Analyzing Adaptive Scaffolds that Help Students Develop Self-Regulated Learning Behaviors

Anabil Munshi[1], Gautam Biswas[2], Ryan Baker[3], Jaclyn Ocumpaugh[4], Stephen Hutt[5], Luc Paquette[6]

[1]*Corresponding author.* Institute for Software Integrated Systems, Vanderbilt University.

[2]Institute for Software Integrated Systems, Vanderbilt University.

[3]Graduate School of Education, University of Pennsylvania.

[4]Penn Center for Learning Analytics, University of Pennsylvania.

[5]Penn Center for Learning Analytics, University of Pennsylvania.

[6]Department of Curriculum & Instruction, University of Illinois at Urbana-Champaign.

Abstract

Background: Providing adaptive scaffolds to help learners develop effective self-regulated learning (SRL) behaviors has been an important goal for intelligent learning environments. Adaptive scaffolding is especially important in open-ended learning environments (OELE), where novice learners often face difficulties in completing their learning tasks.

Objectives: This paper presents a systematic framework for adaptive scaffolding in Betty's Brain, a learning-by-teaching OELE for middle school science, where students construct a causal model to teach a virtual agent, generically named Betty. We evaluate the adaptive scaffolding framework and discuss its implications on the development of more effective scaffolds for SRL in OELEs.

Methods: We detect key cognitive/metacognitive *inflection points*, i.e., moments where students' behaviors and performance change during learning, often suggesting an inability to apply effective learning strategies. At inflection points, Mr. Davis (*a mentor agent in Betty's Brain*) or Betty (*the teachable agent*) provides context-specific conversational feedback, focusing on strategies to help the student become a more productive learner, or encouragement to support positive emotions. We conduct a classroom study with 98 middle schoolers to analyze the impact of adaptive scaffolds on students' learning behaviors and performance. We analyze how students with differential pre-to-post learning outcomes receive and use the scaffolds to support their subsequent learning process in Betty's Brain.

Results and Conclusions: Adaptive scaffolding produced mixed results, with some scaffolds (viz., strategic hints that supported debugging and assessment of causal models) being generally more useful to students than others (viz., encouragement prompts). Additionally, there were differences in how students with high versus low learning outcomes responded to some hints, as suggested by the differences in their learning behaviors and performance in the intervals after scaffolding. Overall, our findings suggest how adaptive scaffolding in OELEs like Betty's Brain can be further improved to better support SRL behaviors and narrow the learning outcomes gap between high and low performing students.

Implications: This paper contributes to our understanding and impact of adaptive scaffolding in OELEs. The results of our study indicate that successful scaffolding has to combine context-sensitive inflection points with conversational feedback that is tailored to the students' current proficiency levels and needs. Also, our conceptual framework can be used to design adaptive scaffolds that help students develop and apply SRL behaviors in other computer-based learning environments.

Keywords

adaptive scaffolding, open-ended learning environments, self-regulated learning, agents, conversational feedback, learning strategies, feedback effectiveness

1. Introduction

An important goal of computer-based learning environments (CBLEs) is to help students develop self-regulated learning (SRL) behaviors that can make them effective life-long learners (Bransford et al., 2000; Zimmerman and Martinez-Pons, 1990). **Self-regulated learning (SRL)** focuses on learners' abilities to understand and control their learning behaviors, which helps them to accomplish their learning and problem-solving goals (Panadero, 2017). This process emphasizes the students' autonomy, strategy use, self-monitoring, and self-reflection during problem-solving. Open-ended learning environments (OELEs) have been designed to support SRL development by providing students with (1) *targeted learning goals* (e.g., construct a causal 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 offers choice in how students combine these tools to achieve their learning goals (Biswas et al., 2016). OELEs often use model-building tasks to help students improve their strategic thinking skills (Segedy, et al, 2015; Basu et al., 2017; Hutchins et al., 2020).

However, open-ended problem-solving can present significant challenges for novice learners (Kinnebrew et al., 2017; Metcalfe and Finn, 2013). They may face difficulties in using the system tools, making it hard to explicitly regulate their own learning processes in these environments (Zimmerman, 2002). These students require timely guidance via targeted *scaffolds* to help them develop effective strategies that address their difficulties and improve learning outcomes.

Frameworks for studying SRL behaviors in computer-based learning environments suggest that researchers need to examine and scaffold for an interacting collection of students' *"CAMM"* processes (Azevedo et al., 2017; Bannert et al., 2017), namely:

- *Cognition*, which includes the use of prior knowledge, skills, and strategies to develop solutions for the learning task (Entwistle and Ramsden, 2015);

- *Affect,* which is the ability to identify and regulate one's emotions during learning (Linnenbrink, 2007);
- *Metacognition,* which involves monitoring progress toward goals, invoking and applying the appropriate cognitive strategies, and periodically reflecting on their outcomes to inform further strategizing towards goals (Schraw et al., 2006); and
- *Motivation,* which is the perceived value of the learning task and the subject matter being learned (task value), the self-perceived ability to accomplish the task (self-efficacy) and one's personal goals (intrinsic versus extrinsic) for doing the task (Pintrich, 1999).

Learning environments that scaffold one or more of students' CAMM processes can empower them to become more strategic in their learning process and successful in achieving their learning tasks (Azevedo et al., 2017; Taub et al., 2020). This form of scaffolding requires *online adaptation*, where the system infers students' behaviors and performance in the OELE and uses this information to adapt and provide feedback (Dabbagh and Kitsantas, 2012; Moreno and Mayer, 2000). Plass et al. (2015) discuss the importance of providing adaptive scaffolds where the feedback is contextualized by the learner's current tasks, intent, and capabilities.

This paper develops and implements an adaptive scaffolding framework in the Betty's Brain OELE (Biswas et al., 2016; Leelawong and Biswas, 2008) to support certain CAMM dimensions of students' SRL behaviors, viz., their cognitive-metacognitive strategy development process, and to some extent, their emotion regulation. More specifically, our scaffolding framework includes methods for detecting and understanding students' learning behaviors around key *'inflection points'*, or moments during learning when they undergo a change in their cognitive/metacognitive processes. We hypothesize that changes are often linked to their inability to apply productive strategies, thus leading to a drop in their performance. In Betty's Brain, this translates to their inability to correct errors and improve their causal maps. So, the inflection points provide a basis for generating contextualized *in-the-moment* scaffolds that can help students to learn and apply more effective strategies that improve their overall outcomes in Betty's Brain. We provide the scaffolding to students through conversations initiated at the inflection points by one of the virtual agents present in Betty's Brain: (1) a mentor agent, Mr. Davis, or (2) the teachable agent, Betty.

We evaluate our adaptive scaffolds by first conducting a design study and then analyzing the effectiveness of our scaffolds using an exploratory data analysis approach. Since the pedagogical objective of our scaffolds is to help the learner adopt more effective cognitive/metacognitive strategies that improve their learning outcomes in Betty's Brain, our data analysis approach explores the relations between learning outcomes and scaffold effectiveness. As a framework, we use the overall differences in students' learning outcomes (i.e., their pre-post learning gains) from the study to conduct an exploratory comparison study on how effective our scaffolds are for two groups of students – *the High Learning Gain Group* and *the Low Learning Gain Group*. We use a temporal log analysis approach to compare students' learning behaviors and performance in the time intervals *before* and *after* they receive an adaptive scaffold, assessing major changes in their cognitive behaviors, task performance and affect.

