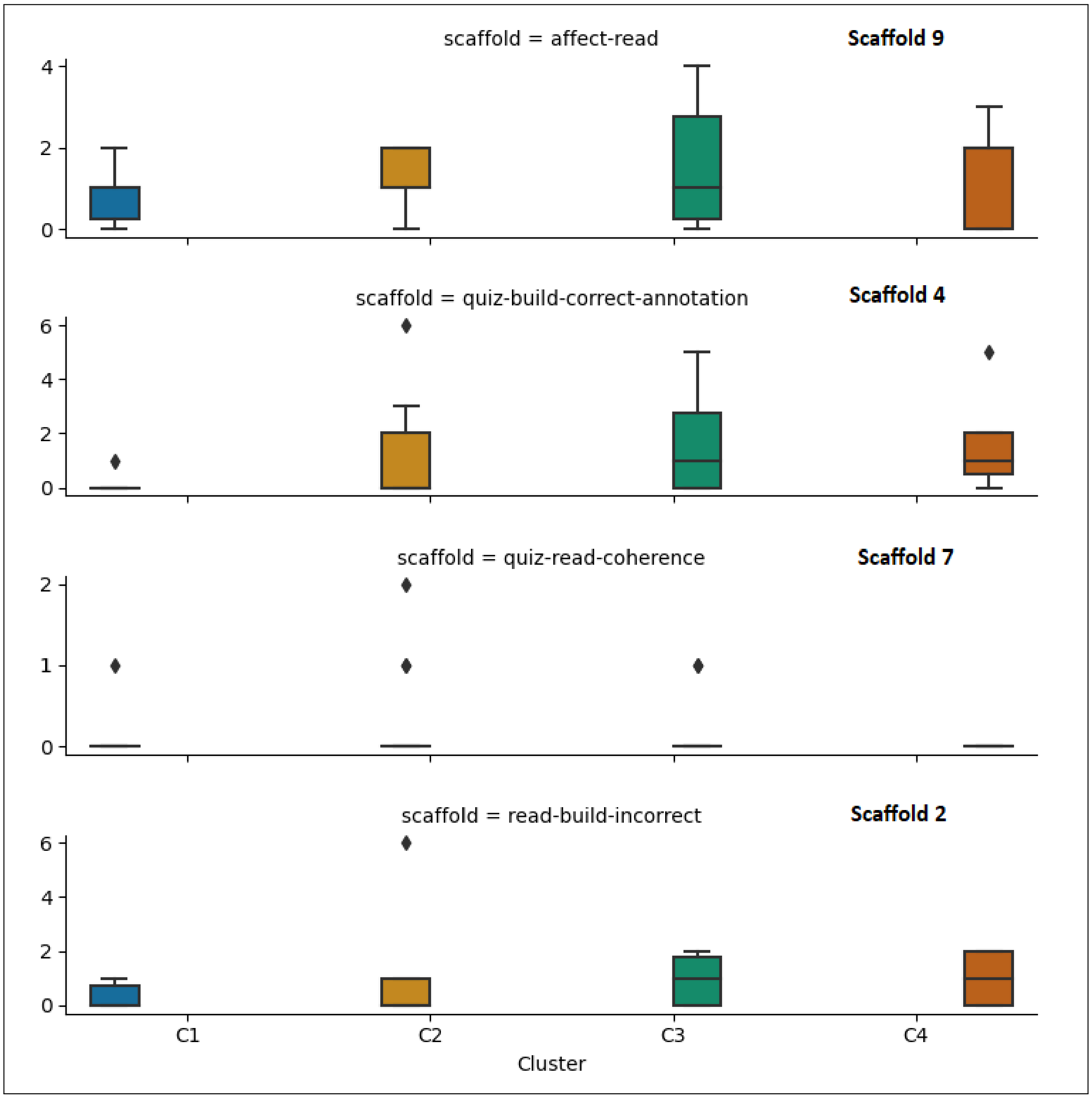
Table VI.5: Number of adaptive scaffolds received by the study participants, all students (n=55) and within in each group (C1, C2, C3 and C4)

Adaptive Scaffold		Total count received by all students (n=55)	Cluster/ Group	Within each group						
				Count received by group	No. of times a student got the scaffold		No. of students (% of group) who got the scaffold			
Type	Trigger				Range	Mean (SD)	never	1 time	2 times	3+ times
Strategic: R→ B	Read→ Shortcut Link Scaffold 1	26 (16 level-0, 10 level-1)	C1 (n=6)	1 (Level-0)	0-1	0.17 (0.4)	5 (83%)	1 (16.7%)	0	0
			C2 (n=19)	7 (5 Level-0, 2 Level-1)	0-2	0.4 (0.7)	14 (74%)	3 (16%)	2 (10%)	0
			C3 (n=22)	18 (10 Level-0, 8 Level-1)	0-3	0.8 (0.98)	12 (54.5%)	3 (13%)	6 (27%)	1 (4.5%)
			C4 (n=8)	0	0	0	8 (100%)	0	0	0
	Read → Incorrect Link (Non-shortcut) Scaffold 2	42 (28 level-0, 14 level-1)	C1 (n=6)	2 (Level-0)	0-1	0.3 (0.45)	4 (67%)	2 (33%)	0	0
			C2 (n=19)	12 (7 Level-0, 5 Level-1)	0-6	0.7 (1.4)	12 (63%)	6 (31%)	0	1 (6%)
			C3 (n=22)	21 (15 Level-0, 6 Level-1)	0-2	0.95 (0.8)	7 (32%)	9 (41%)	6 (27%)	0
			C4 (n=8)	7 (4 Level-0, 3 Level-1)	0-2	1 (0.9)	4 (50%)	1 (12.5%)	3 (37.5%)	0
	Read → Incoherent Link Scaffold 3	35 (23 level-0, 12 level-1)	C1 (n=6)	3 (Level-0)	0-1	0.5 (0.5)	3 (50%)	3 (50%)	0	0
			C2 (n=19)	8 (5 Level-0, 3 Level-1)	0-2	0.5 (0.8)	14 (74%)	2 (10%)	3 (16%)	0
			C3 (n=22)	13 (10 Level-0, 3 Level-1)	0-2	0.6 (0.7)	12 (54%)	7 (32%)	3 (14%)	0
			C4 (n=8)	11 (5 Level-0, 6 Level-1)	0-3	1.6 (1.2)	3 (37.5%)	1 (12.5%)	2 (25%)	2 (25%)
Strategic: Q→ B	Quiz → Correct Link Annotation Scaffold 4	62	C1 (n=6)	1	0-1	0.2 (0.1)	5 (83%)	1 (17%)	0	0
			C2 (n=19)	18	0-6	1.1 (1.5)	11 (58%)	3 (16%)	3 (16%)	2 (10%)
			C3 (n=22)	32	0-5	1.4 (1.6)	10 (45%)	2 (9%)	4 (18%)	6 (27%)
			C4 (n=8)	11	0-5	1.6 (1.6)	3 (37.5%)	2 (25%)	2 (25%)	1 (12.5%)
	Quiz → Incorrect Link Annotation Scaffold 5	64	C1 (n=6)	18	0-10	3 (3.5)	2 (33%)	1 (17%)	0	3 (50%)
			C2 (n=19)	24	0-6	1.4 (1.7)	8 (42%)	6 (32%)	2 (10%)	3 (16%)
			C3 (n=22)	16	0-3	0.7 (0.9)	11 (50%)	7 (32%)	3 (17%)	1 (4%)
			C4 (n=8)	6	0-2	0.9 (0.8)	4 (50%)	2 (25%)	2 (25%)	0
	Quiz → Incoherent Link Edits Scaffold 6	52	C1 (n=6)	6	0-4	1 (1.4)	3 (50%)	2 (33%)	0	1 (17%)
			C2 (n=19)	22	0-7	1.3 (2)	11 (58%)	3 (16%)	3 (16%)	2 (10%)
			C3 (n=22)	9	0-1	0.4 (0.5)	13 (59%)	9 (41%)	0	0
			C4 (n=8)	15	0-12	2 (4)	4 (50%)	3 (37.5%)	0	1 (12.5%)
Strategic: Q→ R	Quiz results → Incoherent Reads Scaffold 7	7	C1 (n=6)	1	0-1	0.2 (0.4)	5 (83%)	1 (17%)	0	0
			C2 (n=19)	4	0-2	0.2 (0.5)	16 (84%)	2 (10%)	1 (6%)	0
			C3 (n=22)	2	0-1	0.09 (0.3)	20 (91%)	2 (9%)	0	0
			C4 (n=8)	0	0	0	8 (100%)	0	0	0
Affective + Strategic	Negative affect in Quiz View + Low Q→B proficiency Scaffold 8	48	C1 (n=6)	6	0-3	1 (1.1)	3 (50%)	1 (17%)	1 (17%)	1 (17%)
			C2 (n=19)	11	0-3	0.65 (0.9)	12 (63%)	4 (21%)	2 (10%)	1 (6%)
			C3 (n=22)	28	0-6	1.3 (1.5)	8 (26%)	7 (32%)	4 (18%)	3 (14%)
			C4 (n=8)	3	0-2	0.4 (0.7)	6 (75%)	1 (12.5)	1 (12.5)	0
	Negative affect in Read View + Low R→B proficiency Scaffold 9	68	C1 (n=6)	5	0-2	0.8 (0.7)	2 (33%)	3 (50%)	1 (17%)	0
			C2 (n=19)	20	0-2	1.2 (0.8)	6 (31%)	6 (31%)	7 (37%)	0
			C3 (n=22)	34	0-4	1.5 (1.3)	6 (27%)	6 (27%)	4 (18%)	6 (27%)
			C4 (n=8)	9	0-3	1.3 (1.2)	4 (50%)	0	3 (37.5%)	1 (12.5%)



(e)

Figure VI.5: Box plots showing the distribution of number of adaptive scaffolds received by students in the four groups (x-axis: Cluster/group, y-axis: Number of scaffolds received)



(f)

Figure VI.5: (Contd...) Box plots showing the distribution of number of adaptive scaffolds received by students in the four groups (x-axis: Cluster/group, y-axis: Number of scaffolds received) strategic reading can help to debug these shortcut links, and (d) provide actionable recommendation on which section of the science book to review next to identify and debug the shortcut links. At the more contextualized level, i.e. Level-1, the mentor agent provided additional information to help students identify a shortcut link located on their map, such as by specifying the source concept of a shortcut link (see Figure IV.4).

To assess students' *responsiveness* to Scaffold 1, we analyzed whether they reviewed the Science Book pages suggested by the agent after receiving the scaffold. Additionally, since the purpose of this scaffold was to encourage the use of a more effective Read→Build strategy, we further evaluated *strategic usage* of the scaffolding by analyzing whether the student followed the review (Read) of the Science Book by then performing effective Build (causal link edit actions) to fix the shortcut links and improve their map scores.

(a) In **C1: Disengaged group** (n=6), only one student received Scaffold 1, at t=16 minutes into their learning session on their first day with Betty's Brain. In the interval *before* scaffolding, this student spent 38% of the time reading the Science Book and 62% of the time editing the causal map. During this time, the student added 9 total links (7 correct and 2 incorrect) on their map which were supported by their prior reading. While the student appeared to be working mostly in a coherent and effective manner, the addition of two incorrect shortcut links (which triggered the adaptive scaffolding) suggested that they needed to review the Science Book more carefully to identify the more complete causal relations between concepts on their map. In the interval *after* receiving Scaffold 1, no marked change in *disengagement percentage* (periods of time with no task-related activity) was observed for this student, who still spent a majority of time (67%) on their causal map, performing 10 correct link edits and 2 incorrect link edits. A deeper analysis into the temporal sequence of activities after receiving the scaffold provides a better assessment of scaffold *responsiveness*. After scaffolding, the student *did not* review the suggested Science Book page but instead deleted only the most recently added causal link (one of the two incorrect shortcut links that had triggered the feedback). Then, without spending further time on strategic reading to debug the other shortcut link present on their map, they proceeded to add new links associated with other concepts on their map. So while they deleted the more recent shortcut link, the other one still remained on the map.

    This suggests a heavy reliance on intuition versus a careful inspection of the mentor's feedback. Similar behavior after Read→Build-correctness scaffolds was reported in Munshi et al. (2022a), where a student had treated such scaffolds as purely *corrective* hints instead of *strategic* feedback, by assuming that the feedback was triggered due to the most recent link addition, and thereby deleting this link and missing other shortcut link(s) on their map which also informed the trigger condition for the scaffold. Based on the results from Munshi et al. (2022a), the conversation of the mentor agent in the current design of Read→Build-correctness scaffolds was refined to emphasize to the learner the strategic aspect of the feedback and how this strategy should be used to debug the multiple shortcut links in the map. However, as reported above, it appears that this modification was still not sufficient to encourage scaffold responsiveness in the student from C1.

(b) In **C2: Inefficient Information Generators group** (n=18), 5 students received Scaffold 1. This includes 3 students who received the feedback one time each (only Level-0) and 2 students who received it two times each (both Level-0 and Level-1). First, *responsiveness* checks showed that this group also mostly did not respond to Scaffold 1 by reading. After the first time receiving the feedback, only 2 of the 5 students from C2 reviewed the Science Book page suggested by the mentor agent, and none of these two students were then able to translate their Reading into Build activities that fixed the shortcut links present on their maps. We note that all 5 students took a quiz shortly after, but were also ineffective in the use of quiz results to fix the shortcut links. (They could have used the quiz grades and Betty's causal explanation to an incorrect grade to identify an erroneous part of their map, and then query Mr Davis about this part of the map, in which case he would have helped them understand and identify that these errors were due to the presence of shortcut links.) For the 2 students who received Scaffold 1 a second time, both followed the feedback this time by reading the suggested pages, but then they added other links from these pages while still not deleting the shortcut link, suggesting that they had not understood the concept of shortcut links explained by the scaffold. We note that this Inefficient information generators group generally had trouble extracting information for effective model-building, as seen in Table VI.4.

(c) In **C3: Strategic Map Builders group** (n=22), 10 students received Scaffold 1, the highest number among all four groups. In the interval *before* scaffolding, this group spent, on average, 49% of time editing their maps, 37% time reading the Science Book, and 14% time checking quiz results. While there was no significant change in view durations after scaffolding, a deeper look into learners' action sequences provide more insights on their scaffold responsiveness.

After receiving Scaffold 1 *the first time*, 7 students (70%) reviewed the suggested Science Book page (spending $>=$ 10 seconds on a page), suggesting that they were *responsive* to the feedback. In terms of strategic use, all of these 7 students followed the Read action by deleting the shortcut links on their map, further demonstrating an strategic and effective use of the Read→Build strategy discussed by the mentor. For the remaining three students who did not respond to the scaffold, their shortcut links were not removed from their maps in the "after" interval. (We note that one of these students changed the label of a shortcut link on their map to include the word "indirectly", i.e., A 'indirectly' increases B, suggesting that they understood the concept of a shortcut link. However, they did not perform Build actions that would have fixed the link, viz., deleting it and replacing it with the more complete causal relations, so the error remained on their map.)

All 7 students who received Scaffold 1 a second time responded to the feedback by reading the suggested Science Book page and successfully fixing their recent shortcut link, however this time a majority of the students spent less time viewing the Science Book. This may be related to the fact that the scaffold they received this time was a Level-1 feedback (more contextualized hint, which clearly specified the source concept of the shortcut link), so they were able to identify the shortcut link more easily and may not have needed to spend as much time reading. The one student who received Scaffold 1 a third time (again, level-1) also demonstrated an effective Read→Build behavior that fixed the erroneous link.

(d) No student from **C4: Experimenters group** received Scaffold 1. Our inferences from this result are outlined in more detail towards the end of the discussion below.

Discussion: The above findings suggest that the Read→Build Shortcut Link Feedback was effective for students who were responsive to the scaffold interpreted the strategy feedback correctly. While the student in C1 who received this scaffold treated it as a corrective hint and was unsuccessful in fixing the shortcut links on their map, the students from C2 also were primarily not very responsive to the mentor's suggestion, in terms of reading the suggested Science Book page, especially the first time they received the scaffold. Even when they read the pages identified by the mentor agent, these students had trouble translating the information to correct causal link edits. (Students from C2 generally not responding to Scaffold 1 with follow-up reading the first time when they received Level-0 feedback versus more responsive the second time when they received more contextualized Level-1 feedback may be related to the finding from Munshi et al. (2018c) which showed that low performers in Betty's Brain generally prefer more direct hints compared to indirect or general ones.) In contrast, a majority of C3 ("Strategic Map Builders") responded to the scaffold and successfully debugged their shortcut links. These findings suggest that the difference in how students from C3 took advantage of the feedback provided by the mentor in Scaffold 1 may have been a possible contributor to their higher final map scores. We also note that no students from C4 (the tinkerers) received this feedback. In Section VI.2, we noted the higher priority assigned to Scaffold 3 compared to Scaffold 1 in our priority assignment algorithm, which led to C4 receiving Scaffold 3 when the trigger conditions for both Scaffolds 1 and 3 were satisfied at the same time. Since C4 had a high percentage of unsupported edits, such a scenario was frequent, leading them to receive higher counts of Scaffold 3 and miss out on Scaffold 1 feedback. As discussed in Section VI.2, in future we will consider using a more dynamic prioritization order so that students who have received sufficient number of a particular type of scaffold are allowed to receive other types of scaffold (e.g., Scaffold 1 here), even if the highest priority trigger (like Scaffold 3) is active at the same time.

2. **Scaffold 2 *(Read→Build Incorrect Link Feedback)*:** From Table VI.5, the Incorrect Link Feedback was received by 28 students (out of 55 study participants) a total of 42 times (28 Level-0 scaffolds and 14 Level-1 scaffolds). 33% of C1 (n=2), 37% of C2 (n=7), 68% of C3 (n=15), and 50% of C4 (n=4) received this scaffold at least once during their time in Betty's Brain.