The rest of this paper is organized as follows. *Section 2* reviews prior research on adaptive scaffolding for SRL in computer-based learning environments. *Section 3* discusses the Betty's Brain system and our previous work on developing adaptive scaffolds. *Section 4* presents the development of our new adaptive scaffolding framework. *Section 5* discusses the empirical methods of the classroom study we conducted to evaluate our scaffolding framework, the findings of which are reported in *Section 6*. Finally, conclusions and implications for future research are provided in *Section 7*.

2. Prior Research

Adapting to the specific needs of students has always been a key goal of intelligent computer-based learning environments (CBLEs; Lajoie and Derry, 1993). But novice learners, who are not proficient in using these tools and lack SRL processes, often adopt sub-optimal strategies in their learning and problem solving tasks. This increases the difficulties they face in their learning tasks, and providing students with relevant *in-time* feedback can help them overcome difficulties and become better learners (Puntambekar and Hubscher, 2005, van der Kleij, et al, 2015).

2.1 Scaffolding in CBLEs

Scaffolds are *“tools, strategies, and guides used to support understanding beyond one’s immediate grasp”* (Graesser et al., 2000; Azevedo and Hadwin, 2005). Research has shown that scaffolds and feedback can improve critical thinking (Wood et al., 1976) and learning outcomes, including those for higher-order constructs (Van der Kleij et al., 2015). Scaffolds developed for Betty’s Brain — namely contextualized conversational feedback from virtual agents — have led to better overall performance by students (Segedy et al., 2013).

Properly designed scaffolds can help foster self-regulation and engagement and reduce frustration (Lepper and Chabay, 1985; Shute, 2008), but there are sometimes unintended consequences. Students may exploit scaffolding features, as Baker et al. (2004) show in their study of

Cognitive Tutors (Koedinger et al., 2006), where students skip learning activity suggestions provided in low-level hints to get to the answers in *bottom-out hints*. In addition, feedback that frequently interrupts workflow or focuses on summative evaluation can also negatively affect learning (Fedor et al., 2001).

Therefore, a sound design process for adaptive scaffolding should guide students towards the optimal use of scaffold content into their learning and problem-solving processes.

2.2 Modeling SRL Behaviors

Early models of SRL defined the construct as a static *"trait"* (Pintrich et al., 1993; Zimmerman & Martinez-Pons, 1986), but by the late 1990s, the concensus shifted toward process-based definitions, including the cyclical phases model (Zimmerman, 2002) and the COPES model (Winne & Hadwin, 1998). The current view on SRL is of a dynamic sequence of cognitive, affective, metacognitive, and motivational (CAMM) events (Azevedo et al., 2017; Panadero et al., 2016).

While SRL models now emphasize *dynamic* processes (Panadero et al., 2016), little research has examined these dynamics from learners' behavioral changes in learning environments, and new methods are needed for detecting and analyzing changes in students' CAMM processes during learning, so that we can develop scaffolds to help students internalize successful SRL processes. In Betty's Brain, prior analysis of interactions between cognitive and affective SRL components showed that virtual agents successfully scaffolded students' learning (Munshi et al., 2018b). In this paper, we extend such earlier findings to design an adaptive scaffolding framework that provides students with agent-initiated (1) guidance on cognitive-metacognitive strategies to support their learning tasks and (2) encouragement messages to support their motivation and affect.

2.3 Designing Adaptive Scaffolds in OELEs

There have been several approaches to designing adaptive scaffolds in computer-based learning environments. Elsom-Cook (1993) proposed that systems could individualize guidance by varying the form and content of the scaffolds according to the cognitive state of the learner. Later work suggested that *"an ongoing diagnosis of the student's current level of understanding of specific and related tasks"* is a pillar of effective scaffold design (Puntambekar and Hubscher, 2005). Basu, et al (2017) demonstrate the effectiveness of providing in-time strategic feedback to students in a computational thinking-based OELE for science learning. In this paper, we build on such previous work, and develop adaptive scaffolds that are *strategic* (help students invoke a procedure or piece of knowledge they are unable to apply properly) and informed by students' past learning behaviors and performance. In addition, we also develop encouragement scaffolds (praise or reassurance) to help learners avoid or overcome emotions that are detrimental to the learning process.

3. The Betty's Brain Open Ended Learning Environment

Betty's Brain, an OELE for middle school science, adopts the *learning-by-teaching* paradigm, where students build *causal models* of scientific processes to "teach" a virtual pedagogical agent, generically named Betty (Biswas et al., 2005; Leelawong and Biswas, 2008). As shown in Figure 1, the system provides students with resources and tools to learn, build, and check their models.

These resources include a **science book**, a set of hypermedia resource pages embedded within the system, that provide the knowledge students need to teach Betty. Students read sections of the book and identify concepts and *causal* (i.e., cause-and-effect) relations between concepts.

Students use the **causal map building tool,** a visual interface with a drag-and-drop menu, to teach Betty. The interface provides students with a visual representation of their causal map encompassing both constructs and subsequent causal links. The menu allows students to add, delete, and modify concepts and links.

Other tools facilitate evaluation of the causal map. The **query and quiz tools** allow students to probe Betty's knowledge of the science concepts and relations, by asking her to take either shorter 'section-specific' (and more targeted) quizzes or a more comprehensive 'Everything' (mastery) quiz. Betty's answers are dynamically generated from the information in the causal map and scored by the **mentor agent**, Mr. Davis. Quiz results direct students to errors in their current causal map, and students can make Betty explain her answers to specific quiz questions by highlighting the links Betty used to answer that question. Effective learners can use this information to make immediate map corrections or to determine which sections of the book to read next. Overall, the quizzes help students track Betty's progress, and by implication their own knowledge of the science concepts and relations.

Betty's Brain adopts a socio-constructivist approach to learning that encourages exploration, strategic thinking, and monitoring skills (Biswas et al., 2016). Mr. Davis, the mentor agent, provides relevant strategy-oriented feedback when students have difficulties building and checking their maps. For Mr. Davis to accomplish this, the system must track student progress, but the open-ended nature of the system can make interpreting and adapting to the student quite challenging.

Over the years, researchers have worked to make Betty's Brain more adaptive (Segedy et al., 2013; Kinnebrew et al., 2017). Segedy et al. (2013) used a conversation tree representation (e.g., Adams, 2010) to deliver agent-initiated conversational scaffolds. Biswas et al. (2016) discuss how the listener interface of Betty's Brain facilitates explicit, contextualized communication between the student, Betty, and Mr. Davis by analyzing the current causal map, the most recent quiz results, and the student's recent interactions with the system. However, additional development must consider scaffolding that reflects the changes in the theoretical understanding of SRL (viz., the shift from SRL as a *static trait* to a *dynamic process*).

4. The Adaptive Scaffolding Framework

Our adaptive scaffolding framework builds off the SRL models mentioned in Section 2.2, to support the design and implementation of contextualized conversational feedback constructs in Betty's Brain.

4.1 The Conceptual Framework

Winne and Hadwin's (1998) COPES model describes self-regulated learners as those who actively manage their learning by adopting behaviors that include monitoring their progress, and executing cognitive and metacognitive strategies to maintain their progress. *Cognitive strategies* are typically goal-directed and situation-specific, e.g., read to find a specific piece of information (Weinstein and Meyer, 1994). *Metacognitive strategies* involve more generally applicable processes that include planning, monitoring, and reflecting (Donker et al., 2014; Zhang et al., 2021). While cognitive strategies focus on our skills and operate on the knowledge of "objects" (Winne, 1995), *metacognitive* learning strategies involve deliberation on the use of particular cognitive processes and combining these processes to accomplish larger tasks (Winne and Hadwin, 2008). *Metacognitive monitoring* bridges the gap between cognition and metacognition, as it involves observing and evaluating one's own execution of cognitive processes to control and improve cognition (Kinnebrew et al., 2017). As suggested by the COPES model, monitoring progress and using cognitive and metacognitive strategies during learning are strong indicators of the development of self-regulated learning (SRL) behaviors.

Therefore, to effectively scaffold SRL behaviors in Betty’s Brain, we study how students monitor their progress, and use cognitive and metacognitive strategies as they work on their causal modeling tasks. Since novice learners are typically not good at monitoring, and reflecting on their strategy use, understanding their cognitive and metacognitive behaviors and possible use of strategies *in context* (of their recent activities and the task they are currently engaged in) can help us to design more *contextualized* scaffolds that better support their strategy development processes. With this understanding, we also monitor students’ affect and performance. By tracking students’ progress and interactions with the system, we can identify opportune times (i.e., inflection points) to provide *strategic hints* that help learners become aware of effective strategies for acquisition, construction, and reasoning with knowledge. *Encouragement hints* in the form of praise or reassuring messages can help them regulate their emotions during learning. We believe that contextualized cognitive and metacognitive strategy feedback will help students acquire the necessary SRL processes to become effective and independent learners (Shyr and Chen, 2018).

4.2 Design

Designing and delivering strategy-focused feedback in Betty’s Brain requires us to consider the different paths that learners may use to accomplish the complex goal of completing a causal map. To be successful, learners must decompose their goal of building a correct map into strategically ordered *tasks* and monitor their progress towards completing these tasks (Winne, 2014). Thus, to design appropriate scaffolds, we need to understand the context of students’ actions logged in the Betty’s Brain system in terms of their current goals and activities.

4.2.1 Detecting Students' Learning Behaviors *in Context*

Our framework considers three factors to understand the *context* of students' activities and behaviors in Betty's Brain: (1) the current task type (i.e., reading, constructing/refining the causal map, or evaluating the causal map); (2) the student's effectiveness in causal modeling tasks (i.e., adding correct versus incorrect links to their causal map); and (3) the relation between the current task and preceding tasks (e.g., adding relevant links to the concept map after reading a page). Taken together, these three factors help us infer students' learning difficulties in different task contexts (e.g., the inability to analyze quiz results to identify correct and incorrect links in their map);

To understand, track, and contextualize student behaviors in Betty's Brain, Kinnebrew et al. (2017) developed a hierarchical task model that maps students' tasks and sub-tasks to higher-level (i.e., more general) cognitive processes in the learning environment. A task model decomposes the complex task (i.e., teaching Betty a scientific process by constructing a causal map) into sub-tasks using cognitive task analysis methods (Schraagen et al., 2000). Figure 2 shows the three primary processes students need to be successful in the Betty's Brain environment: (1) *Information Acquisition* (IA, i.e., reading the hypertext resource pages or taking/organizing notes) (2) *Solution Construction* (SC, i.e., map building/refinement tasks), and (3) *Solution Assessment* (SA, i.e., quiz-related activities). In this paper, we extend our previous model to incorporate an additional task, "Organizing Information" (i.e., taking/editing notes; see Figure 2).

In addition to classifying individual behaviors into these higher-level processes, sequence mining methods can help us derive frequent strategies from logs of students' activities (Kinnebrew et al., 2013). For example, when students read resource pages and add to their map, they are demonstrating an $IA_{(read)} \rightarrow SC_{(build\ map)}$ strategy. Such combinations how students combine cognitive processes and regulate their learning by applying problem-solving strategies (Schwartz et al., 2009). To illustrate in more detail, applying an $IA_{\textit{(read a page)}} \rightarrow SC_{\textit{(add correct link from that page)}}$ strategy shows that a student is able to acquire information from the science book, by (a) identifying the section that contains causal information they need to teach Betty, (b) interpreting the causal relation in the context of their causal map, and (c) then translating the acquired causal relation into a correct increase/decrease link on their map. As students work in Betty's Brain, they may use different combinations of IA, SC, and SA tasks to accomplish their goals.

The system uses *pattern detectors* to track students' use of cognitive and metacognitive strategies combined with information on *effectiveness*, i.e., whether they add correct links to the causal map or remove incorrect ones (effective) or vice versa (ineffective), so these patterns may be classified as productive or unproductive (Munshi et al., 2018b). Prior work has identified a set of productive and unproductive strategies within Betty's Brain (Biswas et al., 2016; Kinnebrew et al., 2017; Munshi et al., 2018a), which we use as the foundation for our adaptive scaffold framework.

4.2.2 Determining the Conditions for Triggering Scaffolds

Our new framework contextualizes a (1) *triggering condition* (i.e., a behavior or sequence) to optimize the selection of the (2) the *content of the adaptive scaffold* so that, when a triggering condition is satisfied, the adaptive scaffolding system provides students relevant *in-the-moment* feedback to help them develop effective strategies and become better learners. Specifically, we formalize the selection of *inflection points*—conditions where prior analysis has typically shown a decrease in students' ability to apply effective strategies—as triggering conditions. We argue that these inflection points represent situations when students' self-regulation (CAMM) processes undergo a change as they work on their learning and problem-solving tasks. Therefore, these points suggest *key transitional moments* in students' learning behaviors and productivity and represent opportune moments for providing *in-the-moment* feedback assistance to students facing difficulties.

Our framework also seeks to address the relationships between cognition and affect (Munshi et al., 2018b) by including scaffolds that deliver *encouragement*—either through *reassurance* (e.g., when students find multiple errors in their model after taking a quiz) or *praise* (e.g., when students teach correct causal links to Betty). Their specific purpose is to help students to manage their affect so that they can continue to engage with the system when they face difficulties.

Table 2 presents a complete list of scaffolds we have implemented, along with their inflection point triggering conditions. For example, when a student produces an $IA \rightarrow SC$ sequence with ineffective map-building behaviors (viz., adding incorrect links or deleting correct links), *in-the-moment* feedback may suggest that the student quiz Betty to assess the effectiveness of their recent map edits. This may help the student combine $SA{\rightarrow}SC$ and $IA{\rightarrow}SC$ strategies, i.e., $SA{\rightarrow}IA{\rightarrow}SC$, to use quiz answers to identify and debug parts of the causal map. Likewise, an inflection point may reflect key affective experiences. For example, *confusion*, (cognitive disequilibrium) to *frustration* (D’Mello & Graesser, 2012) might occur if Betty’s quiz results reflect several ineffective map links. In such situations, triggering affect regulation scaffolding is likely more effective than relying solely on cognitive-metacognitive strategies.