   This scaffold was delivered to support Read→Build-correctness behaviors except the Read→Add Shortcut Link case which was handled separately by Scaffold 1. While delivering Scaffold 2, the mentor agent first listed to the student the reasons for a link between two concepts to be incorrect (viz., adding incorrect sign for a link, directing a link incorrectly, or adding links between unrelated concepts). The agent then suggested that the student read a specific section of the Science Book while critically thinking whether these reasons may apply to links on their map from that section, in which case these should be debugged by appropriate Build actions on the map. Similar to Scaffold 1, we assessed students' *responsiveness* for Scaffold 2 by analyzed whether they reviewed the suggested Science Book pages after scaffolding. *Strategic usage* was assessed by checking whether their subsequent link edits on the causal map were *effective*, i.e., led to an increase in the map score.

   (a) In the **C1: Disengaged group** (n=6), 2 students (33%) received Scaffold 2. In terms of *responsiveness*, one of these students actually read the suggested Science Book page after scaffolding whereas the other did not. The responsive student reviewed the page for > 1 *minute* and followed this Read action by effective Build actions, deleting both incorrect links that had triggered the feedback, thereby suggesting *strategic scaffold use*. A comparison of the general quality of this student's Build actions before versus after scaffolding showed a shift from 100% ineffective link edits in the *before* interval to 100% effective link edits in the *after* interval, suggesting an improved Build-correctness after following the suggestion in the scaffold. Disengagement also decreased from 32% before scaffolding to 17% in the interval after scaffolding.

The other student, who was not responsive to the suggested Read→Build strategy, reviewed their quiz results after scaffolding, suggesting a Quiz→Build behavior, but ineffective Build actions after Quiz suggest they were also unable to apply an effective Quiz→Build strategy. Overall, this student went from 55% effective links before scaffolding to 40% effective links after scaffolding.

(b) In the **C2: Inefficient Information Generators group** (n=19), 7 students (37%) received Scaffold 2, with 6 of these students receiving the feedback one time each and 1 student receiving it a total of 6 times. Tracking the log data for *scaffold responsiveness* showed that, *after* the first time receiving Scaffold 2, only 2 out of 7 students from this group reviewed the suggested Science Book page for $>= 10$ *seconds* - one of these two students then deleted both incorrect links from their map, denoting *strategic use* of the feedback, while the other student still did not fix the links after reading, suggesting an inability to apply the information acquired from the page to the causal model. (This aligns with the general "Inefficient Information Generator" profile of this group.) We note that students who did not read the Science Book page after Scaffold 2 did check their progress after the scaffold by reviewing recent quiz results or querying Betty about recently added links. So while these students did not follow the mentor's suggestion to develop a Read→Build strategy, they exhibited a Quiz→Build behavior. This may correspond to a low ability to read and interpret the Science Book leading to a preference for the use of the Quiz Results. However, the ineffectiveness of their Build actions after Quiz show that they were unable to fix the incorrect links that had prompted the scaffold, thereby also suggesting a Quiz→Build-ineffective behavior.

Only one student from C2 received Scaffold 2 more than once, getting it a total of 6 times - this student did not read but deleted the most recent link every time they got the feedback, thereby not fully eliminating the two incorrect links that triggered the scaffold, and then moved on to other parts of the map, suggesting that they had treated the scaffold as a corrective hint instead of developing an effective model-debugging strategy as intended by our scaffold design. (Again, this result is similar to findings from Scaffold 1 evaluation.)

(c) In **C3: Strategic Map Builders group** (n=22), 15 students (68%) received Scaffold 2, including 9 students who received the feedback once and 6 students who received it two times. The first time these students received Scaffold 2, 11 students (73% of those who received the scaffold) were *responsive* to Mr Davis' suggestion by reading the relevant Science Book pages (that contained incorrect links) for $>= 9$ *seconds* in the "after scaffold" interval. All of these 11 students then showed a similar pattern where they went back to the causal map and deleted the incorrect link, demonstrating *correct use* of the strategy scaffold. The 4 less responsive students reviewed quiz results (75%) or queried Betty on her causal relations after scaffolding, and although they did not fix the specific incorrect links that had triggered the scaffolding, this self-monitoring behavior still led to the addition of other effective links on their map. For the 6 students in this group who received Scaffold 2 a second time (Level-1), 5 students (83%) were both responsive and demonstrated effective scaffold use through their Read→Correct Link Edit behavior. The sixth student responded to the feedback by reviewing the suggested page but added a link different to the one scaffolded for.

(d) In **C4: Experimenters group** (n=8), 4 students (50%) received Scaffold 2, one of them getting the feedback only once while the other three received it 2 times each. After the first time receiving Scaffold 2, 3 students (75%) responded to the scaffold by reading, and 2 of the responsive students fixed then used the Read→Build strategically to fix incorrect links on their map, while the third student just changed the sign of a link that should have been deleted for the edit to be effective. (From Figure IV.4, this scaffold taught the student the three different reasons for a link to be incorrect, and how to read strategically to figure out which reason is applicable in a specific case. The third responsive student here was unable to apply the right reason to their Read→Build, and therefore changed the sign instead of deleting the link.) The fourth student, who did not respond to the scaffold (i.e., did not read after) instead changed the sign of the link they had added and took a quiz to verify. Changing the sign of a link without being guided by information acquired from the Science Book shows the "tinkerer" / "trial-and-error behavior" associated with C4 earlier in our findings.

For the three students in C4 who received this scaffold a second time (level-1 feedback), none of them read the suggested Science Book page but directly started editing their maps, again showing a more tinkering approach compared to the more strategic Read→Build behavior exhibited by C3 after this scaffold. For 2 of the 3 students, the edits were effective in removing the error, whereas for the third student they were not effective.

Discussion: Scaffold 2 was generally useful for the responsive students. In C1, the response for the strategic feedback was 50%, with the responsive student then being able to successfully debug their map. In C2, scaffold responsiveness was low (28%). Even in the case of Scaffold 1, we had observed a low response rate for C2 which may again correspond to their low ability to read resource pages and generate new information for model-building. Even for responsive C2 students, the scaffold use was 50% effective. C3 demonstrated both high (73%) responsiveness and a strategic use of the behavior suggested by the mentor agent in the scaffold. C4 was generally more responsive to the Level-0 (more general) feedback (75%) by reading suggested pages but did not apply a Read→Build strategy for the Level-1 feedback.In future, we plan to build more aggregated models of students' learning behaviors in Betty's Brain, which would use a learner's behavioral profile (e.g., inefficient information generator) at the point of scaffolding to deliver additional feedback or lead a more guided training session to help the student develop the critical monitoring and self-reflection behaviors they lack at that point, for instance, to successfully extract the correct causal relations from Science Book text.

3. **Scaffold 3 *(Read→Build Coherence Feedback)*:** From Table VI.5, the Scaffold 3 was received by 23 students a total of 35 times (23 Level-0 scaffolds and 12 Level-1 scaffolds). 50% of C1 (n=3), 26% of C2 (n=5), 46% of C3 (n=10), and 62.5% of C4 (n=5) received this scaffold at least once during their time in Betty's Brain. It is noted here that unlike the Scaffold 1 and Scaffold 2 cases where there was no statistically significant difference between the number of scaffolds received by students in one group versus another, some significant differences (1-way ANOVA $p < 0.05$) were observed for Scaffold 3, with C4 receiving significantly higher counts of this feedback compared to both C2 and C3.

   The only purpose of Scaffold 3 was to help students be more coherent in their Read→Build behavior, so both responsiveness and strategic use of this scaffold were tested by computing the same metric: *coherence* or *support* (Segedy et al., 2015), and comparing the values in *before* versus *after* scaffolding intervals. Read→Build coherence was computed as *supported edits percentage*, i.e., the percentage of total Build (link edit) actions after Reading that were supported by (in other words, relevant to) the recent Read actions.

   For **C1** and **C2**, no significant difference was observed in the supported edits percentage of scaffolded students after receiving the feedback, therefore no impact of the scaffold can be claimed. But for **C3**, there was a significant increase in the supported edits percentage in the intervals *before* versus *after* scaffolding, with average supported link edits percentage being 69% across *before*-feedback intervals, 89% after receiving Scaffold 3 the first time, and 92% the second time. **C4** showed significant improvements in supported edits percentage after receiving Scaffold 3 for the second and third times, with average unsupported edits moving from 50% before scaffolding to a range of 74-77% in the intervals after the conversation with Betty and Mr Davis.

   Discussion: The above results suggest that the Read→Build coherence scaffold was somewhat useful for both C3 and C4. While the significant difference in the coherence metric was observed when C3 was scaffolded the first time, the difference was significant the second and third times for C4, suggesting that C3 paid closer attention to the feedback the first time while C4 placed higher value on the feedback later in their learning session. Earlier, Table VI.4 had suggested that C4, despite having generally high percentage of unsupported edits and annotations on their map, also had a high (69%) potential usage, implying that this group also spent a large amount of time on coherent link edits after reading or checking quiz results. This, combined with the above result from evaluation of Scaffold 3 (and the fact that students in C4 received significantly higher counts of this feedback even compared to C3), makes it likely that the increased coherence after scaffolding contributed to the potential usage times reported in Table VI.4.

4. **Scaffold 4 *(Quiz→Build Correct Link Annotation Feedback)*:** Table VI.5 reports that this Quiz→Build correctness was received by 26 students a total of 62 times. 17% of C1 (n=1), 42% of C2 (n=8), 54% of C3 (n=12), and 62.5% of C4 (n=5) received this scaffold at least once during their time in Betty's Brain.

   As discussed in Section IV.4.4, this guided scaffold was delivered to help learners adopt a link annotation behavior where they would annotate or mark links (using the "Mark as correct" feature in the causal map in Betty's Brain) that were graded as correct in recent quiz results. The scaffold was triggered by a Quiz→Build-incorrect pattern that suggested an inability to debug the map from quiz assessments. The scaffold was only provided if the learner's most recent quiz results included correct answers whose links had not yet been marked on the map. This was intended to serve as a memory aid which could help students spot incorrect links on their map and debug them more easily. A similar hint was delivered in a previous iteration of the adaptive scaffolding framework analyzed in Munshi et al. (2022b) and findings from this design iteration informed the current trigger conditions and feedback content with a view to increase learner engagement with the scaffold and responsiveness after scaffolding.

   We assessed the responsiveness for this scaffold by tracking students' correct link annotation behaviors before and after scaffolding, i.e., how many links they marked and what percentage of these links were marked correctly.

   (a) In **C1: Disengaged group** (n=6), only one student (17%) received Scaffold 4. In the interval before scaffolding, this student had already marked two links on their map correctly. After the scaffolding, the student correctly marked one more link from their recent quiz result. So, while this student was responsive to the scaffold, their link annotations before scaffolding were already

100% effective, which continued after receiving the feedback. So, we cannot claim any change in their link annotation strategy use due to Scaffold 4. The effectiveness of link edits (which contributed to map score) for this student stayed in a 50-56% range both before and after scaffolding.

(b) In **C2: Inefficient Information Generators group** (n=19), 8 students (42%) received Scaffold 4, 3 of them one time each, 3 other students two times each, and 1 student received it three times. *Before* receiving Scaffold 2, only 5 of these 8 students had marked the correct links from quiz results on their maps, with net effectiveness of their link annotations being 63%. (An effective link annotation here means marking a correct link as ”correct” on the map and an ineffective link annotation implies marking an incorrect link as correct on the map). *After* receiving Scaffold 4 *the first time*, only one student showed *scaffold responsiveness* by adopting the link annotation behavior suggested by the mentor agent. This student had not marked any links in the interval before scaffolding, but after receiving the scaffold marked 23 total links on their map, with 87% effective annotations. So, while there was low responsiveness for Scaffold 4 in C2, the only responsive student appeared to have used the scaffold to their advantage. This was further proven by the fact that 43% of the causal link edits of this student were effective before scaffolding while 64% was effective in the interval after scaffolding. So, responding to the link annotation strategy was also associated with an improvement in map score. 2 more students responded to the scaffold over the second and third rounds of scaffolding, with net effective annotations across students being 83% after both the second and third time the scaffold was delivered. One student from C2 received this scaffold 6 times, with 100% of their link annotations being effective after the last (sixth) scaffold.

(c) In **C3: Strategic Map Builders group** (n=22), 12 students (54%) received this scaffold, with 2 students receiving it once, 4 receiving it two times each, and 6 receiving it three times each (Table VI.5). 5 of these students had marked a total of 63 links before scaffolding, with 98% link annotations being effective. (Therefore, the C3 was already aware of this strategy and had demonstrated successful use even prior to being scaffolded.) After receiving Scaffold 4 the first time, 7 students (58% of the 12 who were scaffolded) responded by marking their links, with 100% effective annotation of 28 links across respondents. This included 2 students who had not marked any links prior to scaffolding, and therefore started using the strategy after receiving the feedback. The net effectiveness of link annotations over the different rounds of scaffolding for this group was 96%. The effectiveness of link edits did not show any substantial change after scaffolding, which may be attributed to the fact that these students were already aware of the link annotation strategy and showed effective use of this strategy even before being scaffolded.

(d) In **C4: Experimenters group** (n=8), 5 students (62.5%) received this scaffold, with 2 students receiving it once, 2 receiving it two times each, and 1 student receiving it three times. *Before* scaffolding, 3 of these students annotated a total of 50 links with 56% effectiveness. This suggests that this group was aware of the link annotation strategy but was not as successful as C3 in their link marking efforts prior to scaffolding. *After* receiving the scaffold *the first time*, 4 students marked a total of 12 links (83% effective): this included the 3 students who had marked links prior to scaffolding and a fourth student who adopted the behavior after scaffolding and marked 2 links (100% effective). Over the next few rounds of scaffolding, these 4 students marked a total of 27 links (85% effective), with the one student who received this scaffold five times marking all links effectively from the fourth time onward. However, there was no substantial change in map scores, with link edit effectiveness remaining in a 55-59% range before and after scaffolding.

To explore this further, we tracked whether students from C4 revisited the 44% incorrectly annotated links and cleared the ”correct” marks from them after being scaffolded on the correct usage of this strategy by the mentor agent. We found that none of the already marked links were cleared after scaffolding. So while the above numbers show that this group effectively annotated their links after being scaffolded, they still did not revisit the links that were already incorrectly annotated on their map, which possibly led to confusion in debugging their map in future, and the observed lack of improvement in map scores.

Discussion: There was low responsiveness for Scaffold 4 in general, but there were also students who were using the link annotation strategy prior to being scaffolded, especially in C3 and C4, although C3 had a higher success rate with the links they annotated in their map prior to scaffolding. After receiving the correct link annotation feedback from the mentor, the responsive students in all groups showed strategic use of the link annotation feature. However, for students (cf., C4) who already had several incorrectly marked links on their map, the scaffold did not make them revisit those marked links and verify whether they were accurately annotated or not. This point needs to be emphasized in a future iteration of this scaffold to further support the learners who show a high number of incorrectly marked links on their map at the moment when the feedback is delivered to them.

5. **Scaffold 5 *(Quiz→Build Incorrect Link Annotation Feedback)*:** This scaffold was received by 30 students a total of 64 times. 67% of C1 (n=4), 58% of C2 (n=11), 53% of C3 (n=11) and 50% of C4 (n=4) received this scaffold at least once.

The Scaffold 5, similar to Scaffold 4, was delivered to encourage the student to adopt a type of link annotation feature in Betty's Brain (Incorrect Link Annotation or "Mark as 'maybe wrong'") into their Quiz→Build behavior. This feedback was typically delivered to students who has marked some correct links on their map using the "Mark as correct" feature but were still exhibiting ineffective Quiz→Build behaviors where they performed incorrect map edits after checking Betty's quiz results. The scaffold was designed to support the student to engage in more effective map assessment and debugging behaviors by teaching students how to use the "Mark as maybe wrong" feature to annotate links that may be wrong based on their understanding of the quiz results, and how using this feature after quiz assessment could make it easier to spot potentially incorrect links that may need to be reviewed and modified and , if needed, removed from the map.

(a) In **C1: Disengaged group** (n=6), 4 students received Scaffold 5, 1 of them getting the feedback one time and the other 3 three times each. Only one of these students had marked potentially incorrect links on their causal map *before* being scaffolded. *After* scaffolding, student activities showed that 3 of the 4 students (75%) were responsive to the scaffold the first time and marked links on their map as "maybe wrong". The responsiveness decreased as the scaffold count increased, with only one student marking links on their map after receiving this feedback the tenth time. To assess the strategic usage of scaffold, we checked whether the "maybe wrong" annotations were followed by more effective link edits on the map (i.e., edits that increased the map score). Results do not show any significant difference in the intervals before versus after scaffolding, with the average proportion of students' ineffective link edits remaining in a 60-75% range both before and after receiving this scaffold. Instead, one of the dominant actions in the after-Scaffold 5 interval for this group of students was *map moves*, i.e., moving elements of the map from one position to the other. While sometimes students in Betty's Brain may move map elements to organize their map, it may also suggest that the learner is disengaged or bored. Overall, unlike the case of Scaffold 4, there was no evidence that the Scaffold 5 was used by learners in C1 to strategically improve the quality of their maps.

(b) In **C2: Inefficient Information Generators group** (n=19), 11 students received this scaffold. 6 of these students received the scaffold once, 2 students got it twice each, and 3 students got it three times each. Only 4 of these students (36%) responded to the scaffolding by performing the suggested link annotations, and while these students reviewed their quiz results after scaffolding, their subsequent link edits in the *after feedback* interval was 63% ineffective, suggesting that they were unable to use Scaffold 5 effectively. No student responded to this scaffold a second or third time, possibly suggesting that they did not find it to be useful.

(c) In **C3: Strategic Map Builders group** (n=22), only one of the 11 students who received the feedback demonstrated responsiveness by marking potentially incorrect links, but this student did not edit any links in the interval after scaffolding, so was difficult to determine if the feedback served its intended purpose in this case.

(d) A similar result was seen in **C4: Experimenters group** where only 1 of the 4 students who received the feedback responded by making link annotations.

Discussion: Overall, there is no evidence to support that Scaffold 5 was useful to the learner. On the contrary, there was low responsiveness with the response decreasing as the scaffold-count increased, suggesting that many students did not find this feedback useful the first time. In future scaffold design iterations, this scaffold may be removed or the low response rates may need to be further explored by conducting in-the-moment interviews to better understand student perspectives about the feedback. Also, this is the only scaffold for which there was a significant difference in the counts received by C1 and C3, with students in C1 receiving higher counts of Scaffold 5. Given that students in C1 showed activities like map moves after scaffolding, an inability to use this scaffold effectively may have even had a negative effect on the engagement of C1, who have generally reported high disengagement (Table VI.4).

6. **Scaffold 6 *(Quiz→Build Coherence Feedback)*:** This scaffold was received by 24 students a total of 52 times. 3 students (50%) from C1, 8 students (42%) from C2, 9 students (41%) from C3, and 4 students (50%) from C4 received this scaffold at least once. This coherence scaffold was designed to help students perform more coherent Quiz (Solution assessment)→Build (Solution construction). In the feedback, Mr. Davis suggested to the student to teach Betty in a more coherent manner by first teaching her about the concepts and links she did not answer correctly in the last quiz. Mr Davis further suggested that the student start a conversation with him anytime they needed help, especially if they wanted to better understand quiz grades and how they might use this understanding to improve their maps.

There was no significant change in students' Quiz→Build coherence in any group in the intervals after receiving Scaffold 6. Also, none of the students started a conversation with Mr Davis to inquire about their quiz grades. In fact, several students did not let the mentor agent complete the full conversation tree, instead closing the conversation towards the beginning by selecting the option that they did not need help at this time. The lack of responsiveness made it difficult to ascertain whether Scaffold 6 had any impact on learning, but it does make it appear that students generally were not very responsive to this specific coherence feedback. (The impact of Scaffold 6 is explored and discussed further in the case study reported later in Section VI.4).

7. **Scaffold 7 *(Quiz→Read Coherence Feedback)*:** This scaffold was received by 6 students a total of 7 times. 1 student (17%) from C1, 3 students (16%) from C2, and 2 students (9%) from C3 received this scaffold at least once. This scaffold was designed to help students perform more strategic Quiz→Read for knowledge refinement. It was triggered when a student showed incoherent Quiz→Read (e.g., after viewing quiz results, the student skipped through multiple Science Book pages, some of which were unrelated to concepts mentioned in the quiz results they viewed). The hypothesis for scaffold design was that students showing this behavioral pattern were having trouble identifying information related to quiz results from the Science Book. The mentor agent suggested that the student make use of the "search box" feature on the Science Book to identify and review quiz-related information more easily and use their refined knowledge to improve their maps.

   Two students, both from C2, responded to the mentor's suggestion by using the search box after scaffolding to look up information about concepts in their quiz results. The first student searched for information about one concept, then reviewed the science book page that came up in the search results for > 1 *minute* and then added four links from this page on their map (50% effective). The other student searched for information about two concepts, reviewed the results about one of them, compared this information to the recent quiz results (thereby showing that they were trying to refine their knowledge using evidence from both quiz results and science book) and then added two coherent but ineffective links. This again shows that while the scaffold helped C2 identify relevant sections of the Science Book for information acquisition, they still had difficulties updating their knowledge structure using evidence from the science book, leading to an ineffective application of this knowledge in the subsequent causal map edits.

8. **Scaffold 8 *(Negative Affect + Quizzing Proficiency Feedback)*:** Scaffold 8 was received by 26 students a total of 48 times. 3 students from C1 (51%), 7 students from C2 (37%), 14 students from C3 (53%), and 2 students from C4 (25%) received this scaffold.

   This was one of the two scaffolds designed to support students' cognitive-affective regulation process, at moments when their activities and causal modeling performance in Betty's Brain and their affect predictions from the BROMP-based affect detectors suggested (a) a high likelihood that the student was frustrated or bored, (b) the student was currently in the Quiz Results view, and (c) the Betty's Brain learner model suggested a low Quiz→Build proficiency with $< 30\%$ effective Build actions in the last 5 minutes. In the scaffold, Mr Davis gave suggestions that targeted the resolution of their cognitive difficulties that formed the basis of their appraisals of affect. To help the student utilize their quiz results more effectively, Mr Davis suggested in the scaffold that the student start a conversation with him about quiz grades if they wanted to better understand how to interpret certain grades, else they should review the relevant section of the science book while reflecting on the links that needed to be updated for a better grade.

Only one student from C4 was responsive to Mr Davis' suggestion to start a conversation with him about quiz grades. *Before* scaffolding, this student showed frustration as their dominant affect prediction. In the interval *after* scaffolding, they requested a conversation with the mentor agent and asked him how to interpret 'incorrect' quiz grades. Following the mentor's response, the student reviewed a graded question in their quiz, then went back to the map and correctly changed the sign of a link related to the reviewed quiz question. This suggests that the student was able to understand how to interpret an incorrect grade, and apply this interpretation to identify and debug incorrect links associated with the graded quiz answer. The student then took a quiz to verify the effectiveness of the update to their map. The dominant affect prediction from the BROMP-based affect detectors at this stage was confusion, followed closely (in order of the likelihood score from the corresponding binary affect detector model) by frustration.

Prior studies with Betty's Brain have shown interactions between cognitive and affective states (Munshi et al., 2018c), so the resolution of the cognitive obstacles is also likely to support the learner's affect regulation. The predictions from the BROMP-based affect detectors also suggest a transition from a negative affect state that signals an "impasse" or blocked goals (D'Mello and Graesser, 2012) of frustration before the scaffold to a relatively positive state of confusion after the scaffold. However, since the reliability of these affect detectors could not be established very clearly (note the discussion in Section VI.1.2), and in the absence of affect data validated by ground truth labels in our study, we refrain from making claims on the affective implications of Scaffold 8.

9. **Scaffold 9 *(Negative Affect + Reading Proficiency Feedback)*:** This final scaffold was similar to Scaffold 8 but targeted the resolution of negative affect (frustration or boredom) during information acquisition from reading the Science Book. Prior research has shown that low performers in Betty's Brain often get stuck in a reading loop (Munshi et al., 2022b), so this scaffold (triggered when a student was in the Read view and exhibited low Read$\rightarrow$Build proficiency and a likelihood of frustration or boredom) was designed to break the "reading loop", by suggesting the student to take a quiz to check their progress. Scaffold 9 was received by 37 students a total of 68 times. All students had frustration predictions prior to receiving the scaffold.

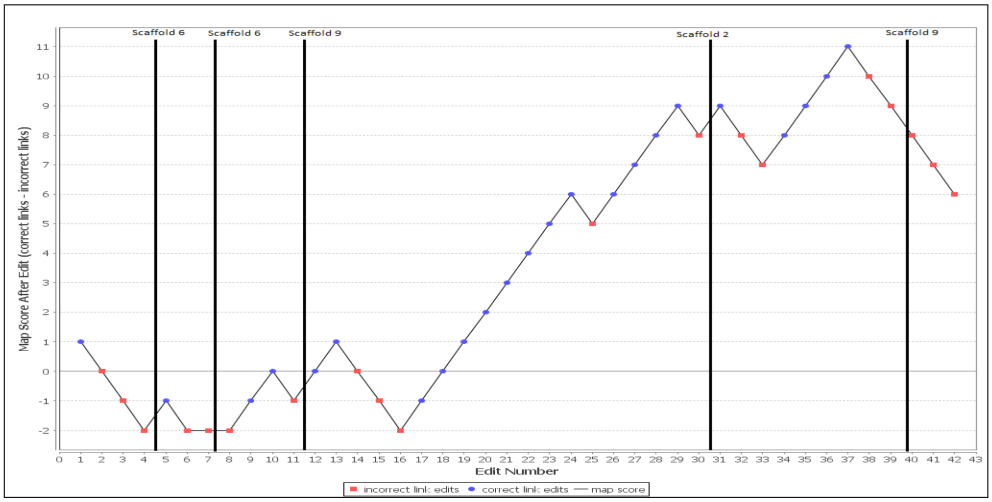
For the 4 students from **C1** who received this scaffold, 2 students (50%) responded to the feedback and took quizzes. The affect prediction scores from the BROMP-based detectors for these students in the *after scaffold* interval showed a mix of confusion (40%) and engagement (37%) as the most dominant emotions. In **C2**, 12 out of 13 students were responsive to the scaffold and took quizzes. This group still showed dominant frustration in the *after* interval, which may be indicative of a difficulty applying quiz results, as observed in the profile of this group (Table VI.4) and also in the discussion on the Quiz→Build scaffolds. In **C3**, 13 out of 16 students responded to the mentor's suggestion, with dominant emotion likelihood scores after scaffolding being 42% confusion, followed by engagement and frustration. In **C4**, all 4 students took quizzes after the mentor's feedback, showing 46% confusion, 27% delight and 19% engagement. Again, as in the case of Scaffold 8, we do not make any claims on the affective implications of Scaffold 9, given the lack of validated affect data from the detectors. (In the case studies in SectionVI.4, we explored the change in student affect after Scaffold 9 in some more detail, using emotion predictions obtained by AffDex models from facial video data.)

VI.4 Case Studies

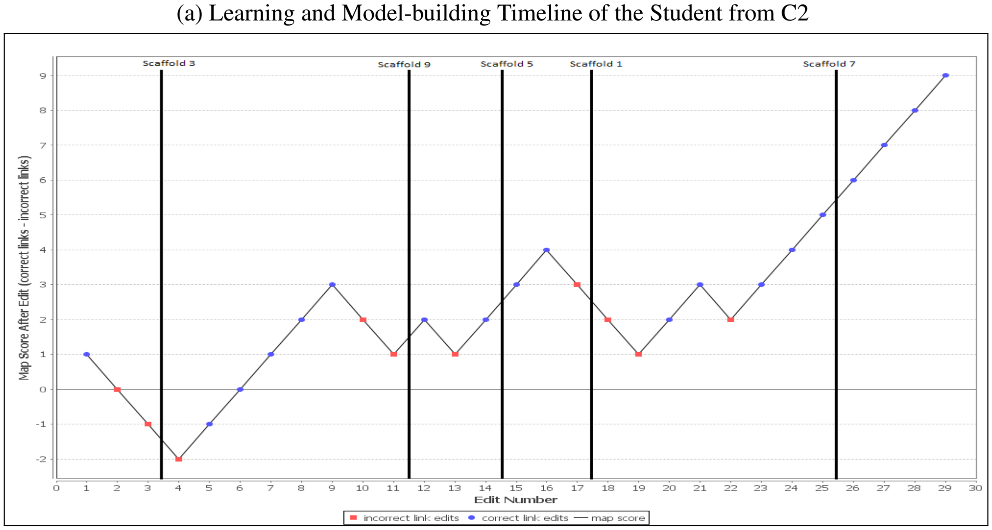
To derive more fine-grained insights into the cognitive and affective implications of adaptive scaffolding, we conducted case studies that tracked the cognitive activities/behaviors, affective states, and model-building performance of two students from our study, as they worked on their causal modeling task in the Betty's Brain environment and received timely adaptive scaffolds from the mentor agent. Since one of the objectives of this analysis was to explore the affective changes associated with adaptive scaffolding, we only considered the subset of students (n=18) for whom facial video data was available (see Section V.1.2). We then filtered out the students who received no scaffolds (n=1), and students whose faces were partially obstructed by face masks (n=13) and prevented the AffDex model from predicting facial emotions. This resulted in 2 students from C2, and 2 students from C3. One representative student from each group was then randomly selected to conduct the case study. The model-building/scaffolding timeline of these two students is presented in Figure VI.6 and discussed below.

Student 1, from the C2 (Inefficient Information Generators) group, received 5 scaffolds during her time in Betty's Brain, as shown in Figure VI.6 (a). This included 1 count of Scaffold 2 (Read→Build Incorrect Link Feedback), 2 counts of Scaffold 6 (Quiz→Build Coherence Feedback), and 2 counts of Scaffold 9 (Negative Affect + Reading Proficiency Feedback).

She received her *first scaffold*, a Scaffold 6 (Quiz→Build Coherence Feedback), after performing a series of ineffective link edits that were not related to wrong answers in the quiz results she viewed prior to the edits. In the feedback delivered to the student, Mr Davis suggested (a) that she first focus on teaching Betty



(a) Learning and Model-building Timeline of the Student from C2



(b) Learning and Model-building Timeline of the Student from C3

Figure VI.6: Case Study of Two Students Who Received Adaptive Scaffolds While Engaged With the Causal Modeling Task in Betty's Brain the links she answered incorrectly in her last quiz, and (b) that she start a conversation with him if she required information on understanding specific quiz grades. After receiving this scaffold, the student continued editing her map without reviewing the quiz results as suggested by the mentor. (This behavior aligns with the general lack of responsiveness to Scaffold 6 observed in Section VI.3.) Frequent head movements of the student prevented affect predictions by the AffDex models at this time. (The BROMP-based affect detectors suggested a shift in affect from confusion before and during scaffolding to frustration in the interval after scaffolding.)

After some time, the student received her *second scaffold*, again a Scaffold 6, due to her continuation of incoherent and ineffective link edits. This time, she did not even let Mr Davis complete the conversation, informing him midway that she did not require help at this time (again following the general trend reported in the previous section). However, this time the student took a quiz shortly after being scaffolded, and then started exhibiting a coherent and effective Quiz→Build behavior. This suggests that Scaffold 6 may be more useful and meaningful to the student *if Mr Davis first asks them to take another quiz*, before going into the Quiz→Build coherence behavior. This would ensure that the recent incoherent and ineffective link edits performed by the student since their last quiz is also reflected in the new quiz results, thereby giving them a better understanding of the current state of their map and the next steps to debug errors in a more coherent manner from the quiz results.

The *third scaffold* received by this student was a Scaffold 9 (Negative Affect + Reading Proficiency Feedback), intended to break the "inefficient reading" loop by suggesting that the student should take a quiz from time to time or even ask him questions if needed. The student was responsive to the mentor's feedback after scaffolding. In addition to using the quiz results to check her progress, she also started a conversation with the mentor agent asking him how to interpret the "? (right so far)" quiz grade. Mr Davis then explained how this grade was connected to missing links that she needed to identify and add to her causal map. But after this conversation, instead of investigating the missing information further, the student started deleting links from her map, leading to the deletion of both correct and incorrect links from the map. The student only returned to a more strategic evidence-driven link editing behavior the next day (which can be seen from edit number 17 onward, in Figure VI.6(a)). Again, continuous affect data was not available from the facial videos in this period due to the student's head movements, which limited a study of the affective implications of Scaffold 9 at this point. However, the findings above still present an opportunity to improve the adaptive scaffolding framework. We surmise that the system should monitor students' use of the feedback after scaffold delivery in a more fine-grained manner, and identify situations like the consecutive deletion of multiple links (without obtaining evidence from the Science Book or Quiz Results). Such a situation should trigger *additional scaffolding*, including *diagnostic* feedback to understand why the learner is not collecting evidence about link correctness before deleting them, followed by *strategic* feedback to teach them a more evidence-driven solution construction approach.

On the second day of working in Betty's Brain, this student exhibited a generally coherent and effective model-building approach, but also added two incorrect links on their map due to which she received her *fourth scaffold* - a Scaffold 2 (Read→Build Incorrect Link Feedback). After being scaffolded, the student followed Mr Davis' suggestion by reviewing the suggested page but was unsuccessful in extracting the information needed to fix the incorrect links from this page, showing the inefficient Read→Build behavior that is characteristic of C2. Now at this point, the student should have received a follow-up (i.e., Level-1) Scaffold 2. Instead, she received Scaffold 9 as her *fifth scaffold* from the mentor. This gives us another scope for improving the feedback triggering conditions, especially the priority assignment algorithm, which accorded a higher priority to Scaffold 9 compared to Scaffold 2: Level 1. In the next design iteration, the design framework should ensure that, if the Level-0 version of a specific scaffold has already been delivered to a student and if the Level-1 trigger for the same scaffold is currently active, then this Level-1 scaffold should receive a higher priority than other different types of feedback, to ensure that the root cause of an obstacle is sufficiently scaffolded before moving on to a different issue.

The fact that Scaffold 9 was not the best feedback at this stage is further validated by the changes observed after scaffolding - a decrease in map score (three incorrect link additions) and an increase in confusion as predicted from the facial indicators (average likelihood scores increasing from 5.5 to 10.4 in this period).

Student 2, from the C3 (Strategic Map Builders) group, received 5 scaffolds (Figure VI.6 (b)), including 1 count each of a Scaffold 1 (Read→Build Shortcut Link Feedback), a Scaffold 3 (Read→Build Coherence Feedback), a Scaffold 5 (Quiz→Build Incorrect Link Annotation Feedback), a Scaffold 7 (Quiz→Read Coherence Feedback), and a Scaffold 9 (Negative Affect + Reading Proficiency Feedback).

After the *first* scaffold, a Scaffold 3 (Read→Build Coherence Feedback), the student responded successfully to the mentor's suggestion by seeking evidence from the Science Book and then using this evidence to build their causal model in a coherent and mostly effective manner. The average emotional valence score, as determined from the AffDex facial emotion detection model, also changed from -0.5 before scaffolding to $+2.6$ after scaffolding, suggesting that receiving the mentor's feedback and adopting the suggested cognitive strategy was also associated with a net improvement in affect.

The student did not respond to the *second scaffold*, a Scaffold 9 (Negative Affect + Reading Proficiency Feedback) by performing any of the suggested cognitive activities. Again, the lack of affect information from the facial affect detectors in this interval prevented a fine-grained study of the affective component of this scaffold on the student's emotions. However, the BROMP-based detectors suggested a state of confusion both before and after scaffolding.

The *third scaffold* received by this student was a Scaffold 5 (Quiz→Build Incorrect Link Annotation Feedback). The student was also not responsive to Scaffold 5, which aligns with the general response to this scaffold for C3 that was observed in Section VI.3. Since she did not respond to this feedback by performing suggested link annotation activities, no claims can be made regarding the impact of this scaffold on the consequent effective link edits observed in Figure VI.6(b). Combined with the findings from the previous section, the Scaffold 5 may need to be removed from the adaptive scaffolding framework or redesigned completely to encourage responsiveness.