4.2.3 Providing Conversational Scaffolds at Trigger Conditions

After identifying inflection points to serve as triggering conditions, we deliver scaffolds using a back-and-forth conversation format (Figure 4) between the student and one of the two virtual agents, Mr. Davis or Betty. We have shown that this engages students in more authentic social interactions (Segedy et al., 2013), allowing them to be more active participants in the conversation (D’Mello et al., 2006). Students can direct the discussion toward topics/information they feel are most relevant. The next section discusses our approach for implementing the scaffolding framework in Betty's Brain.

4.3 Implementation

An overview of the scaffolding framework implemented in Betty's Brain is shown in Figure 3. Students' primary actions logged by the system (Table 1) are: (1) *Reading* the resources; (2) *Making notes* as a memory aid and to organize the information read; (3) *Building* and refining the causal map; (4) *Requesting* Betty to take *quizzes*; and (5) *Checking explanations* to quiz answers to identify the links used to answer questions. These logged actions are mapped onto the higher-level cognitive processes using the task model in Figure 2.

Map-edit activities associated with an increase or decrease in the causal map score (computed as the *number of correct links* – *number of incorrect links* in the map) are identified in the logs by marking them with *-Eff* (effective) and *-Ineff* (ineffective) tags, respectively. For example, an $Edit_{\text{-}Ineff}$ is used for causal map edits that decrease students' map scores. These (*-Eff* and *-Ineff*) labels were also applied to pre-defined task sequences of cognitive and metacognitive strategies, as derived from work by Kinnebrew et al., (2014, 2017), as were labels of *coherence* (i.e., relevant or supported by the information they just received; Segedy et al., 2015).

We applied these labels to data collected from two Betty's Brain classroom studies (March 2017 and Dec 2018), and used two methods to determine candidates for triggering conditions. We used a combination of (1) sequential pattern mining (Kinnebrew et al., 2014) to identify frequent strategies that lead to a decline in performance, and (2) student interviews, to identify times where students articulated difficulties they encountered while working with Betty's Brain. This resulted in nine cognitive/metacognitive inflection points, for which we developed adaptive scaffolds.

Each scaffold (see Table 2) is structured as a *conversation tree* and is delivered to the student by Betty or Mr. Davis. Each step in the conversation is represented by a node in the conversation tree, with opportunities for students to respond at each step. Their responses help guide the subsequent feedback to meet their specific needs. Students have autonomy to exit the feedback at will, controlling the amount of feedback they want based on their judgements of relevance. Two example inflection points are shown in Figure 4 with their corresponding conversation trees: (a) $Edit_{\text{-}Ineff} \rightarrow Quiz$ (when a student edits the causal map incorrectly and then takes a quiz), and (b) $Read\text{-}Long \rightarrow Edit_{\text{-}Ineff}$ (when a student spends a long time reading and then edits the causal map incorrectly). In each figure the triggering condition (inflection point) is shown in green, and example conversation text is given in blue.

5. Methodology

To evaluate our new scaffolding's effectiveness, we ran a design study in February 2019 with sixth-grade students in an urban public school in the southeastern US. The school's population was 60% White, 25% Black, 9% Asian, and 5% Hispanic, with 8% enrolled in the free/reduced-price lunch program. (Individual classroom demographics were not collected.) During the study, 98 students built a causal model of the human thermoregulation system (regulation of human body temperature, Figure 5) using the updated version of Betty's Brain.

5.1 Study Design and Data Collection

The study was conducted over 6 consecutive days. On Day 1, students took a paper-based pre-test that used both multiple-choice & short-answer questions to evaluate students' domain understanding and causal reasoning skills. On Day 2, students worked on a practice unit to familiarize themselves with the Betty's Brain environment. On Days 3-5, students constructed causal models of thermoregulation in Betty's Brain. On Day 6, students took a post-test that was identical to the pre-test.

The paper-based pre- and post-tests included multiple-choice and short-answer questions (examples in *Appendix A*) that were designed in consultation with middle school educators in Nashville, TN, keeping the sixth-grade public school curriculum in mind. These tests have previously been used in multiple Betty's Brain classroom studies, such as Segedy et al., 2015, Munshi et al., 2018. Test reliability was determined by applying Flanagan's Formula (Chakrabartty, 2013) on students' pre-test scores, which resulted in a *split-half reliability coefficient* $r = 0.75$, suggesting that the tests were reliable.

Betty's Brain logged students' activities and affective states (as detected with trace data) with time stamps as they worked on the system. Specifically, affect detectors captured 5 achievement emotions in 20 second intervals: (1) engaged concentration, (2) boredom, (3) delight, (4) confusion, and (5) frustration using affect detection models (Jiang et al., 2018). All of Mr. Davis' and Betty's conversations were also logged in the system with time stamps. Students' *map scores*, used as a measure of performance, were updated every time students modified their map.

5.2 Research Questions for Exploratory Data Analysis

Our primary research objective for analyzing the data collected in this study was to evaluate our adaptive scaffolds, more specifically, to study how the scaffolds affected students' SRL behaviors and performance. To achieve this goal, we used an exploratory data analysis approach that combined students' performance, behaviors, and affect, as logged with time stamps in the learning environment.

First, an exploration of students' overall learning outcomes from the Betty's Brain intervention (reported in Section 6.1) revealed that the study participants could be categorized into two groups - *High Learning Gain* (High) and *Low Learning Gain (Low*) - based on differences in their pre-to-post-study gains in domain knowledge and causal reasoning skills. We used this observation to frame our overarching research question to study and compare the impact of scaffolds on these two groups of students.

RQ: *How did students from High and Low Learning Gain groups respond to receiving the different types of adaptive scaffolds (listed in Table 2) during their learning and causal model-building process in the Betty's Brain environment?* More specifically, if the type of scaffold was a *strategic hint*, did students in each group follow the suggestion given in the hint? Did that appear to influence their subsequent use of SRL strategies, as evidenced by changes in their cognitive behaviors and task performance? If the type of scaffold received was an *encouragement prompt* did it have a positive impact on student emotions?

The next section reports the main findings from our exploratory data analysis approach to answer the above research questions.

6. Results and Discussion

6.1 Exploratory Analysis of Students' Learning Outcomes

We operationalize learning outcomes in Betty's Brain using two measures: *(1) Normalized Pre-to-posttest Learning Gains (NLG)*, which provides us with a summative measure of students' learning of the science content during the Betty's Brain intervention; *(2) Map Scores (MS),* which is a sequence of formative map score measures that update each time a student makes a change to their causal map. The summative score is computed as a *learning gain*, i.e., $\frac{Post\ score - Pre\ score}{Max\ score - Pre\ score}$, while the map scores are calculated as $\#\ number\ of\ correct - \#\ of\ incorrect\ causal\ links$ in a student's map after every change they make to their maps.

The distribution of the Normalized pre-post Learning Gains (NLG) was close to normal, with only mild skew and no evidence of kurtosis, justifying the use of parametric statistical tests in Table 3. One-way ANOVA tests of the students' pre-test and post-test scores in Table 3 show statistically significant ($p < .05$) pre-to-post learning gains for the science content, with high effect size ($Cohen's\ d = 1.28$). This suggests that the Betty's Brain intervention helped students learn the science content. However, Table 3 and Figure 6 (which shows the distribution of students' NLG scores) together indicate that there was a considerable variation in the learning gains (range $[-0.16, +0.74]$($median = 0.18$; $mean\ = 0.2$; $SD\ = 0.19$).

Therefore, to add additional context to our analysis on the impact of feedback on students SRL behaviors, we divided the students ($N = 98$) into High Learning Gain or *"High"* ($N = 44$) and Low Learning Gain or *"Low"* ($N = 45$) groups using a median split on their NLG scores. We excluded 9 students who had NLG scores that were equal to or differed by 0.2 from the median (i.e., their scores were between $[0.16, 0.2]$) from this analysis to create separation between the two groups.

6.2 Impact of Adaptive Scaffolds on Students' SRL Process

To answer the research question from Section 5.2, we delved deeper into the impact of the scaffolds in Table 2 on *High* and *Low* students' cognitive processes and their use of strategies. In addition, we also tracked students' affect states and map building performance, especially around the inflection points that triggered the adaptive scaffolds presented to the students.

6.2.1 Differences in Scaffolds Received by High and Low Groups

Of the six strategy-related adaptive scaffolds (Hints 1-6), Hint3 and Hint4 were triggered very infrequently for all students ($\leq\ 5$), so we excluded them from further analyses. For the remaining strategy and encouragement scaffolds, we computed: (1) the average number of times students in the *High* and *Low* groups received a scaffold; and (2) the number of times students in each group received each type of scaffold during the intervention. Table 4 lists the number of times (0 to 4+) an adaptive scaffold was received.

Table 4 shows that the High group received more scaffolds than the Low group. For three of the seven types of scaffolds, Hints 2, 5, and 6, this difference was statistically significant ($p < .05$), with t(43)=1.6 for Hint2, t(76)=2.3 for Hint5, t(78)=4.7 for Hint6. Upon further observation, we realized that this was because the triggering conditions for a number of these scaffolds (see Table 2) required students to take quizzes to assess their progress, and the High group took quizzes more often (spending 26% of their total time on the system in taking quizzes and looking at quiz results) than the Low group (who spent 16% of their time in quiz-related activities). This distinct difference in the number of hints received by the two groups may imply that the students in the High group benefited from receiving more feedback than the Low group, therefore, giving them more chances to learn and apply learning strategies.. In the next section, we investigate whether they benefited more from the help they received, and if so, whether this was related to how effectively they interpreted and used the scaffolds.

None of the encouragement hints had any substantial impact on students' affective states (determined by changes in affect likelihood values) or their performance; therefore, we do not include them in subsequent discussion. While students did not show any negative transitions in their affective states after the feedback, they also did not show any positive changes. In a complex open-ended learning environment like Betty's Brain, it is possible that the reassurance would have been more useful if it was associated with actionable (strategic) information that the student could consciously use to improve their current maps (Tan and Biswas, 2006). We need to redesign our encouragement scaffolds in view of the above findings to make them more useful towards improving students' affective experiences in the learning environment.

6.2.2 Impact of Scaffolds on SRL Behaviors of the High & Low Groups

To evaluate how the adaptive scaffolds affected students' learning outcomes and behaviors, we tracked the change in their performance in the causal modeling task (*by tracking the change in their map scores (MS)*), their related cognitive and strategic processes, and their affect after receiving scaffolds. For this temporal analysis, we created sequences of *scaffold-triggered 'before' and 'after' intervals*, where the *after* interval for an adaptive scaffold started just after the adaptive scaffold was given to the student and extended till the student received the next scaffold from the system. Similarly, the *before* interval started from when students received the last adaptive scaffold to when the current scaffold was provided. To illustrate this, we consider a student who received two adaptive scaffolds during the course of their learning session – Hint2 at time t_i and Hint5 at time t_j. For the Hint2 scaffold, the student's before interval was $[0, t_i]$ and after interval was $[t_i, t_j]$, where the time 0 represents the start of the current session. Similarly, for Hint5, the before interval was $[t_i, t_j]$ and after interval was $[t_j, end]$, where end represents the end time of the session.

We used *"average map-score slope"* as a measure of their causal modeling performance in a *before* or *after* interval (Kinnebrew et al., 2014). Map-score slope in an interval is calculated as the slope of a regression line fitted to a student's map scores as a function of their map edits in that interval over time. By analyzing the Map-score slope and the students' strategic behaviors in the intervals before and after they received each scaffold, we analyzed the effectiveness of the scaffolds on students' performance and learning behaviors over time.

Next, we discuss our findings of the impact of the different Hint and Enc scaffolds on High and Low students' learning behaviors and performance.

Hint1 (Mark Correct Links on map): This hint reminded students to *mark (annotate) the correct causal links on their map* so they could keep track of the links that have been graded as correct by Mr. Davis versus the other links (some of which may be incorrect). This hint was triggered when the student took a quiz in which at least one of the answers was graded as 'correct', indicated by the *green checkmark* in Figure 1(c). Mr. Davis delivered this feedback if a student did not follow up by annotating the links associated with correct answers using the "Mark as correct" feature on their map. Table 4 shows that 57% of the High group ($n = 25$) and 75% of the Low group ($n = 35$) did not receive this hint. The remaining 19 High and 10 Low students received the hint once or twice during the entire intervention. Since many students did not mark their links, the trigger condition for this hint may need to be revised to ensure that more students receive and use it.

***Behavior*:** For the 19 High and 10 Low group students that got this hint at least once, we study if students used the hint strategically and adopted the link-marking behavior to improve their learning process and causal modeling performance (map scores).

In the interval *before* receiving Hint1, only one High student and one Low student had marked link(s) on their maps. In the interval after receiving Hint1 for the first time, 26 total links were marked by students on their maps (9 links marked by the High students and 17 links marked by the Low students). Within the High group, 13 of 19 students did not mark any links after getting the hint, 5 students marked 1 link each, and 1 student marked 4 links on their map. The student who marked the 4 successive links followed the link-marking actions by deleting an incorrect link from their map, suggesting that keeping track of the correct links through marking them on the map may have aided their map-debugging process.

Within the Low group, 4 of the 10 students who got the hint did not mark any links, three students marked 1 link each, one student marked 2 links, one marked 4 links, and one marked 8 links on their map upon receiving the hint. The student who marked 8 links switched between looking at the quiz results and marking the correct links and then deleted two incorrect links from the map, suggesting that this student was systematically applying this hint and marking the correct links also helped the student identify incorrect links that needed to be deleted from the map. 4 High students and 4 Low students got Hint1 a second time during their learning session, but none of these students marked any links following the second time they received the hint.