Next, the student received her *fourth scaffold*, a Scaffold 1 (Read→Build Shortcut Link Feedback). In the interval following the receipt of Scaffold 1, she showed a clear pattern of effective and coherent Read→Build behavior that fixed the shortcut link and led to the addition of other correct links. The highest average affective valence score of 4.6 for this student was also observed in the period after receiving Scaffold 1, providing circumstantial evidence that the adoption of the Read→Build strategy after scaffolding was also associated with an effective emotion regulation process.

The *fifth scaffold* received by this student was a Scaffold 7 (Quiz→Read Coherence Feedback). She was already in an effective model-building phase by the time of receiving this scaffold, and the map score kept increasing after the scaffold was delivered. However, no discernible impact of this scaffold was observed, and the student did not use the "search box" behavior prescribed by the mentor agent but was still successful in her model-building and debugging activities. The results in Section VI.3 showed that only students from C2 responded to Scaffold 7, so this feedback may need to be targeted only to specific groups of students who show clear evidence of being inefficient information generators.

VI.5 General Discussions

Section VI.2 discussed our inferences on the scaffold statistics, i.e., the number of adaptive conversational scaffolds received by students with different behavioral profiles, as reported in Table VI.5. This section discussed our plans about redesigning or improving certain trigger conditions in future research, to account for between-group differences and to ensure even more meaningful and task-relevant adaptive scaffolding. Section VI.3 reported on the responsiveness and strategic usage of scaffolds by tracking student and group behaviors before and after scaffolding. The case studies in Section VI.4 provide further clues on how to refine certain scaffolds to benefit students in future design iterations.

The general similarity observed in inter-scaffold intervals between the four groups, as observed in Section VI.1.2, is an improvement on previous scaffold design iterations (Munshi et al., 2022b,a), and may be attributed to more uniform triggering parameters and the inclusion of inter-scaffold interval as a separate parameter within the current design framework, which ensured that students did not receive their next scaffold unless a minimum set amount of time had passed since their previous scaffolding.

The results from the case studies (Section VI.4) also provide opportunities for improving scaffold triggering factors like the priority assignment algorithm, for instance, with respect to Scaffold 2: Level-1 and Scaffold 9 (see discussion in Section VI.4). The temporal analysis in Section VI.3, beyond helping us understand the between-group differences in the impact of scaffolding, also provided insights into temporal changes in such impact, for instance, the case of C2 after receiving Scaffold 5 multiple times. Such results present further scope to introduce *fading* for certain types of scaffolds based on student responses, to maintain engagement with the task and future responsiveness to scaffolding. We also found that the Read→Build scaffolds (Scaffolds 1, 2 and 3) and the Quiz→Build Correct Link Annotation Feedback (Scaffold 5) were generally followed by high response rates, with certain groups exhibiting more effective cognitive strategies and metacognitive monitoring behaviors in the *after* intervals, eventually linking the strategic use of scaffolds to improvements in their causal models.

We now summarize some of our future plans on revising the current adaptive scaffolding framework based on the findings from this chapter. **Scaffold 1** was used by C1 as a corrective hint. C2 showed low response rates to this feedback, instead checking quiz results and being unsuccessful at using Quiz→Build to fix their shortcut links. C3 used the scaffold strategically and effectively, with the case study even showing very high affective valence alongside the strategy use for this group. But interestingly, C4 did not receive this scaffold despite the presence of shortcut links at certain points on their maps. In future design iterations, we may use out understanding of a student's current behavioral profile, as discussed above, to provide more feedback that is tailored to the characteristics of such behavioral groups. For instance, if a student is identified to be an inefficient information generator (C2) and does not respond to Scaffold 1, they may receive additional scaffolding that guides them through the Read→DebugShortcutLink process. If a student is detected as an experimenter (C4), then their priority assignment algorithm may be modified dynamically (as discussed in Section VI.2) if it is observed that they have shortcut links on their map but are not receiving any Shortcut Link Feedback due to the higher priority assigned to Scaffold 3. **Scaffold 2** was more effective for the disengaged group, who used the feedback strategically and also showed a decrease in disengagement after scaffolding. So, the trigger condition for Scaffold 2 may be placed at a higher priority order for a student who is identified as disengaged in a future study. Tinkering behavior after scaffolding may also be monitored, so that students can receive additional feedback (e.g., we noted that C4 students changed the sign of incorrect links instead of deleting them in the current results). **Scaffold 3** was useful for both C3 and C4 groups to develop a more coherent Read→Build strategy. We discussed how this feedback may also have accounted for the high potential usage durations in the experimenter group. In Section VI.2, we discussed how the trigger condition for **Scaffold 4** may be improved to include an additional component which tracks incorrect link annotations already present on the student's map prior to scaffolding. This is based on the finding for C4, who used the scaffold to correctly perform new link annotations but did not review previous incorrectly annotated links. Otherwise, this scaffold was successful in helping C2 develop *quiz→link annotation* as a strategy to debug and improve their maps. For **Scaffold 5**, which was received a large number of times by C1, we noted higher disengagement after successive rounds of scaffolding. Therefore, this scaffold may be reduced to a one-time feedback, or faded with time if disengagement is diagnosed, or may even be redesigned to provide additional feedback on how to combine the two types of link annotations (Scaffolds 5 and 6) to isolate potentially incorrect links. **Scaffold 6** generally showed low response rates but the case study also suggested that this feedback may be made more meaningful to the student if Mr Davis first asks them to take another quiz, before going into the Quiz→Build coherence behavior, thereby ensuring that Build actions performed since their last quiz is also reflected in the new quiz results, further giving them a better understanding of the current state of their map and form a plan to debug errors in a more coherent manner from the quiz results. The trigger condition for **Scaffold 7** may be assigned a higher priority when students show inefficient information generator characteristics, with a follow-up Read→Build scaffold after the Quiz→Read that then helps the student to extract the correct causal relations from the text they are reading. For **Scaffold 8** and **Scaffold 9**, an important component of our future plans is to perform a better evaluation of the affective components of this feedback by further validating the detector predictions using ground truth labels, as discussed above. Scaffold 8 had low response rates while Scaffold 9 had high response rates in the current study. The case study suggested that the system should monitor students' use of Scaffold 9 in a more fine-grained manner after the feedback is delivered, to identify the type of situation observed in the case study (viz., the consecutive deletion of multiple links without obtaining evidence from the Science Book or Quiz Results). As discussed in Section VI.4, such situations should trigger additional *diagnostic* followed by *strategic* feedback to teach students how to develop a more evidence-driven solution construction process.

CHAPTER VII

Conclusions and Future Work

This dissertation contributes to the field in two primary directions: (1) Design & development, and (2) Research. We discuss the specific contributions along these two directions in some more detail below.

VII.1 Design and Development Contributions

The dissertation has presented the **design and development of an adaptive scaffolding framework** to model and support cognitive-metacognitive and cognitive-affective components of K-12 learners' SRL processes in an open-ended science learning environment.

The **conceptual framework for scaffold design** was informed by a theoretical review of SRL and scaffolding literature (Chapter II) and by findings from an empirical design-based research process conducted in the context of the Betty's Brain learning environment (Chapter III). The **literature review** involved a critical analysis of prominent SRL models (Zimmerman; Boekaerts; Winne and Hadwin; Pintrich; Efklides; Hadwin et al.) to understand that SRL is a *dynamic learning process* with interacting 'CAMM' components. Emerging evidence from empirical research on SRL in advanced learning technologies (Azevedo et al., 2015, 2017; Bannert et al., 2017) further established that the regulation of CAMM processes was integral to students' SRL process in OELEs. To understand how to *detect CAMM processes* from observable information in OELEs, we studied cognitive-metacognitive (Winne, 1995) and cognitive-affective (Pintrich, 2000; D'Mello and Graesser, 2012) relations, and the procedures applied by researchers for the detection of cognitive-metacognitive (Kinnebrew et al., 2013b; Biswas et al., 2016; Munshi et al., 2018a) and affective states (Jiang et al., 2018; McDuff et al., 2016) in learning environments. We further reviewed the features (Elsom-Cook, 1993; Puntambekar and Hubscher, 2005) and objectives (Self, 1988) of successful scaffold design, to infer that our adaptive scaffolds needed to be *flexible*, be based on an *online diagnosis* of students' learning process (cognitive-metacognitive strategy use and affect appraisals), and provide *tailored assistance* to achieve the *strategic* (and in some cases, diagnostic, and elaborative) objectives of scaffolding. A set of **design-based research (DBR) studies** with Betty's Brain (Section III.2), conducted over a five-year period, helped to further understand students' cognitive-metacognitive behaviors and strategy use and their cognitive-affective processes in the context of this learning environment. This allowed us to situate the design of the scaffolding environment in the context of students' learning tasks, activities, and behavior in Betty's Brain. The DBR process involved cycles of *Problem Analysis* →*Design* →*Evaluation* phases. Reflecting on the results from each phase of this process (Munshi et al., 2018c,b; Rajendran et al., 2018b; Jiang et al., 2018;

Munshi and Biswas, 2019; Munshi et al., 2020, 2022b,a) led to step-by-step improvements in the design and implementation approach of our adaptive scaffolding framework in Betty's Brain.

The **design framework** presented in Chapter IV of this dissertation was informed by the findings from this DBR process, and included a set of *triggering conditions* for online adaptive scaffolding of students as they worked in the Betty's Brain environment. The adaptive scaffolding was designed as *conversational trees* that allowed for delivery of the feedback in a step-by-step manner to make it easier for students to assimilate and apply the feedback provided. Our triggering conditions were linked to a **strategy detection** process, which tracked key binary relations from students' observable activity sequences modeled as cognitive processes, that were derived from a modified version of the Kinnebrew et al. (2017) task model as a reference framework. Measures of coherence adopted from (Segedy et al., 2015) provided the metrics for identifying ineffective or sub-optimal strategy use. in mode-building behaviors in Betty's Brain, and identified a *pattern* of ineffective cognitive-metacognitive strategy use. Sensor-free **affect detectors** (Jiang et al., 2018) provided the basis for measuring affect likelihood values, and were used to identify the moments during learning when a student's' affect likelihood shifted to a state of dominant frustration or boredom. We interpreted these signals to indicate *obstacles* or *disengagement* (D'Mello and Graesser, 2012) and we developed appropriate strategy scaffolds to help students recover from these negative valence states (Baker et al., 2021). To deliver meaningful strategy scaffolds at these moments, the *cognitive attributions* of affect appraisal were also inferred by checking students' associated strategy use and their effectiveness.

Since students' activity patterns could activate multiple scaffold triggering conditions at the same point of time, each type of trigger condition included in our design framework was also assigned a priority value by a **prioritization algorithm** (Appendix C). The priority order between trigger conditions was determined primarily by checking the state of the student's causal model, with higher priority assigned to Read→Build patterns when the student had a sparse map (so they could receive feedback that encouraged information acquisition and model construction) and higher priority assigned to Quiz→Build patterns when the student had a denser map (so they received more adaptive feedback for model debugging purposes). Each trigger condition was also assigned a **time to live** value, to ensure that students only received scaffolds relevant to their current task and recent activities.

The learner scaffolding module in Betty's Brain periodically checked for trigger conditions, pulled the highest priority trigger condition, and used the **conversation tree** structure (Segedy et al., 2015) to deliver conversational adaptive feedback through the mentor agent in Betty's Brain. The conversations included diagnostic, elaborative and strategic components, and provided actionable recommendations that students could use to develop cognitive-metacognitive strategies for model construction or debugging tasks in Betty's Brain. An **inter-feedback interval** parameter in the learner scaffolding module checked that sufficient time had elapsed since the student received their last scaffold, thus ensuring that their learning process was not interrupted by too frequent interventions. Some of the scaffolds were also offered at different **levels**, with more contextualized (level-1) feedback provided to learners who were unable to use the generalized (level-0) feedback in an effective manner. Nine types of adaptive scaffolds IV.5 were included in the latest design and implementation of our adaptive scaffolding framework in Betty's Brain.

VII.2 Research Contributions

Chapter V of this dissertation presented an approach for **scaffold evaluation**, that combined a study of the relationship between students' pre-post learning gains, their map building performance, their use of effective strategies, and their ability to assimilate and use the strategy scaffolds provided by the system. The results of our analyses are presented in Chapter VI, and show the nuanced differences in how groups of students with differing learning behaviors/outcomes *received*, *responded to* and *used* the adaptive feedback received from the mentor agent, for their subsequent learning tasks in Betty's Brain. The results helped us derive insights on the strengths and weaknesses of specific aspects of our scaffold design framework, which provide opportunities for future research to further improve adaptive scaffolding in OELEs like Betty's Brain.

Our study involved middle school students who worked on a model-building task in Betty's Brain and received feedback when they had difficulties or they performed sub-optimally. We explored students' learning outcomes, affect and behaviors from this study, and applied a cluster analysis algorithm that discovered four groups of students with differences in behavior and learning outcomes and behaviors - *(1) disengaged learners, (2) inefficient information generators, (3) strategic map builders, and (4) trial-and-error experimenters and tinkerers*. The findings from this exploratory analysis enabled us to formulate more targeted research questions (Section V.2.3) and data analysis (Section V.2.4) for scaffold evaluation. In Sections VI.2, VI.3 and VI.4, we reported the findings from this analysis, with specific focus on (1) the **count**, i.e., number of adaptive scaffolds of each type received by students in each of the four groups; (2) students' **responsiveness** to the scaffolds they received, as observed from comparing their subsequent activities with respect to the actionable suggestions given by the mentor agent in the feedback; and (3) their **strategic use** of these scaffolds, as evidenced by subsequent changes in their learning behaviors and outcomes in Betty's Brain. We focused our discussions on **between-group differences** in scaffold counts, student response and usage, to (1) **infer** how the behavioral profiles of these different groups may have been mediated by the receipt of in-time conversational scaffolding, and (2) **decide** how this improved understanding (of learner behaviors/outcomes, and the impact of adaptive scaffolds on their behaviors/outcomes) may be used to improve adaptive scaffold design in such OELEs in future. Our conclusions in this regard have been presented in Section VI.5, with more specific cases (where the findings point to opportunities for future research) discussed across Sections VI.2

to VI.4. Our scaffold evaluation analysis also revealed the limitations of our current approach (viz., the lack of validated and continuously available affect data) and prompted a discussion (Section VI.5) on the future research directions to overcome these limitations.

To summarize, the findings from the scaffold evaluation analysis presented in this dissertation show *which scaffolds were useful*, and *for which groups* of students, and *how the strengths and weaknesses of our current design can inform the design of an even more improved adaptive scaffolding framework in future*, to provide learners with differing behavioral profiles with more nuanced, contextualized and meaningful guidance.

In addition to the results from assessing the current scaffold design, **the DBR process (Section III.2) culminating in this dissertation has also resulted in several research contributions to the field** of SRL and adaptive scaffolding in OELEs. This multi-year research process included multi-modal data collection and analysis (using information from questionnaires, log traces, face videos, screen recordings, eye-trackers, ML models, audio interviews, etc.) Some of the specific contributions from this set of research studies were along the following directions: understanding cognitive-affective relations in OELEs (Munshi et al., 2018c,b), modeling learners' temporal behaviors and performance (Rajendran et al., 2018b), designing a learner model based scaffolding framework for SRL (Munshi and Biswas, 2019), modeling the relationships between different types of affect states (Munshi et al., 2020), designing a learner-model-based scaffolding framework for SRL (Munshi and Biswas, 2019), analyzing adaptive scaffolds to better understand their impact on self-regulated learning (Munshi et al., 2022b,a). A more complete set of co-authored publications resulting from the research towards this dissertation is reported in Appendix A, and provide evidence on the impact of this work in extending the state-of-the-art.

VII.3 Future Work

Future work in this field can use the findings from this dissertation research to build learner models that use a more aggregated understanding of a student's behaviors (and the evolution of behaviors across time in the OELE) to assign them into behavioral profiles (viz., a disengaged student, or an inefficient information generator), which would then be used to offer a more guided support, to re-engage the student in their learning task, or to train them on the use of monitoring and self-reflection processes for information generation and successful application in their causal models. The current set of scaffolds included in our design framework will also be refined based on our findings, for instance, by removing or redesigning scaffolds which incurred low responsiveness (like Scaffold 5) or by refining the conversations to better communicate the purpose of scaffolding (e.g., strategic versus corrective hint in the case of Scaffold 1 or 2). Our more specific plans in relation to the refinement of different aspects of the adaptive scaffolding framework based on the findings from the latest evaluation study is discussed in more detail in Section VI.5.

More generally, the adaptive scaffolding framework developed for this dissertation can also be applied to scaffold student learning in other open-ended STEM learning environments. In particular, this approach may be especially useful to understand and support model-building and debugging behaviors in other *learning-by-modeling* OELEs like C2STEM (Hutchins et al., 2020), which use problem-based scientific model construction to promote K-12 STEM learning.

Additionally, we note that our scaffold evaluation study was limited by a lack of validated and consistently available affect data from the detector models, as discussed in Section VI.1.2. Also, since the evaluation study was conducted in a real classroom versus a more controlled laboratory setting, factors like frequent head movements or face masks led to data loss and an unavailability of continuous temporal face data for the AffDex models. So, while the observed improvements in cognitive behaviors and performance after scaffolding may also suggest more positive affect appraisals, given the type of cognitive-affective relationships found from prior Betty's Brain studies (Munshi et al., 2018c)), the unavailability of accurate ground truth labels of student affect did not allow us to derive evidence on such cognitive-affective states and their relation to the received scaffolds. Future work in this area may use affect detector mapping procedures such as Munshi et al. (2020) in conjunction with human coded affect data for validation, to gain a more complete picture of the affect regulation process of different groups of students who receive adaptive scaffolds in Betty's Brain. Moreover, keeping in mind the data collection and other practical constraints in classroom studies, multimodal data sources such as eye-tracking and in-the-moment interviews or in-system self-reports may also be used to fill in the gaps in data availability from sources like log data and detector models.

Appendix A

List of Relevant Publications (in reverse chronological order)

1. Munshi, A., Biswas, G., Davalos, E., Logan, O., Narasimham, G., and Rushdy, M. (2022a). **Adaptive scaffolding to support strategic learning in an OELE**. In *Proceedings of the 30th International Conference on Computers in Education (ICCE 2022)*, Malaysia

   Abstract: This paper discusses the effectiveness of adaptive conversational scaffolds implemented in the Betty's Brain open-ended learning-by-teaching environment. Our scaffolding framework utilizes detectors based on student's activity patterns to infer their suboptimal strategic cognitive-metacognitive behaviors and a conversation tree structure to facilitate the delivery of in-the-moment contextualized feedback. We conducted an experimental lab study to collect data (activity logs, screen recordings and interviews) while students worked on the Betty's Brain system. Our initial findings suggest that some scaffolds helped learners develop strategic behaviors that helped them overcome their individual learning difficulties. We also discuss improvements to some of the scaffolds in our framework to better support learners as they build causal models to teach Betty.

2. Munshi, A., Biswas, G., Rushdy, M., Baker, R., Ocumpaugh, J., and Paquette, L. (2022b). **Analyzing Adaptive Scaffolds to Help Students Develop Self-Regulated Learning Behaviors**. *Journal of Computer Assisted Learning*

   Abstract: Providing adaptive scaffolds to help learners develop self-regulated learning (SRL) processes has been an important goal of intelligent learning environments. In this paper, we develop a systematic framework for adaptive scaffolding in Betty's Brain, an open-ended learning-by-teaching environment that helps middle school students learn science by constructing a causal model to teach a virtual agent, generically named Betty. Given the open-ended nature of the environment, novice learners often face difficulties in their learning and teaching tasks. We detect key cognitive/meta-cognitive inflection points, i.e., instances where students' behaviors and performance change as they work on learning and teaching tasks. At such inflection points, Mr. Davis (a mentor agent), or Betty (the teachable agent) provide conversational feedback on SRL strategies to help students become more productive learners. We analyze data collected from a classroom study with 98 middle school students to study the impact of the scaffolds on students' SRL behaviors and learning performance. We discuss how our findings will support the next iteration of our adaptive scaffolding framework to help students develop their SRL behaviors when working in OELEs.

3. Zhang, Y., Paquette, L., Baker, R., Ocumpaugh, J., Bosch, N., Biswas, G., and Munshi, A. (2021b). **Can strategic behavior facilitate confusion resolution?** *Journal of Learning Analytics*

   Abstract: Confusion may benefit learning when it is resolved or partially resolved. Metacognitive strategies (MS) may help learners to resolve confusion when it occurs during learning and problem solving. This study examined the relationship between confusion and MS that students invoked in Betty's Brain, a computer-based learning-by modeling environment where elementary and middle school students learn science by building causal models. Participants were sixth graders. Emotion data were collected from real-time observations by trained researchers. MS and task performance information were determined by analyzing the action logs. Pre- and post-tests were used to assess learning gain. The results revealed that the use of MS was a function of the students' state of confusion. However, confusion resolution was not related to MS behavior, and MS did not moderate the effect of confusion on students' task performance in Betty's Brain and learning gain.

4. Baker, R., Nasiar, N., Ocumpaugh, J., Hutt, S., Andres, A., Slater, S., Schofield, M., Moore, A., Paquette, L., Munshi, A., and Biswas, G. (2021). **Affect-Targeted Interviews for Understanding Student Frustration**. In *Proceedings of the 22nd International Conference on Artificial Intelligence in Education*. Springer, LNCS

   Abstract: Frustration is a natural part of learning in AIED systems but remains relatively poorly understood. In particular, it remains unclear how students' perceptions about the learning activity drive their experience of frustration and their subsequent choices during learning. In this paper, we adopt a mixed-methods approach, using automated detectors of affect to signal classroom researchers to interview a specific student at a specific time. We hand-code the interviews using grounded theory, then distill particularly common associations between interview codes and affective patterns. We find common patterns involving student perceptions of difficulty, system helpfulness, and strategic behavior, and study them in greater depth. We find, for instance, that the experience of difficulty produces shifts from engaged concentration to frustration that lead students to adopt a variety of problem-solving strategies. We conclude with thoughts on both how this can influence the future design of AIED systems, and the broader potential uses of data mining-driven interviews in AIED research and development.

5. Hutt, S., Ocumpaugh, J., Andres, A., Munshi, A., Bosch, N., Baker, R., Zhang, Y., Paquette, L., Slater, S., and Biswas, G. (2021). **Who's Stopping You? – Using Microanalysis to Explore the Impact of Science Anxiety on Self-Regulated Learning Operations**. In *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society (CogSci)*

Abstract: Research shows that anxiety can disrupt learning processes, but few studies have examined anxiety's relationships to online learning behaviors. This study considers the interplay between students' anxiety about science and behavior within an online system designed to support self-regulated science inquiry. Using the searching, monitoring, assessing, rehearsing, and translating (SMART) classification schema for self-regulated learning (SRL), we leverage microanalysis of self-regulated behaviors to better understand how science anxiety inhibits (or supports) different learning operations. Specifically, we show that while science anxiety is positively associated with searching behaviors, it is negatively associated with monitoring behaviors, suggesting that anxious students may avoid evaluation, opting instead to compensate with information-seeking. These findings help us to better understand SRL processes and may also help us support anxious students in developing SRL strategies.

6. Munshi, A., Mishra, S., Zhang, N., Paquette, L., Ocumpaugh, J., Baker, R., and Biswas, G. (2020). **Modeling the Relationships Between Basic and Achievement Emotions in Computer-Based Learning Environments**. In Bittencourt, I. I., Cukurova, M., Muldner, K., Luckin, R., and Millán, E., editors, *Artificial Intelligence in Education*, pages 411–422. Springer

   Abstract: Commercial facial affect detection software is typically trained on large databases and achieves high accuracy in detecting basic emotions, but their use in educational settings is unclear. The goal of this research is to determine how basic emotions relate to the achievement emotion states that are more relevant in academic settings. Such relations, if accurate and consistent, may be leveraged to make more effective use of the commercial affect-detection software. For this study, we collected affect data over four days from a classroom study with 65 students using Betty's Brain. Basic emotions obtained from commercial software were aligned to achievement emotions obtained using sensor-free models. Interpretable classifiers enabled the study of relationships between the two types of emotions. Our findings show that certain basic emotions can help infer complex achievement emotions such as confusion, frustration and engaged concentration. This suggests the possibility of using commercial software as a less context-sensitive and more development-friendly alternative to the affect detector models currently used in learning environments.

7. Zhang, Y., Paquette, L., Baker, R., Ocumpaugh, J., Bosch, N., Munshi, A., and Biswas, G. (2020). **The relationship between confusion and metacognitive strategies in Betty's Brain**. In *Proceedings of the 10th International Learning Analytics and Knowledge (LAK) Conference*, Frankfurt

Abstract: Confusion has been shown to be prevalent during complex learning and has mixed effects on learning. Whether confusion facilitates or hampers learning may depend on whether it is resolved or not. Confusion resolution, behind which is the resolution of cognitive disequilibrium, requires learners to possess some skills, but it is unclear what these skills are. One possibility may be metacognitive strategies (MS), strategies for regulating cognition. This study examined the relationship between confusion and actions related to MS in Betty's Brain, a computer-based learning environment. The results revealed that MS behavior differed during and outside confusion. However, confusion resolution was not related to MS behavior, and MS did not moderate the effect of confusion on learning.

8. Sharma, K., Mishra, S., Papamitsiou, Zacharoula, M. A. D. B., Biswas, G., and Giannakos, M. (2020). **Towards obtaining facial proxies for gaze behaviour in TEL**. In *Proceedings of the International Conference of the Learning Sciences (ICLS)*, volume 5, pages 2621–2622, Nashville

   Abstract: Current multimodal studies have a common limitation of not being able to scale up the implications since the apparatus used is not scalable. In this paper, we propose a simple method to find measurements from scalable data modes such as facial data and examine the measures in richer and more granular data modes like eye-tracking that they correspond most closely to. In other words, we find pervasive proxies to the measurements that have been reported to be obtrusive. We exemplify this approach using eye-tracking and facial data from two different studies.

9. Munshi, A. and Biswas, G. (2019). **Personalization in OELEs: Developing A Data-Driven Framework to Model and Scaffold SRL Processes**. In *Artificial Intelligence in Education*, pages 354–358, Chicago. Springer, LNCS

   Abstract: This research focuses on developing a data-driven framework for modeling and scaffolding learners' self-regulated learning (SRL) processes in open-ended learning environments (OELE). The aim of this work is to offer a personalized and productive learning experience by adapting scaffolds to help learners develop self-regulation skills and strategies. This research applies mining techniques on data collected from multiple channels to track learners' cognitive, affective, metacognitive and motivational (CAMM) processes as they work in Betty's Brain, a computer-based OELE. The CAMM information is used to derive online models of learners' SRL processes. These learner models inform the design of personalized scaffolds that help students develop the required SRL process and become more proficient learners. The significance of this research lies in developing and using data-driven learner SRL models to personalize and contextualize the scaffolds provided to learners within the OELE.

10. Andres, J. M. A. L., Ocumpaugh, J., Baker, R. S., Slater, S., Paquette, L., Jiang, Y., Karumbaiah, S., Bosch, N., Munshi, A., Moore, A., and Biswas, G. (2019). **Affect sequences and learning in Betty's Brain**. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pages 383–390, Arizona

Abstract: Education research has explored the role of students' affective states in learning, but some evidence suggests that existing models may not fully capture the meaning or frequency of how students transition between different states. In this study we examine the patterns of educationally-relevant affective states within the context of Betty's Brain, an open-ended, computer-based learning system used to teach complex scientific processes. We examine three types of affective transitions based on similarity with the theorized D'Mello and Graesser model, transition between two affective states, and the sustained instances of certain states. We correlate of the frequency of these patterns with learning outcomes and our findings suggest that boredom is a powerful indicator of students' knowledge, but not necessarily indicative of learning. We discuss our findings within the context of both research and theory on affect dynamics and the implications for pedagogical and system design.

11. Mishra, S., Munshi, A., Rushdy, M., and Biswas, G. (2019). **LASAT: Learning Activity Sequence Analysis Tool**. In *TEEL Workshop at the 9th International Learning Analytics and Knowledge (LAK) Conference*, Arizona

Abstract: Learning Activity Sequence Analysis Tool (LASAT) is a collection of sequence analysis algorithms developed at Institute for Software Integrated Systems, Vanderbilt University, with the purpose of extracting and interpreting students' learning behaviors extracted as frequent patterns (sequence of activities) from their activity traces logged in computer-based learning environments. LASAT includes several algorithms for analyzing temporal sequence data – such as, sequential pattern mining (SPM), differential sequence mining (DSM) and Hidden Markov Model-based learner modeling. LASAT also includes tools for pre-processing and organizing log data for analysis. In this paper, we present the LASAT toolkit with an aim of making these algorithms accessible to the wider community of researchers and practitioners. We review cases from the learning analytics literature, which have employed LASAT algorithms to demonstrate the use of the tool in supporting evidence-based pedagogical decision making, specifically in the context of learner modeling in computer-based learning environments (CBLE). This paper demonstrates the applicability of LASAT for a range of applications that span from studying learners' cognitive and strategic processes to affect transitions that together form the basis for understanding self-regulated learning processes.

12. Munshi, A., Rajendran, R., Ocumpaugh, J., Biswas, G., Baker, R. S., and Paquette, L. (2018c). **Modeling Learners' Cognitive and Affective States to Scaffold SRL in Open-Ended Learning Environments**. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, pages 131–138, Singapore. ACM

Abstract: The relationship between learners' cognitive and affective states has become a topic of increased interest, especially because it is an important component of self-regulated learning (SRL) processes. This paper studies sixth grade students' SRL processes as they work in Betty's Brain, an agent-based open-ended learning environment (OELE). In this environment, students learn science topics by building causal models. Our analyses combine observational data on student affect to log files of students' interactions within the OELE. Preliminary analyses show that two relatively infrequent affective states, boredom and delight, show especially marked differences among high and low performing students. Further analysis shows that many of these differences occur after receiving feedback from the virtual agents in the Betty's Brain environment. We discuss the implications of these differences and how they can be used to construct adaptive personalized scaffolds.

13. Munshi, A., Rajendran, R., Moore, A., Ocumpaugh, J., and Biswas, G. (2018b). **Studying the Interactions between Components of Self-Regulated Learning in Open Ended Learning Environments**. In *Proceedings of the 13th International Conference of the Learning Sciences (ICLS)*, pages 1691–1692, London, England

Abstract: This paper investigates the interactions between learners' cognitive strategies and affective states; both important components of self-regulated learning (SRL) processes that influence student learning. We study cognitive-affective relationships in high versus low performing students as they worked on a model building task to teach their agent in Betty's Brain, an open-ended science learning environment. Our initial results allow for some interesting discussions , but they also emphasize the need for fine grained affective data to match up against cognitive states to determine how they influence performance or vice versa.

14. Rajendran, R., Munshi, A., Emara, M., and Biswas, G. (2018b). **A Temporal Model of Learner Behaviors in OELEs using Process Mining**. In *Proceedings of the 26th International Conference on Computers in Education (ICCE)*, Philippines

Abstract: Open-ended learning environments (OELEs) present learners with complex problems and a set of tools for solving these problems. Developing logging mechanisms that capture learners' interactions with the system provide a wealth of trace data that can be employed for studying relations

between their behaviors and performance. Such analyses provide a framework for making the OELE intelligent in that it can adapt its feedback to meet the needs of individual learners. In our previous research, we have developed learner modeling schemes that are based on sequential pattern mining (SPM) and Hidden Markov models (HMMs) to represent and track the temporal sequence of learners' interactions with the OELE. We briefly discuss the pros and cons of these models, and then propose a process modeling approach to capture the temporal nature of learners' behaviors. We apply the process modeling method to data collected from students working with the Betty' Brain OELE, where students learn about scientific processes by building causal models.

Appendix B

Supplementary Figures and Tables

Table B.1: From Munshi et al. (2022b): List of the inflection point triggers, and scaffolds provided at triggers, in the first scaffold design iteration leading to this dissertation (From DBR Cycle 1, evaluated in the Feb 2019 study)

(a) When the trigger condition is related to unproductive/ineffective activities

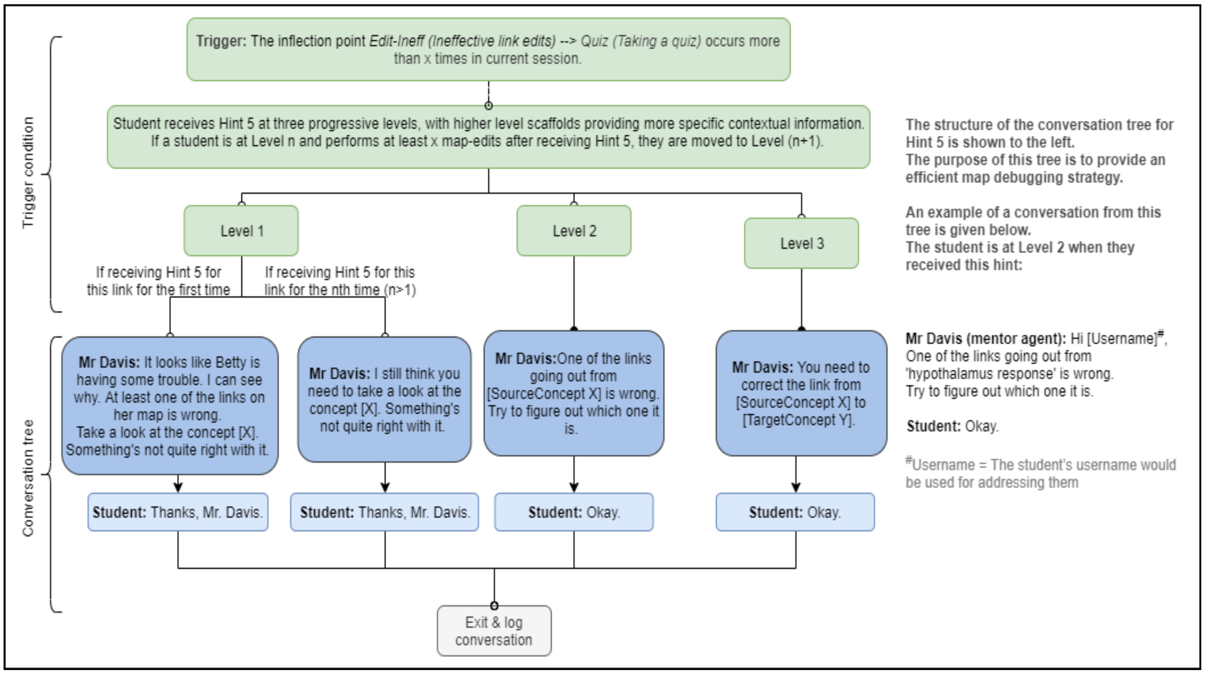
<table>
<tr><th colspan="2">Inflection Point Trigger</th><th colspan="2">Provided Scaffold</th></tr>
<tr><th colspan="2">Task/Activity Context</th><th>Scaffold Type</th><th>Content Overview & Excerpts</th></tr>
<tr><td colspan="2">Information acquisition (Read-Long) → Ineffective model construction (Edit-Ineff)</td><td>Strategic hint:
Assess by Quiz
Hint2</td><td>Betty suggests taking a quiz, as a model assessment strategy, to help debug errors in the map.
"Hi, I think you just added a causal link on your map after looking at the science book. ... Do you think I am ready for a quiz now?"</td></tr>
<tr><td rowspan="4">Ineffective model construction (Edit-Ineff) → Model assessment (Quiz)</td><td>Case 1:
AND The student has not marked the recently edited incorrect links.</td><td>Strategic hint:
Mark Wrong
Hint3</td><td>Mr Davis suggests marking the possibly incorrect links on map as "could be wrong", as an efficient map organization strategy.
"From the quiz results, looks like Betty may have some incorrect links on her map. You can mark those links as 'could be wrong'. Do you want to know more? ..."</td></tr>
<tr><td>Case 2:
WHERE The Edit-Ineff was a shortcut link addition (e.g.: an A→C link instead of an A→B→C link)</td><td>Strategic hint:
Shortcut Link
Hint4</td><td>Mr Davis explains how to identify & correct shortcut links.
"From the quiz, it seems you may have an incorrect shortcut link on your map. Do you want to know more about shortcut links? ..."</td></tr>
<tr><td>Case 3</td><td>Strategic hint:
Debug from Map
Hint5</td><td>Mr Davis provides map debugging strategies to fix model errors identified from quizzes, progressing from high-level feedback to more specific corrective hints.
"One of the links going out of 'hypothalamus response' is wrong. Try to find out which one it is."</td></tr>
<tr><td>Case 4</td><td>Encouragement:
Reassure
Enc3</td><td>Betty provides an encouragement message to ensure that the student is not demotivated after seeing their errors in the quiz results.
"... Sometimes I find all of this a little tricky. But with you to teach me, I'm sure we can do it."</td></tr>
<tr><td colspan="2">Information acquisition (Read-Long) → Model assessment (Quiz)</td><td>Strategic hint:
Debug from Read
Hint6</td><td>Mr Davis provides progressive hints to support reading the pages relevant to map errors, as an efficient map debugging strategy.
"You are missing a link that comes out of 'heat loss'. Try reading up on Page 'Response 1: Skin Contraction' and see if you can find the link."</td></tr>
</table>