Performance: The average map-score slope in the interval before Hint1 was 0.5 for the High group (n=19) and 0.04 for the Low group (n=10). After students received Hint1 for the first time, the average map-score slope changed to 0.3 for the High group and 0.06 for the Low group. While these values may initially suggest that the hint was not very effective in helping students improve the quality of their causal maps, we get a clearer insight on feedback effectiveness by studying the students within each group who followed the feedback by marking correct links, versus the students who did not. The average map-score slope of High group students who followed Hint1 the first time ($n = 6$) changed from 0.4 before the hint to 0.6 after, whereas the High group students who did not follow the mentor's suggestion ($n = 13$) changed from 0.5 to 0.2. In the Low group, the average map-score slope changed from 0.05 to 0.2 for the students who followed the feedback (n=6) and from 0.02 to -0.15 for students who did not follow the feedback ($n = 4$).

Discussion*:* Overall, the findings on students' behavioral and performance changes after receiving Hint1 suggest that following the feedback and marking correct links may have had a marginally positive effect on students' subsequent ability to keep track of their correct versus incorrect links. We did obtain evidence of some students using the feedback strategically to monitor and improve their maps, but there were others who, despite receiving the scaffold, did not use it in an effective manner. We will have to improve the feedback in subsequent design iterations and provide additional information to help students understand the advantages of marking links correctly. In past studies, we have seen students correct links on their map, but later delete/change the link when some of the other quiz answers are incorrect (Kinnebrew et al., 2013). Therefore, marking links may be a useful memory aid to ensure correct links are not deleted or changed to be incorrect.

Hint2 (Assess map by taking Quiz): This adaptive scaffold was designed to inform students that having Betty take a quiz from time to time is an effective strategy to assess the correctness and completeness of their map. The hint was triggered when students read multiple science book pages but added incorrect links to the map. Betty delivered this hint to encourage students to check on how much she was learning. Table 4 shows that 30 High and 18 Low students received the hint at least once. A few High students received the hint up to seven times and two Low students received the hint four times. 14 High and 27 Low students never received this hint.

***Behavior*:** We study the impact of the hint on students' relevant cognitive behaviors, i.e., taking quizzes and then assessing the quiz results by viewing the answers and checking the explanations. We also look at the changes in MS values *before* to *after* they got this adaptive scaffold. 30 High and 18 Low students got Hint2 at least once, but only 12 High students and two Low students had taken a quiz before they received Hint2. After receiving Hint2 for the *first time*, 26 of the 30 High students and 13 of the 18 Low students took a quiz. Four of the High students and one Low student took multiple quizzes. When students got Hint2 a second time, they took a quiz immediately after.

This suggests that the majority of the students who received Hint2 responded to the feedback by taking a quiz, but it is not clear that they internalized this assessment strategy and used it on their own as they progressed further in their map building activities. We also investigate the subsequent activities of students who got Hint2 and followed it by taking a quiz. Of the 26 High students who took a quiz after getting Hint2 the first time, 10 students then went on to view the explanations to specific quiz answers, suggesting that these students were engaged in extended map assessment behaviors by analyzing the correct and incorrect answers in their quiz. Within the Low group, only two students clicked on quiz answers and checked quiz explanations after receiving the hint for the first time, but the numbers increased the next time. Unlike the 10 High group students, the Low group students did not engage in deeper map assessment behaviors the first time they received the hint. Over time, more Low students started analyzing quiz behaviors more extensively.

***Performance*:** The average map-score slope in the interval before Hint2 was -0.02 for the High group and -0.29 for the Low group. After receiving Hint2 for the first time, the average map-score slope for the High group increased to 0.45, but the average map-score slope for the Low group decreased further to -0.42. This implies that the High group was more effective in using the feedback the first time to assess and correct errors in their maps than the Low group, who had difficulties in assessing and correcting errors in their maps. However, when the Low students received Hint2 multiple times, their after-hint map-score slope increased, and students who received Hint2 a third time achieved an average slope of 0.33 in the after phase. This suggests that it took multiple hints for the Low group to understand and effectively apply the map assessment strategy.

Discussion: Overall, Hint2 was effective for both groups. However, the High performers were more adept at using the explanations for analyzing quiz answers to improve their map-building performance. In contrast, it took multiple hints for the Low group to eventually develop an effective strategy using Hint2. This suggests that more details on how and why Hint2 is useful may help the Low students develop effective debugging strategies faster.

Hint5 (Debug from Map) and Hint6 (Debug from Read). Hint5 and Hint6 were both designed to have Mr. Davis make more detailed suggestions on how students could debug the errors in their causal map after they had taken a quiz. Hint5 pointed them to specific erroneous links on the map (SC), whereas Hint6 focused on going back and reading specific pages in the science book to find information to correct their erroneous links (IA). Table 4 shows that all students received Hint5 (trigger: *SC-Ineff→SA*) and Hint6 (trigger: *SA→IA (multiple reads)*) more often than the other scaffolds. There could be two reasons why these two hints dominate: (1) the use of less stringent filtering criteria imposed by the pattern detectors for triggering these hints (see Section 4.2.1); and (2) students frequent inability to use the quiz results effectively. Section 6.2.3 showed that the High group used this this strategy more often than the Low group. In other words, they were using quizzes to debug their maps more often than the Low group, and, therefore, received the quiz-triggered hints more often. In the future, we may need to take into account students' performance and their current cognitive abilities in terms of their map checking behaviors in specifying the hint triggering conditions to better match student needs. We may also have to suggest to Low performers who check their maps very infrequently to take quizzes more frequently.

Hints 5 and 6 were often delivered in succession (38% of the time students received either hint) because both hints originated from quiz-taking episodes Therefore, we studied the impact of these two hints for three different cases: (a) when students received *Hint5 only*, (b) when students received *Hint6 only*, and (c) when students received both *Hint5 and Hint6 in succession*.

Hint5 only. Table 4 shows that all High group students received Hint5 at least once, with 41 students (93%) getting the hint four times or more during their learning session. 43 of the 45 students in the Low group, got Hint5 at least once, with 34 students (75%) receiving this scaffold four times or more during the intervention.

***Behavior*:** In the interval *before* receiving Hint5, the High group spent 59% of their time and the Low group 50% of their time on map-building activities. *After* receiving Hint5 for the first time, the High group spent an average of 57% of their time on map edits. This number increased to 65% after the third time they received the hint. For the Low group, the map editing time increased from 58% after the first time to 80% (a significant increase) after the third time they got the hint. Therefore, as their causal maps became more complex, Hint5 seemed to have a greater impact on students' map-building efforts, especially for the Low group.

***Performance*:** After receiving Hint5 for the first time, the average map-slope score changed from 0.17 to 0.14 for the High group and from -0.2 to 0.18 for the Low group. This suggests a marked improvement in performance for the Low group. Receiving this hint more than once had a positive effect on both groups, with the net value of the average map-slope score after getting Hint5 being 0.2 for the High group and 0.36 for the Low group. This suggests that the students used the information provided in Hint5 to successfully find and correct incorrect links on their map.

Hint6 only:. All students in the High and Low groups received Hint6 at least once during the intervention. 41 High students and 43 Low students got this hint four times or more.