(b) When the trigger condition is related to productive/effective activities

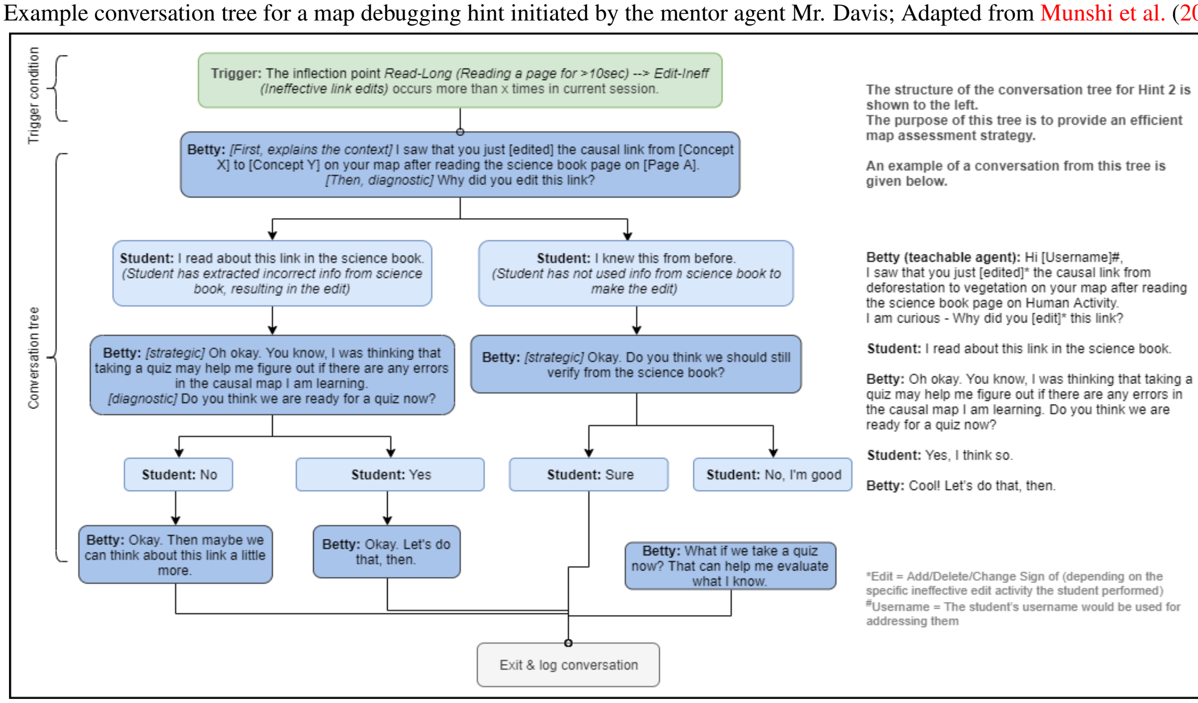
Inflection Point Trigger		**Provided Scaffold**	
Task/Activity Context		***Scaffold Type***	***Content Overview & Excerpts***
Information acquisition (Read-Long) → Efficient model construction (Edit-Eff)		Encouragement: Praise & Quiz **Enc2**	Mr Davis praises the student for teaching her well, and suggests taking a quiz to find evidence for their teaching progress. *"Looks like you're doing a good job teaching correct causal links to Betty. ... Make sure you check her progress ... by asking her to take a quiz"*
Efficient model construction (Edit-Eff) → Model assessment (Quiz)	Case 1	Strategic hint: Mark Correct **Hint1**	Mr Davis suggests marking the possibly correct links on the map as "correct", as an efficient map organization strategy. *"If Betty got an answer graded correct, remember to mark those links as 'correct' in the map. This can help you keep track of what you have taught her correctly so far. Do you know how to ..."*
	Case 2	Encouragement: Praise **Enc1**	Betty praises the student for doing a good job of teaching her an efficient causal model. *"Wow! I think I have some correct links on the map. This is fun! Thanks, A."*

Table B.2: List of scaffolds provided in the second design iteration, evaluated in the Sept 2021 pilot study

Scaffold Type	**Excerpt from Conversation Tree**
Read to Build Correctness: Shortcut Link	Level 0: Mr Davis: "... Your map may have shortcut link(s) from this part of your science book. Do you want me to tell you more about shortcut links? ..."
	Level 1: Mr Davis: "Looks like you still have a shortcut link coming out of [Concept A]. .. Review pages [X, Y] to find out more about the missing link(s) to add to your map."
Read to Build Correctness: Incorrect Link	Level 0: Mr Davis: "You may have incorrect links on your map from this part. ... There are three ways a causal link may be incorrect: if (a) you are linking two unrelated concepts; (b) the sign of the link (increase/decrease) is wrong; (c) the direction is wrong. ... Review Pages [X, Y] to figure out the incorrect links."
	Level 1: Mr Davis: [Review of Level 0] + "...You have an incorrect link coming out of the [Concept A]. Review the relevant pages to figure it out."
Read to Build Coherence	Level 0: Betty: "... teaching me concepts and links that are not related to what we just read." Mr Davis: "... A good learning strategy is to work on one topic at a time ... read a page and add all correct links from it to your map before moving on to a different one. This will help teach ..."
	Level 1: Mr Davis: "... you just added a link [A→B] from Page [Y] but you were reading Page [X]. Try to add all links from Page [X] first ..."
Quiz to Build Correctness: Correct Link Annotation	Mr Davis: "... There are correct links from your recent quiz that you have not marked on your map yet. Would you like me to teach you how to mark the 'correct' links?... Select a quiz question graded correct (green checkmark). ..."
Quiz to Build Correctness: Incorrect Link Annotation	Mr Davis: "... Select a quiz question graded incorrect (red X). ... This means that at least one of the links ... is wrong. You can mark these links as 'maybe wrong' on the map..."
Quiz to Build Coherence	Mr Davis: "... After Betty takes a quiz, ... first teach her the concepts she did not answer correctly. This way, she can get these right the next time she takes a quiz. ..."
Quiz to Read Coherence	Mr Davis: "... I can help you find information from the science book to correct Betty's wrong answers. ... you can type about a concept in the search bar of the page ..."



(a) Example conversation tree for a map debugging hint initiated by the mentor agent Mr. Davis; Adapted from Munshi et al. (2022b)



(b) Example conversation tree for a map assessment hint, initiated by the teachable agent Betty; Adapted from Munshi et al. (2022b)

Figure B.1: Two example trees that illustrate the conversation tree representation in Betty's Brain, developed for the first iteration of adaptive scaffolding.
In each figure: *(left) the triggering condition and tree structure; (right) example of a student-agent conversation from the tree*

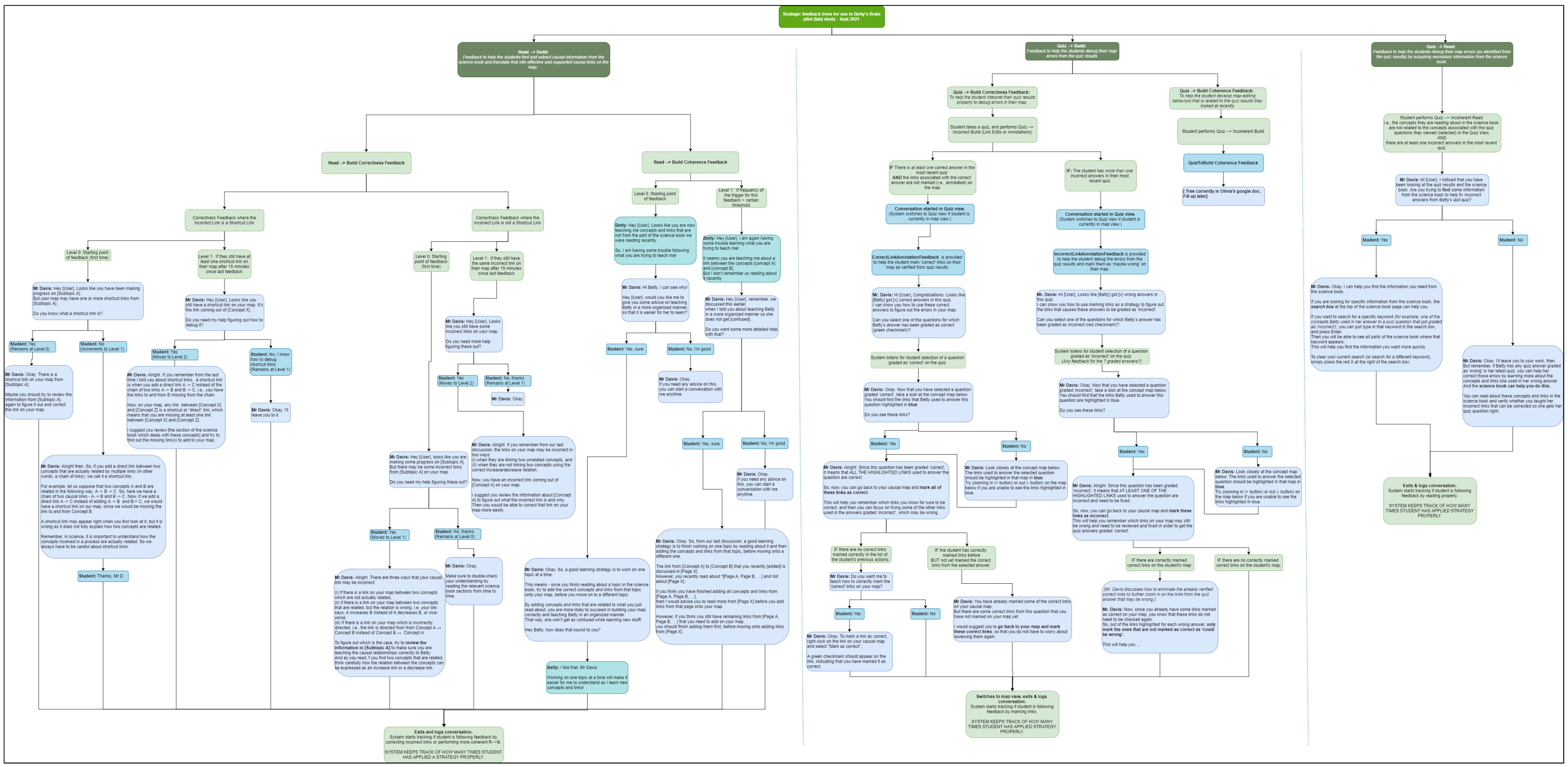


Figure B.2: Conversation trees used in the pilot lab study for evaluating the adaptivity framework

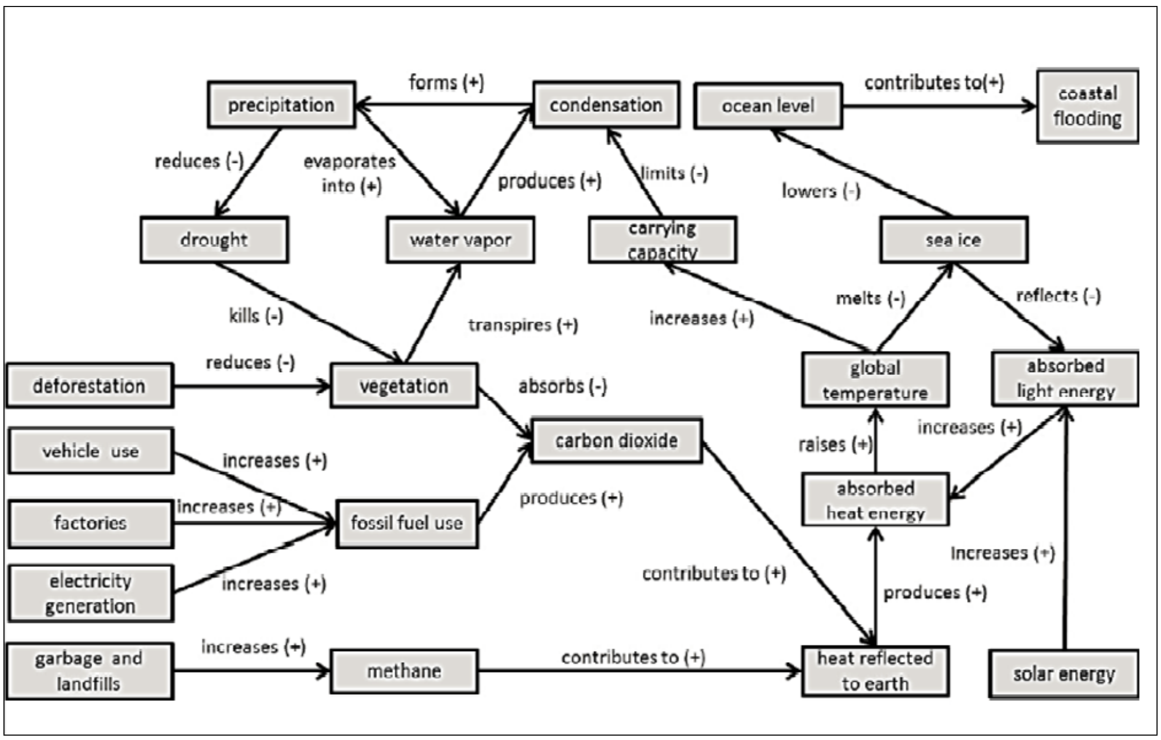


Figure B.3: The expert map that learners attempt to build in the climate change unit of Betty's Brain

Appendix C

Other Supplementary Material

The Priority Assignment Algorithm (Scaffold Design Component)

```
#############################################
# FRAMEWORK FOR TRIGGERING FEEDBACK
#############################################
1. Within Learner Model Controller
   a. Pattern (Trigger) Detection
   b. Pattern Priority Assignment For Insertion Into Priority Queue
2. Maintaining references to the priority queue
   a. Maintain minimum_time_between_feedbacks
   b. IF current_time - last_feedback_time >= minimum_time_between_feedbacks:
       Request feedback (pattern_msg) at top of queue
3. Deliver feedback
   a. Deliver conversation_tree (pattern_msg, type, level)
   b. Log <pattern_msg, type, level, conversation_tree>
   c. Update last_feedback_time

################################################
# Pattern (Trigger) Detection (In Controller) ##
################################################
current_pattern_msg = Event sequences triggered by the student
pattern_threshold = x          #threshold can be same for all patterns or different for different patterns
IF current_pattern_msg.getCount > pattern_threshold:        # If pattern detection exceeds threshold
   pattern_priority = getPriority(current_pattern_msg):
   current_pattern_msg.setPriority(pattern_priority)            # priority is assigned to the pattern
   priorityQueue.append(current_pattern_msg)          # the pattern is inserted into a priority queue


####################################################################################
#
# Pattern Priority Assignment For Insertion Into Priority Queue (Also in Controller) #
####################################################################################

mapOfReceivedFeedback = {R→B-Shortcut: 0,   # map {type, level} of feedbacks the student has received
                  Q→B-Coh: 1,    .... }

def getPriority(pattern_msg):
   IF mapOfReceivedFeedback.get(pattern_msg.getType) == 0  # student has received level 0 of same
feedback
      return 0;  # currentPatternType is assigned the highest priority = 0 (supersedes others in queue)
   ELSE
      return priority_lookup (correct_links, incorrect_links)        # assigns priority based on lookup table


def priority_lookup(correct_links, incorrect_links):
   standard = 25                              # standard (of the no. of links) to compare against
```

```
IF correct_links <= standard:
   IF incorrect_links <= standard:       # LOW correct_links & LOW incorrect_links (LOW NO. OF BUILDS)
                 Affect + R-->B: 1
                 R-->B Coherence: 2
                 R-->B-Shortcut: 3
                 R-->B-Incorrect: 4
                 Affect + Q-->B: 5
                 Q-->B-Correct Link Annotation: 6
                 Q-->B-Incorrect Link Annotation: 7
                 Q-->B-Coherence: 8
                 Q-->R-Coherence: 9
   ELSE:                          # LOW correct_links & HIGH incorrect_links (UNSUCCESFUL BUILDS)
                 Q→ B-Correct Link Annotation: 1
                 Q-->B-Incorrect Link Annotation: 2
                 Q-->B-Coherence 3
                 Q-->R-Coherence: 4
                 Afect + R-->B: 5
                 R-->B Coherence: 6
                 R-->B-Shortcut: 7
                 R-->B-Incorrect: 8
                 Q-->R-Coherence: 9
ELSE:
   IF  incorrect_links <= standard:  #HIGH correct_links & LOW incorrect_links (SUCCESSFUL BUILDS)
                 Affect Q-->B: 1
                 Q-->B-Incorrect Link Annotation: 2
                 Q-->B-Correct Link Annotation: 3
                 Q-->B-Coherence: 4
                 Afect + R-->B: 5
                 R-->B Coherence: 6
                 R-->B-Incorrect: 7
                 R-->B-Shortcut: 8
                 Q→ R-Coherence: 9
   ELSE:                          # HIGH correct_links & HIGH incorrect_links (HIGH NUMBER OF BUILDS)
                 Affect Q-->B: 1
                 Q-->B-Correct Link Annotation: 2
                 Q-->B-Incorrect Link Annotation: 3
                 Q-->B-Coherence: 4
                 Q→ R-Coherence: 5
                 Afect + R-->B: 6
                 R-->B Coherence: 7
                 R-->B-Shortcut: 8
                 R-->B-Incorrect: 9
```

Pre and Post Assessments Used in Scaffold Evaluation Study

Betty’s Brain PRE-SURVEY

Name ____

Date _______________

VU DoD – Mar 2022

Motivation

Instructions

This survey will help us to know your views on learning about climate change with Betty's Brain. Please **circle the number corresponding to your response** for each statement.

Serial no.	***Statement***	***Response to statement*** **Strongly Disagree (1)**	**Disagree (2)**	**Neutral (3)**	**Agree (4)**	**Strongly Agree (5)**
1.	I think that learning about climate change is important because I can use it in my daily life.	1	2	3	4	5
2.	In science, I think that it is important to learn to solve problems.	1	2	3	4	5
3.	It is important to be curious when learning science.	1	2	3	4	5
4.	Whether the content is difficult or easy, I am sure that I can understand it.	1	2	3	4	5

Serial no.	*Statement*	*Response to statement* Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)
5.	I am not confident about understanding difficult science concepts.	1	2	3	4	5
6.	I am sure that I can do well on science tests.	1	2	3	4	5
7.	I am willing to participate in this course about climate change because it is challenging.	1	2	3	4	5
8.	I am willing to participate in this course about climate change because the content is exciting.	1	2	3	4	5

Climate Change

Instructions

This survey will help us find out what you know about climate change. Please answer every question as best as you can.

Multiple choice questions: Choose the Best Answer

1. **What is the greenhouse effect?**
   a. The atmosphere of the earth traps some radiated heat energy and reflects it back to the earth. This makes the earth warmer.
   b. The atmosphere of the earth is reflective and keeps sunlight away from the earth's surface. This light reflection keeps the earth from getting too hot.
   c. The atmosphere acts like a magnifying glass. This makes the light stronger and makes the earth hotter.
   d. The atmosphere traps pollution from cars and factories. Over time, the air will become more polluted and the earth will get warmer.

2. **Light from the sun comes to the earth and its energy is absorbed by the atmosphere. What is the relation between this *absorbed light energy* and the heat energy absorbed by the earth?**
   a. Absorbed light energy increases the amount of absorbed heat energy.
   b. Absorbed light energy decreases the amount of absorbed heat energy.
   c. Absorbed light energy does not change the amount of absorbed heat energy.
   d. Absorbed light energy is not related to absorbed heat energy.

3. **Clouds are made up of water vapor. How does condensation of water vapor in clouds affect precipitation?**
   a. Condensation and precipitation are not related.
   b. Condensation decreases precipitation.
   c. Condensation increases precipitation.
   d. An increase in condensation may increase or decrease precipitation.

4. **What is the main greenhouse gas created in landfills?**
   a. Carbon dioxide
   b. Methane
   c. Oxygen
   d. All of the above

5. **How does vegetation affect the amount of carbon dioxide in the atmosphere?**
    a. Vegetation produces carbon dioxide through the process of photosynthesis, which increases the amount of carbon dioxide in the atmosphere.
    b. Vegetation releases water vapor through the process of photosynthesis. The vapor bonds with carbon dioxide, which reduces the amount of carbon dioxide in the atmosphere.
    c. Vegetation absorbs carbon dioxide as part of the process of photosynthesis, which reduces the amount of carbon dioxide in the atmosphere.
    d. Vegetation produces oxygen because of photosynthesis, but it does not affect the amount of carbon dioxide in the atmosphere.

6. **How does an increase in carbon dioxide affect sea ice?**
    a. Carbon dioxide is absorbed by sea ice. This increases the melting point of sea ice and more sea ice melts.
    b. Carbon dioxide reflects heat radiated from the earth back to earth. This radiation keeps the earth cool, and so it decreases the amount of sea ice that melts.
    c. Carbon dioxide forms a shield around the earth. This protects solar energy from heating the ice caps, so less sea ice melts.
    d. Carbon dioxide reflects heat radiated from the earth back to earth. This radiation increases the earth's temperature, and more sea ice melts.

7. **Which statement <u>best</u> explains how driving more cars affects global temperature?**
    a. Car engines run hot. This increases the surrounding temperatures. The more cars we drive, the higher the global temperature.
    b. Cars burn fossil fuels, and this produces carbon dioxide. The carbon dioxide prevents solar energy from entering the earth's atmosphere. This reduces global temperature.
    c. Cars burn fossil fuels, and this produces carbon dioxide. The carbon dioxide prevents radiated heat energy from leaving the earth's atmosphere. This increases global temperature.
    d. Car engines run hot, and this produces carbon dioxide. But the carbon dioxide cools quickly, and it does not affect the global temperature.

Short Answer Questions: Answer the following questions clearly, and use a <u>complete step-by-step approach as shown in the example below to answer your question</u>. Start each step on a separate line. NOTE: You may not need as many steps as are available on the answer lines.

EXAMPLE: *Because of interdependence in an ecosystem, a change in the population of one species can have effects through the ecosystem. Explain, step-by-step, how wolves affect the amount of grass in the ecosystem?*

Step 1: *Wolves eat deer, so more wolves would reduce the number of deer.*

Step 2: *Deer eat grass, so fewer deer would increase the amount of grass.*

Step 3: *<EXAMPLE PROBLEM – DO NOT FILL IN>*

Step 4: *<EXAMPLE PROBLEM – DO NOT FILL IN>*

Therefore, *when the wolf population increases the amount of grass will increase.*

8. We know that deforestation (cutting down a large number of trees) increases the earth's absorbed heat energy.
   Explain, step-by-step, how deforestation increases the earth's absorbed heat energy.

Step 1: *Deforestation reduces the number of trees on the earth, so more deforestation would decrease vegetation.*

Step 2: *When vegetation decreases, …*

Step 3:

Step 4:

Therefore, *deforestation causes an increase in the earth's absorbed heat energy.*

9. Explain, step-by-step, how increases in global temperature would affect coastal flooding.

Step 1: ____________________

____

Step 2: ____________________

____

Step 3: ____________________

____

Step 4: ____________________

____

Therefore, ____________________

____

10. Explain, step-by-step, how carbon dioxide affects global temperature.

Step 1: ______________________________

____

Step 2: ______________________________

____

Step 3: ______________________________

____

Step 4: ______________________________

____

Therefore, ______________________________

____

Betty's Brain POST-SURVEY

Name ____

Date ______________

VU DoD - March 2022

Motivation

Instructions

This survey will help us to know your views on learning about climate change with Betty's Brain. Please **circle the number corresponding to your response** for each statement.

Serial no.	***Statement***	***Response to statement*** **Strongly Disagree (1)**	**Disagree (2)**	**Neutral (3)**	**Agree (4)**	**Strongly Agree (5)**
1.	I think that learning about climate change is important because it stimulates my thinking.	1	2	3	4	5
2.	In science, I think it is important to participate in inquiry activities.	1	2	3	4	5
3.	When science activities are too difficult, I give up or only do the easy parts.	1	2	3	4	5
4.	During science activities, I prefer to ask other people for the answer rather than think for myself.	1	2	3	4	5

Serial no.	*Statement*	*Response to statement* Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)
5.	When I find the science content difficult, I do not try to learn it.	1	2	3	4	5
6.	No matter how much effort I put in, I cannot learn science.	1	2	3	4	5
7.	I enjoyed teaching my “student” about climate change in Betty’s Brain.	1	2	3	4	5
8.	It was challenging to work on Betty’s Brain.	1	2	3	4	5
9.	The hints given by Mr. Davis were helpful in teaching my “student”.	1	2	3	4	5

Climate Change

Instructions

This survey will help us find out what you know about climate change. Please answer every question as best as you can.

Multiple choice questions: Choose the Best Answer

1. **What is the greenhouse effect?**
    a. The atmosphere of the earth traps some radiated heat energy and reflects it back to the earth. This makes the earth warmer.
    b. The atmosphere of the earth is reflective and keeps sunlight away from the earth's surface. This light reflection keeps the earth from getting too hot.
    c. The atmosphere acts like a magnifying glass. This makes the light stronger and makes the earth hotter.
    d. The atmosphere traps pollution from cars and factories. Over time, the air will become more polluted and the earth will get warmer.

2. **Light from the sun comes to the earth and its energy is absorbed by the atmosphere. What is the relation between this *absorbed light energy* and heat energy absorbed by the earth?**
    a. Absorbed light energy increases the amount of absorbed heat energy.
    b. Absorbed light energy decreases the amount of absorbed heat energy.
    c. Absorbed light energy does not change the amount of absorbed heat energy.
    d. Absorbed light energy is not related to absorbed heat energy.

3. **Clouds are made up of water vapor. How does condensation of water vapor in clouds affect precipitation?**
    a. Condensation and precipitation are not related.
    b. Condensation decreases precipitation.
    c. Condensation increases precipitation.
    d. An increase in condensation may increase or decrease precipitation.

4. **What is the main greenhouse gas created in landfills?**
    a. Carbon dioxide
    b. Methane
    c. Oxygen
    d. All of the above

5. **How does vegetation affect the amount of carbon dioxide in the atmosphere?**
    a. Vegetation produces carbon dioxide through the process of photosynthesis, which increases the amount of carbon dioxide in the atmosphere.
    b. Vegetation releases water vapor through the process of photosynthesis. The vapor bonds with carbon dioxide, which reduces the amount of carbon dioxide in the atmosphere.
    c. Vegetation absorbs carbon dioxide as part of the process of photosynthesis, which reduces the amount of carbon dioxide in the atmosphere.
    d. Vegetation produces oxygen because of photosynthesis, but it does not affect the amount of carbon dioxide in the atmosphere.

6. **How does an increase in carbon dioxide affect sea ice?**
    a. Carbon dioxide is absorbed by sea ice. This increases the melting point of sea ice and more sea ice melts.
    b. Carbon dioxide reflects heat radiated from the earth back to earth. This radiation keeps the earth cool, and so it decreases the amount of sea ice that melts.
    c. Carbon dioxide forms a shield around the earth. This protects solar energy from heating the ice caps, so less sea ice melts.
    d. Carbon dioxide reflects heat radiated from the earth back to earth. This radiation increases the earth's temperature, and more sea ice melts.

7. **Which statement <u>best</u> explains how driving more cars affects global temperature?**
    a. Car engines run hot. This increases the surrounding temperatures. The more cars we drive, the higher the global temperature.
    b. Cars burn fossil fuels, and this produces carbon dioxide. The carbon dioxide prevents solar energy from entering the earth's atmosphere. This reduces global temperature.
    c. Cars burn fossil fuels, and this produces carbon dioxide. The carbon dioxide prevents radiated heat energy from leaving the earth's atmosphere. This increases global temperature.
    d. Car engines run hot, and this produces carbon dioxide. But the carbon dioxide cools quickly, and it does not affect the global temperature.

Short Answer Questions: Answer the following questions clearly, and use a <u>complete step-by-step approach as shown in the example below to answer your question</u>. Start each step on a separate line. NOTE: You may not need as many steps as are available on the answer lines.

EXAMPLE: *Because of interdependence in an ecosystem, a change in the population of one species can have effects through the ecosystem. Explain, step-by-step, how wolves affect the amount of grass in the ecosystem?*

Step 1: *Wolves eat deer, so more wolves would reduce the number of deer.*

Step 2: *Deer eat grass, so fewer deer would increase the amount of grass.*

Step 3: *<EXAMPLE PROBLEM – DO NOT FILL IN>*

Step 4: *<EXAMPLE PROBLEM – DO NOT FILL IN>*

Therefore, *when the wolf population increases the amount of grass will increase.*

8. We know that deforestation (cutting down a large number of trees) increases the earth's absorbed heat energy.
   Explain, step-by-step, how deforestation increases the earth's absorbed heat energy.

Step 1: *Deforestation reduces the number of trees on the earth, so more deforestation would decrease vegetation.*

Step 2: *When vegetation decreases, …*

Step 3:

Step 4:

Therefore, *deforestation causes an increase in the earth's absorbed heat energy.*

9. Explain, step-by-step, how increases in global temperature would affect coastal flooding.

Step 1: ____

Step 2: ____

Step 3: ____

Step 4: ____

Therefore, ____

10. Explain, step-by-step, how carbon dioxide affects global temperature.

Step 1:

Step 2:

Step 3:

Step 4:

Therefore,

Feedback

Instructions

Through this survey, we want to get some feedback from you about your experience with the Betty's Brain climate change unit and your feelings about science learning. This will help us improve Betty's Brain to better support your needs.

Please **circle your choice** for each question/statement.

Serial no.	***Statement***	***Choice***				
1.	How hard was the Betty's Brain unit on climate change?	Very Difficult	Difficult	Neither	Easy	Very Easy
2.	How familiar were you with the science concepts in the Climate Change unit?	Very Difficult	Difficult	Neither	Easy	Very Easy
3.	Mr. Davis's feedback during the Climate Change unit was helpful.	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
4.	It wouldn't bother me at all to take more science classes.	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
5.	I have usually been at ease during science tests.	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
6.	I have usually been at ease in science courses.	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
7.	I usually don't worry about my ability to solve science problems.	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

8.	I almost never get uptight while taking science tests.	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
9.	I get really uptight during science tests.	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
10.	I get a sinking feeling when I think of trying hard science problems.	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
11.	My mind goes blank and I am unable to think clearly when working on science.	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
12.	Science makes me feel uncomfortable and nervous.	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
13.	Science makes me feel uneasy and confused.	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

BIBLIOGRAPHY

Adams, E. (2010). Fundamentals of game design. berkeley. *CA: New Riders*.

Aleven, V. and Koedinger, K. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, 26:147– 179.

Aleven, V., Stahl, E., Schworm, S., Fischer, F., and Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of educational research*, 73(3):277–320.

Andres, J. M. A. L., Ocumpaugh, J., Baker, R. S., Slater, S., Paquette, L., Jiang, Y., Karumbaiah, S., Bosch, N., Munshi, A., Moore, A., and Biswas, G. (2019). **Affect sequences and learning in Betty's Brain**. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pages 383–390, Arizona.

Azevedo, R., Behnagh, R., Duffy, M., Harley, J., and Trevors, G. (2012). Metacognition and self regulated learning in student centered leaning environments. *Theoretical foundations of student centered learning*, pages 171–197.

Azevedo, R., Cromley, J., Winters, F., Moos, D., and Greene, J. (2005). Adaptive human scaffolding facilitates adolescents' self-regulated learning with hypermedia. *Instructional science*, 33(5-6):381– 412.

Azevedo, R., Johnson, A., Chauncey, A., and Burkett, C. (2010). Self-regulated learning with metatutor: Advancing the science of learning with metacognitive tools. In *New science of learning*, pages 225–247. Springer.

Azevedo, R., Taub, M., and Mudrick, N. (2015). Technologies supporting self-regulated learning. *The SAGE Encyclopedia of educational technology*, page 731– 734.

Azevedo, R., Taub, M., and Mudrick, N. V. (2017). Understanding and reasoning about real-time cognitive, affective, and metacognitive processes to foster self-regulation with advanced learning technologies.

Baker, R., Corbett, A., and Koedinger, K. (2004). Detecting student misuse of intelligent tutoring systems. In *Proceedings of the 7th International Conference on Intelligent Tutoring Systems*, pages 531–540.

Baker, R., Nasiar, N., Ocumpaugh, J., Hutt, S., Andres, A., Slater, S., Schofield, M., Moore, A., Paquette, L., Munshi, A., and Biswas, G. (2021). **Affect-Targeted Interviews for Understanding Student Frustration**. In *Proceedings of the 22nd International Conference on Artificial Intelligence in Education*. Springer, LNCS.

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., and Koedinger, K. (2008). Why students engage in gaming the system behavior in interactive learning environments. *Journal of Interactive Learning Research*, 19 (2):185–224.

Bannert, M., Molenaar, I., Azevedo, R., Järvelä, S., and Gašević, D. (2017). Relevance of learning analytics to measure and support students' learning in adaptive educational technologies. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, pages 568–569. ACM.

Basu, S., Biswas, G., and Kinnebrew, J. S. (2017). Learner modeling for adaptive scaffolding in a computational thinking-based science learning environment. *User Modeling and User-Adapted Interaction*, 27(1):5–53.

Beaudoin, L. and Winne, P. (2009). Nstudy: An internet tool to support learning, collaboration and researching learning strategies. *Canadian e-Learning conference*.

Belland, B. R., Glazewski, K. D., and Richardson, J. C. (2011). Problem-based learning and argumentation: Testing a scaffolding framework to support middle school students' creation of evidence-based arguments. *Instructional Science*, 39(5):667–694.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., and at Vanderbilt, T. T. A. G. (2005). Learning by teaching: a new agent paradigm for educational software. *Applied Artificial Intelligence*, 19:363–392.

Biswas, G., Segedy, J. R., and Bunchongchit, K. (2016). From design to implementation to practice a learning by teaching system: Betty's brain. *International Journal of Artificial Intelligence in Education*, 26(1):350–364.

Boekaerts, M. (1996). Self-regulated learning at the junction of cognition and motivation. *Eur. Psychol. 2*, page 100–112.

Boekaerts, M. (2011). Emotions, emotion regulation, and self-regulation of learning. *Handbook of Self-Regulation of Learning and Performance eds Zimmerman B. J., Schunk D. H. (New York, NY: Routledge*, page 408–425.