***Behavior*:** Before receiving Hint6, the High group spent 37% of their time and the Low group spent 44% of their time reading the science book. After receiving Hint6 for the first time, the High group spent 39% of the time reading, while the Low group, who were already reading more than the High group, spent 57% of their time on the reading task. The time allocated to reading by the High group did not change much after receiving Hint6 multiple times. For the Low group, the reading time was the highest (57%) after the first time they received Hint6, and decreased thereafter to a stable value in the range 35−37% after they received the hint three or more times. This suggests that the High group, who were better at finding information in the science book, did not have to devote additional time to reading after they received Hint6, but they probably were more strategic in their approach. By contrast, the Low group spent more time reading after they got the hint for the first time. The change in map-slope score from *before* to *after* the hint gives us more insight on whether the Low students were able to use this additional strategic reading to debug their maps.

Performance**:** The average map-slope for the High group was 0.14 before they received Hint6 and did not change significantly upon receiving the hint, suggesting that Hint6 by itself did not result in a performance change for this group. For the Low group, the map-slope score changed from -0.07 in the interval before the hint to an average of -0.15 after the hint, with the score dropping to -0.43 as they got additional hints even though they read more. Despite students increasing their reading time after receiving the hint (especially the first time they received it), the Low group did not become more effective readers. This may be attributed to an inability to extract relevant knowledge when reading the science book text. This will be investigated further in future work.

<u>**Hints5+6:**</u> When students received the two hints in quick succession, they always received Hint5 before they received Hint6, prompting us to label these situations as Hints5+6. There was no significant change in student activities or performance trends after receiving Hints5+6. Their *before* and *after* interval map slopes showed fluctuations across intervals instead of a uniform upward/downward trend. On further inspection, we found that different students even within the same group, and at different points during their learning session, reacted differently to receiving Hints5+6, with some resorting to more reading and others to more map-editing activities, with no overall substantial differences in behavior between the High and Low groups. We believe that students became confused upon receiving the two hints in succession. hence their non-uniform reactions to the scaffolding. Emotion likelihood values generated by the affect detectors (Jiang et

al., 2018) provide circumstantial evidence here. Confusion likelihood scores increased after receiving Hints5+6; from 8% to 13% for the High group and for 8% to 12% for the Low group. This suggests a need for a “minimum inter-scaffold time” in the scaffold framework so that students have sufficient time to process and act upon a feedback before receiving another one.

7. Conclusions and Future Work

In this paper, we have developed a conceptual framework and used it to design and implement an adaptive scaffolding framework to help students develop and refine their SRL behaviors in the Betty’s Brain learning environment. Our system continually monitors inflection points, i.e., changes in students’ cognitive activities and map-building performance as they learn by building their causal maps in the environment. Online inflection point detection linked to strategy use and map building performance, support the delivery of in-time scaffolds (strategic hints and encouragements) adapted to the students’ current performance and behaviors. Results from an exploratory study run in a 6th-grade classroom showed that the students achieved significant pre- to post-test learning gains from a Betty’s Brain intervention that provided adaptive scaffolds. However, we also observed large differences in students’ pre-to-post learning gain scores, which prompted us to investigate if there were differences in how the two groups (*High* and *Low*) used adaptive scaffolds during learning in Betty’s Brain.

Our exploratory analysis is inconclusive, partly because we need to rethink how we define inflection points and define scaffolds. For example, the High group received more feedback on taking quizzes to assess their maps, because they took the quiz more often. Similarly, they received very important feedback on how to use the results of the quiz to debug their map more often for the same reason they took the quiz more often. It is clear we will have to use inflection points to detect the lack of use of effective SRL processes, and encourage and guide students on how to use them. This will be very useful for Low performers who may need this feedback more often to become more familiar with SRL processes, and, as a consequence, better learners.

In contrast, some of the High group's success may be attributed to their previous knowledge of or the ability to learn SRL strategies over time as they used the Betty's Brain system.. Overall, the findings reported in Section 6.2 show that some of our adaptive scaffolds were useful for students, whereas some others did not serve their intended purpose and need to be refined to make them more effective. For example, feedback suggesting link annotation by marking correct links (Hint1), while effective to an extent, may be further improved by demonstrating to learners how link annotation could help them spot incorrect links in their map more easily.

Overall, the differences in High and Low students' behaviors after they take a quiz based on the mentor's suggestion also presents opportunities for improving the feedback for the Low group, by providing additional scaffolding to help them interpret their quiz results and develop more effective $SA \rightarrow SC$ strategies. In fact, some scaffolds, such as Hint5, helped the Low group with map debugging after they took a quiz, and this helped them to improve their causal models.

More generally, this study and past studies (e.g., Leelawong & Biswas (2008); Schwartz, et al (2009); Roscoe, et al (2013)) demonstrate that in-time adaptive scaffolding directed toward learning and applying SRL strategies while problem solving in OELEs is essential to help students become better learners. However, the adaptive scaffold triggers need to be designed to be more cognizant of students' needs, and the evolution of such needs as they learn in the environment. An important role of adaptive scaffolding is to close the gap between students with High and Low learning outcomes, and careful design considerations and additional empirical analyses are needed to discover how to narrow this gap. We will address all of these issues in future work.

References

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

Azevedo, R. (2005) Using hypermedia as a metacognitive tool for enhancing student learning? the role of self regulated learning. Educational Psychologist, 40(4), 199–209.

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, 171–197.

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

Azevedo, R. and Hadwin, A. F. (2005) Scaffolding self-regulated learning and metacognition–implications for the design of computer-based scaffolds. Instructional Science, 33, 367–379.

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, 225–247. Springer.

Azevedo, R., Taub, M. and Mudrick, N. V. (2017) Understanding and reasoning about real-time cognitive, affec-tive, 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 Pro-ceedings of the 7th International Conference on Intelligent Tutoring Systems, 531–540.

Bannert, M., Molenaar, I., Azevedo, R., Järvelä, S. and Gašević, D. (2017) Relevance of learning analytics to meas-ure and support students’ learning in adaptive educational technologies. In Proceedings of the Seventh Interna-tional Learning Analytics & Knowledge Conference, 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, 5–53.

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

Boekaerts, M. and Niemivirta, M. (2000) Self-regulated learning: Finding a balance between learning goals and ego-protective goals. In Handbook of self-regulation, 417–450. Elsevier.

Bransford, J. D., Brown, A. L., Cocking, R. R. et al. (2000) How people learn, vol. 11. Washington, DC: National acad-emy press. Chi, M., Siler, S., Jeong, H., Yamauchi, T. and Hausmann, R. (2001) Learning from human tutoring. Cognitive Science, 25, 471–534.

Chakrabartty, S. N. (2013). Best split-half and maximum reliability. IOSR Journal of Research & Method in Educa-tion, 3(1), 1-8.

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, 3–8.

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

DMello, S. and Graesser, A. (2012) Dynamics of affective states during complex learning. Learning and Instruc-tion, 22(2), 145–157.

D’Mello, S. K., Craig, S. D., Sullins, J. and Graesser, A. C. (2006) Predicting affective states expressed through an emote-aloud procedure from autotutor’s mixed-initiative dialogue. International Journal of Artificial Intelligence in Education, 16, 3–28.

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

D’Mello, S., Lehman, B., Pekrun, R. and Graesser, A. (2014) Confusion can be beneficial for learning. Learning and Instruction, 29, 153–170.

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.