Boekaerts, M. and Corno, L. (2005). Self regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology*, 54(2):199–231.

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

Borkowski, J. G., Chan, L. K. S., and Muthukrishna, N. (2000). A process oriented model of metacognition: links between motivation and executive functioning. *Issues in the Measurement of Metacognition eds Schraw G., Impara J.*

Bransford, J. D., Brown, A. L., Cocking, R. R., et al. (2000). *How people learn*, volume 11. Washington, DC: National academy press.

Bransford, J. D. and Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *A. Iran-Nejad P. D. Pearson (Eds.). Review of Research in Education*, 24:61–101.

Brusilovski, P., Kobsa, A., and Nejdl, W. (2007). *The adaptive web: methods and strategies of web personalization*, volume 4321. Springer Science & Business Media.

Calinski, T. and Harabasz, J. (1974). A dendrite method for cluster analysis: Communications in statistics. *Theory and Methods*, 3:1–27.

Chrysafiadi, K. and Virvou, M. (2013). Student modeling approaches: A literature review for the last decade. *Expert Systems with Applications*, 40:4715–4729.

Dabbagh, N. and Kitsantas, A. (2012). Personal learning environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. *The Internet and higher education*, 15(1):3–8.

Devolder, A., van Braak, J., and Tondeur, J. (2012). Supporting self-regulated learning in computer-based learning environments: systematic review of effects of scaffolding in the domain of science education. *Journal of Computer Assisted Learning*, 28(6):557–573.

Dinsmore, D. L., Alexander, P. A., and Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational psychology review*, 20(4):391–409.

D'Mello, S. and Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2):145–157.

Donker, A., de Boer, H., Kostons, D., van Ewijk, C., and van der Werf, M. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review*, 11:1–26.

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: the masrl model. *Educ. Psychol.*, 46:6–25.

Ekman, P. and Friesen, W. (1978). Facial action coding system: A technique for the measurement of facial movement. *Consulting Psychologists Press, Palo Alto*.

Elsom-Cook, M. (1993). Student modelling in intelligent tutoring systems. *Artif. Intell. Rev.*, 7(3–4):227–240.

Entwistle, N. and Ramsden, P. (2015). *Understanding student learning (routledge revivals)*. Routledge.

Greene, J. A. and Azevedo, R. (2007). Emotions, emotion regulation, and self-regulation of learning. *A theoretical review of Winne and Hadwin's model of self-regulated learning: new perspectives and directions. Rev. Educ. Res. 77.*, page 334–372.

Hadwin, A. F., Järvelä, S., and Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. *Handbook of Self-Regulation of Learning and Performance eds Zimmerman B. J., Schunk D. H.*, page 65–84.

Hannafin, M., Hill, J., Land, S., and Lee, E. (2014). Student-centered, open learning environments: research, theory, and practice. *Handbook of Research on Educational Communications and Technology*, pages 641–651.

Hickey, D. T. (1997). Motivation and contemporary socio-constructivist instructional perspectives. *Educational Psychologist*, 32(3):175–193.

Hutchins, N. M., Biswas, G., Maróti, M., Lédeczi, Á., Grover, S., Wolf, R., Blair, K. P., Chin, D., Conlin, L., Basu, S., et al. (2020). C2stem: A system for synergistic learning of physics and computational thinking. *Journal of Science Education and Technology*, 29(1):83–100.

Hutt, S., Ocumpaugh, J., Andres, A., Munshi, A., Bosch, N., Baker, R., Zhang, Y., Paquette, L., Slater, S., and Biswas, G. (2021). **Who's Stopping You? – Using Microanalysis to Explore the Impact of Science Anxiety on Self-Regulated Learning Operations**. In *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society (CogSci)*.

Jiang, Y., Bosch, N., Baker, R. S., Paquette, L., Ocumpaugh, J., Andres, J. M. A. L., Moore, A. L., and Biswas, G. (2018). Expert feature-engineering vs. deep neural networks: Which is better for sensor-free affect detection? In *International Conference on Artificial Intelligence in Education*, pages 198–211. Springer.

Karumbaiah, S., ANDRES, J., BOTELHO, A., Baker, R., and OCUMPAUGH, J. (2018). **The Implications of a Subtle Difference in the Calculation of Affect Dynamics**. *Proceedings of the 26th International Conference on Computers in Education*.

Kinnebrew, J., Loretz, K., and Biswas, G. (2013a). A contextualized, differential sequence mining method to derive students' learning behavior patterns. *Journal of Educational Data Mining*, 5(1):190–219.

Kinnebrew, J., Segedy, J., and Biswas, G. (2017). Integrating model-driven and data-driven techniques for analyzing learning behaviors in open-ended learning environments. *IEEE Trans. Learn. Technol.*, 10(2):140–153.

Kinnebrew, J. S., Loretz, K. M., and Biswas, G. (2013b). A contextualized, differential sequence mining method to derive students' learning behavior patterns. *JEDM— Journal of Educational Data Mining*, 5(1):190–219.

Kinnebrew, J. S., Segedy, J. R., and Biswas, G. (2014). Analyzing the temporal evolution of students' behaviors in open-ended learning environments. *Metacognition and learning*, 9(2):187–215.

Lajoie, S. and Derry, S. (1993). *"Computers as Cognitive Tools"*. Lawrence Erlbaum Associates, Mahwah, NJ.

Land, S. (2000). Cognitive requirements for learning with open-ended learning environments. *Educational Technology Research and Development*, 48:61–78.

Larson, R. and Richards, M. (1991). Boredom in the middle school years – blaming schools versus blaming students. *American Journal of Education*, 99(4):418–443.

Leelawong, K. and Biswas, G. (2008). Designing learning by teaching agents: The betty's brain system. *International Journal of Artificial Intelligence in Education*, 18(3):181–208.

Luckin, R. and du Boulay, B. (2016). Reflections on the ecolab and the zone of proximal development. *International Journal of Artificial Intelligence in Education*, 26(1):416–430.

Luckin, R. and Hammerton, L. (2002). Getting to know me: Helping learners understand their own learning needs through metacognitive scaffolding. *International Conference on Intelligent Tutoring Systems*, page 759–771.

McDuff, D., Amr, M., Mahmoud, A., Turcot, J., Mavadati, M., and Kaliouby, R. (2016). Affdex sdk: a cross-platform realtime multi-face expression recognition toolkit. ACM.

Metcalfe, J. and Finn, B. (2013). Metacognition and control of study choice in children. *Metacognition and learning*, 8(1):19–46.

Mishra, S., Munshi, A., Rushdy, M., and Biswas, G. (2019). **LASAT: Learning Activity Sequence Analysis Tool**. In *TEEL Workshop at the 9th International Learning Analytics and Knowledge (LAK) Conference*, Arizona.

Mitrovic, A. (2012). Fifteen years of constraint-based tutors: what we have achieved and where we are going. *User Model. User Adapt. Interact*, 22(1–2):39–72.

Moreno, R. and Mayer, R. E. (2000). Engaging students in active learning: The case for personalized multimedia messages. *Journal of educational psychology*, 92(4):724.

Munshi, A. and Biswas, G. (2019). **Personalization in OELEs: Developing A Data-Driven Framework to Model and Scaffold SRL Processes**. In *Artificial Intelligence in Education*, pages 354–358, Chicago. Springer, LNCS.

Munshi, A., Biswas, G., Davalos, E., Logan, O., Narasimham, G., and Rushdy, M. (2022a). **Adaptive scaffolding to support strategic learning in an OELE**. In *Proceedings of the 30th International Conference on Computers in Education (ICCE 2022)*, Malaysia.

Munshi, A., Biswas, G., Rushdy, M., Baker, R., Ocumpaugh, J., and Paquette, L. (2022b). **Analyzing Adaptive Scaffolds to Help Students Develop Self-Regulated Learning Behaviors**. *Journal of Computer Assisted Learning*.

Munshi, A., Mishra, S., Zhang, N., Paquette, L., Ocumpaugh, J., Baker, R., and Biswas, G. (2020). **Modeling the Relationships Between Basic and Achievement Emotions in Computer-Based Learning Environments**. In Bittencourt, I. I., Cukurova, M., Muldner, K., Luckin, R., and Millán, E., editors, *Artificial Intelligence in Education*, pages 411–422. Springer.

Munshi, A., Rajendran, R., Moore, A., Ocumpaugh, J., and Biswas, G. (2018a). Studying the interactions between components of self-regulated learning in open ended learning environments. In *Proceedings of the 13th International Conference of the Learning Sciences (ICLS)*, pages 1691–1692, London, England.

Munshi, A., Rajendran, R., Moore, A., Ocumpaugh, J., and Biswas, G. (2018b). **Studying the Interactions between Components of Self-Regulated Learning in Open Ended Learning Environments**. In *Proceedings of the 13th International Conference of the Learning Sciences (ICLS)*, pages 1691–1692, London, England.

Munshi, A., Rajendran, R., Ocumpaugh, J., Biswas, G., Baker, R. S., and Paquette, L. (2018c). **Modeling Learners' Cognitive and Affective States to Scaffold SRL in Open-Ended Learning Environments**. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, pages 131–138, Singapore. ACM.

Narciss, S., Proske, A., and Koerndle, H. (2007). Promoting self-regulated learning in web-based learning environments. *Computers in human behavior*, 23(3):1126–1144.

Ocumpaugh, J., Baker, R. S., and Rodrigo, M. M. T. (2015). Baker rodrigo ocumpaugh monitoring protocol (bromp) 2.0 technical and training manual. *Technical Report. New York, NY: Teachers College, Columbia University. Manila, Philippines: Ateneo Laboratory for the Learning Sciences*.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in psychology*, 8:422.

Panadero, E., Klug, J., and Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: when measurement and intervention come hand in hand. *Scandinavian Journal of Educational Re-search*.

Pintrich, P. (2000). The role of goal orientation in self regulated learning. *Handbook of Self Regulation, eds M. Boekaerts, P. R. Pintrich, and M. Zeidner*, pages 452–502.

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International journal of educational research*, 31(6):459–470.

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into practice*, 41(4):219–225.

Pintrich, P. R., Smith, D. A. F., Garcia, T., and Mckeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (mslq). *Educational and Psychological Measurement*, 53(3):801–813.

Plass, J. L., Homer, B. D., and Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, 50(4):258–283.

Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., and Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational psychologist*, 27(1):91–109.

Puntambekar, S. and Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: what have we gained and what have we missed? *Educ. Psychol.*, 40(1):1–12.

Rajendran, R., Kumar, A., Carter, K. E., Levin, D. T., and Biswas, G. (2018a). Predicting learning by analyzing eye-gaze data of reading behavior. In *EDM*.

Rajendran, R., Munshi, A., Emara, M., and Biswas, G. (2018b). **A Temporal Model of Learner Behaviors in OELEs using Process Mining**. In *Proceedings of the 26th International Conference on Computers in Education (ICCE)*, Philippines.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

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

Schraagen, J. M., Chipman, S. F., and Shalin, V. L. (2000). *Cognitive task analysis*. Psychology Press.

Schraw, G., Crippen, K. J., and Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in science education*, 36(1-2):111–139.

Schunk, D. H. and Zimmerman, B. J. (2012). *Motivation and self-regulated learning: Theory, research, and applications*. Routledge.

Schwartz, D. L., Chase, C., Chin, D. B., Oppezzo, M., Kwong, H., Okita, S., Biswas, G., Roscoe, R., Jeong, H., and Wagster, J. (2009). Interactive metacognition: Monitoring and regulating a teachable agent. *Handbook of metacognition in education*, pages 340–358.

Segedy, J., Kinnebrew, J., and Biswas, G. (2013). The effect of contextualized conversational feedback in a complex open-ended learning environment. *Educational Technology Research and Development*, 61(1):71–89.

Segedy, J. R. (2014). Adaptive scaffolds in open-ended computer-based learning environments.

Segedy, J. R., Kinnebrew, J. S., and Biswas, G. (2015). Using coherence analysis to characterize self-regulated learning behaviours in open-ended learning environments. *Journal of Learning Analytics*, 2(1):13–48.

Self, J. (1988). Student models: what use are they. *Artificial intelligence tools in education*, pages 73–86.

Self, J. (1998). The defining characteristics of intelligent tutoring systems research: Itss care, precisely. *Int. J. Artif. Intell. Educ.*, 10:350–364.

Sharma, K., Mishra, S., Papamitsiou, Zacharoula, M. A. D. B., Biswas, G., and Giannakos, M. (2020). **Towards obtaining facial proxies for gaze behaviour in TEL**. In *Proceedings of the International Conference of the Learning Sciences (ICLS)*, volume 5, pages 2621–2622, Nashville.

Taub, M., Azevedo, R., Bouchet, F., and Khosravifar, B. (2014). Can the use of cognitive and metacognitive selfregulated learning strategies be predicted by learners' levels of prior knowledge in hypermedia-learning environments? *Computers in Human Behavior*, 39:356– 367.

Taub, M., Sawyer, R., Smith, A., Rowe, J., Azevedo, R., and Lester, J. (2020). The agency effect: The impact of student agency on learning, emotions, and problem-solving behaviors in a game-based learning environment. *Computers Education*, 147.

Tuan, H., Chin, C., and Shieh, S. (2005). The development of a questionnaire to measure students' motivation towards science learning. *International Journal of Science Education*, 27(6):639–654.

U.S. Department of Education, O. o. E. T. (2016). Future ready learning: Reimagining the role of technology in education. *Office of Educational Technology, US Department of Education*.

Verpoorten, D., Glahn, C., Kravcik, M., Ternier, S., and Specht, M. (2009). Personalisation of learning in virtual learning environments. In Cress, U., Dimitrova, V., and Specht, M., editors, *Learning in the Synergy of Multiple Disciplines*, pages 52–66, Berlin, Heidelberg. Springer Berlin Heidelberg.

Vygotsky, L. S. (1978). Lmind in society: The development of higher psychological processes. *Cambridge, Mass.: Harvard University Press.*

Wang, F. and Hannafin, M. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, 53:5–23.

Weinstein, C. and Meyer, D. (1994). Learning strategies, teaching and testing. *The international encyclopedia of education*, 2:3335–3340.

Winne, P. (1995). Inherent details in self-regulated learning. *Educational psychologist*, 30:173–187.

Winne, P. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition Learning*, 9:229–237.

Winne, P. and Hadwin, A. (2013). nstudy: Tracing and supporting self-regulated learning in the internet. *International handbook of metacognition and learning technologies*, page 293–308.

Winne, P. H. (2001). Self regulated learning viewed from models of information processing. *Self Regulated Learning and Academic Achievement, eds B. J. Zimmerman and D. H. Schunk*, pages 153–190.

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational psychologist*, 45(4):267–276.

Winne, P. H. (2011). A cognitive and metacognitive analysis of self-regulated learning. *Handbook of Self-Regulation of Learning and Performance, eds B. J. Zimmerman and D. H. Schunk*, page 15–32.

Winne, P. H. (2017). Learning analytics for self-regulated learning. *Handbook of learning analytics*, pages 241–249.

Winne, P. H. and Hadwin, A. F. (1998). Studying as self-regulated engagement in learning. *Metacognition in educational theory and practice*, pages 277–304.

Winne, P. H. and Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. *Motivation and self-regulated learning: Theory, research, and applications*, 2:297–314.

Winne, P. H., Hadwin, A. F., and Gress, C. (2010). The learning kit project: Software tools for supporting and researching regulation of collaborative learning. *Computers in Human Behavior*, 26(5):787–793.

Wittgenstein, L. (1968). Philosophical investigations. *New York: Macmillan*.

Wolters, C. A. and Hussain, M. (2015). Investigating grit and its relations with college students' self-regulated learning and academic achievement. *Metacognition and Learning*, 10(3):293–311.

Wood, D., Bruner, J., and Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 17:89–100.

Yun, H., Fortenbacher, A., Helbig, R., Geißler, S., and Pinkwart, N. (2019). Emotion recognition from physiological sensor data to support self-regulated learning. In *International Conference on Computer Supported Education*.

Zhang, N., Biswas, G., and Hutchins, N. (2021a). Measuring and analyzing students' strategic learning behaviors in open-ended learning environments. *Int J Artif Intell Educ (2021)*.

Zhang, Y., Paquette, L., Baker, R., Ocumpaugh, J., Bosch, N., Biswas, G., and Munshi, A. (2021b). **Can strategic behavior facilitate confusion resolution?** *Journal of Learning Analytics*.

Zhang, Y., Paquette, L., Baker, R., Ocumpaugh, J., Bosch, N., Munshi, A., and Biswas, G. (2020). **The relationship between confusion and metacognitive strategies in Betty's Brain**. In *Proceedings of the 10th International Learning Analytics and Knowledge (LAK) Conference*, Frankfurt.

Zimmerman, B. (1989). A social cognitive view of self-regulated academic learning. *J. Educ. Psychol*, 81:329–339.

Zimmerman, B. (1994). Dimensions of academic self regulation: A concep-tual framework for education. *Schunk, D.H.; Zimmerman, B.J. (Eds.), Self regulation of Learning and Performance Issues and Educational Applications*, pages 3–21.

Zimmerman, B. (2001). Theories of self regulated learning and academic achievement: An overview and analysis. *B. J. Zimmerman D. H. Schunk (Eds.), Self regulated learning and academic achievement*, pages 1–37.

Zimmerman, B. (2002a). Becoming a self regulated learner: An overview. *Theory Into Practice*, 41(2):64–70.

Zimmerman, B. J. (2002b). Becoming a self-regulated learner: An overview. *Theory into practice*, 41(2):64–70.

Zimmerman, B. J. and Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of educational Psychology*, 82(1):51.

Zimmerman, B. J. and Moylan, A. R. (2009). Self-regulation: where metacognition and motivation intersect. *Handbook of Metacognition in Education, eds D. J. Hacker, J. Dunlosky, and A. C. Graesser (New York, NY: Routledge)*, page 299–315.

Zimmerman, B. J. and Pons, M. M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American educational research journal*, 23(4):614–628.