Fedor, D., Davis, W., Maslyn, J. and Mathieson, K. (2001) Performance improvement efforts in response to negative feedback: the roles of source power and recipient self-esteem. J. Manag., 27(1), 79–97.

Graesser, A., Wiemer-Hastings, K., Wiemer-Hastings, P., Kruez, R. and Group, T. T. R. (2000) Autotutor: A simulation of a human tutor. Journal of Cognitive Systems Research, 1, 35–51.

Grau, V. and Whitebread, D. (2012) Self and social regulation of learning during collaborative activities in the classroom: The interplay of individual and group cognition. Learning and Instruction, 22, 401–412.

Hadwin, A. F., Wozney, L. and Pantin, O. (2005) Scaffolding the appropriation of self-regulatory activity; a so-cio-cultural analysis of changes in teacher-student discourse about a graduate research portfolio. Instructional science, 33(5-6), 413– 450.

Huang, Y.-M., Liang, T.-H., Su, Y.-N. and Chen, N.-S. (2012) Empowering personalized learning with an interac-tive e-book learning system for elementary school students. Educational Technology Research and Develop-ment, 60, 703–722.

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 Ed-ucation and Technol- ogy, 29, 83–100.

Hutt, S., Baker, R.S., Ocumpaugh, J., Munshi, A., Andres, A., Karumbaiah, S., Slater, S., Biswas, G., Paquette, L., Bosch, N., and Van Velsen, M. (in press) Quick Red Fox: An App Supporting a New Paradigm in Qualitative Research on AIED for STEM. Artificial Intelligence in STEM Education: The Paradigmatic Shifts in Research, Education, and Technology.

Hutt, S., Grafsgaard, J. F. and D'Mello, S. K. (2019) Time to scale: Generalizable affect detection for tens of thousands of students across an entire school year. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–14.

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, 198–211. Springer.

Kanevsky, L. and Keighley, T. (2003) To produce or not to produce? understanding boredom and the honor in un-derachievement. Roeper Review, 26:1, 20–28.

Kinnebrew, J., Loretz, K. and Biswas, G. (2013) A contextualized, differential sequence mining method to de-rive 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., 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.

Koedinger, K. R., Corbett, A. et al. (2006) Cognitive tutors: Technology bringing learning sciences to the classroom. na.

Lajoie, S.P. and Azevedo, R. (2006) Teaching and learning in technology-rich environments. Handbook of Educational Psychology, 2nd edition, 803–821. Lajoie, S. and Derry, S. (1993) "Computers as Cognitive Tools". Lawrence Erlbaum Associates, Mahwah, NJ.

Leelawong, K., & Biswas, G. (2008). Designing learning by teaching agents: The Betty's Brain system. International Journal of Artificial Intelligence in Education, 18(3), 181-208.

Lepper, M. R. and Chabay, R. W. (1985) Intrinsic motivation and instruction: Conflicting views on the role of motivational processes in computer-based education. Educational Psychologist, 20, 217–230.

Linnenbrink, E. A. (2007) The role of affect in student learning: A multi-dimensional approach to considering the interaction of affect, motivation, and engagement. In Emotion in education, 107–124. Elsevier.

Martinez, M. (2001) Key design considerations for personalized learning on the web. Journal of Educational Technology & Society, 4, 26–40.

Metcalfe, J. and Finn, B. (2013) Metacognition and control of study choice in children. Metacognition and learning, 8, 19–46. Moreno, R. and Mayer, R. E. (2000) Engaging students in active learning: The case for personalized mul-timedia messages.

Journal of educational psychology, 92, 724.

Munshi, A., Mishra, S., Zhang, N., Paquette, L., Ocumpaugh, J., Baker, R. and Biswas, G. (2020) Modeling the re-lationships between basic and achievement emotions in computer-based learning environments. In Artificial In-telligence in Education (eds. I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin and E. Millán), 411–422. Cham: Springer International Publishing.

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 Inter-national Conference of the Learning Sciences (ICLS), 1691–1692. London, England.

Munshi, A., Rajendran, R., Ocumpaugh, J., Biswas, G., Baker, R. S. and Paquette, L. (2018b) Modeling learners' cognitive and affective states to scaffold SRL in open-ended learning environments. In Proceedings of the 26th Con-ference on User Modeling, Adaptation and Personalization, 131–138. ACM.

O'Leary, K. and O'Leary, S. (1977) Classroom management: The successful use of behavior modification. Pergamon Press, New York, NY.

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. R. (1999) The role of motivation in promoting and sustaining self-regulated learning. International jour-nal of educational research, 31, 459–470.

Pintrich, P., Smith, D., Duncan, T., and Mckeachie, W. (1993). Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (MSLQ). Educational and Psychological Measurement - EDUC PSYCHOL MEAS. 53. 801-813.

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.

Rathvon, N. (2008) Effective school interventions.

Roscoe, R. D., Segedy, J. R., Sulcer, B., Jeong, H. and Biswas, G. (2013) Shallow strategy development in a teacha-ble agent environment designed to support self-regulated learning. Computers & Education, 62, 286–297.

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, 111–139.

Schwartz, D. L., Chase, C., Chin, D. B., Oppezzo, M., Kwong, H., Okita, S., Biswas, G., Roscoe, R., Jeong, H. and Wag-ster, J. (2009) Interactive metacognition: Monitoring and regulating a teachable agent. Handbook of metacog-nition in education, 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., 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, 13–48.

Self, J. (1988) Student models: what use are they. Artificial intelligence tools in education, 73–86. Shute, V. (2008) Focus on formative feedback. Rev. Educ. Res., 78(1), 153–189.

Shyr, W.-J. and Chen, C.-H. (2018) Designing a technology-enhanced flipped learning system to facilitate stu-dents' self- regulation and performance. Journal of Computer assisted learning, 34, 53–62.

Skipper, Y. and Douglas, K. (2012) Is no praise good praise? effects of positive feedback on children’s and university students’ responses to subsequent failures. British Journal of Educational Psychology, 82, 327–339.

Steinkuehler, C. and Duncan, S. (2008) Scientific habits of mind in virtual worlds. Journal of Science Education and Technology, 17, 530–543.

Tan, J. and Biswas, G. (2006) The role of feedback in preparation for future learning: A case study in learning by teaching environments. In International Conference on Intelligent Tutoring Systems, 370–381. Springer.

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

Van der Kleij, F.M, Feskens, R. and Eggen, T. (2015). Effects of feedback in a computer-based learning environment on students’ learning outcomes: A meta-analysis. Review of educational research 85 (4), 475-511.

Van der Kleij, F. M., Feskens, R. C. and Eggen, T. J. (2015) Effects of feedback in a computer-based learning envi-ronment on students’ learning outcomes: A meta-analysis. Review of educational research, 85, 475–511.

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

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, 153–190.

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

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

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

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

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

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, 293–311.

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

Yen, M., Chen, S., Wang, C., Chen, H., Hsu, Y. and Liu, T. (2018) A framework for self-regulated digital learning (srdl). Journal of Computer Assisted Learning, 34(5), 580–589.

Zhang, N., Biswas, G. and Hutchins, N. (2021) Measuring and analyzing students' strategic learning behaviors in open-ended learning environments. International Journal of Artificial Intelligence in Education.

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

Zimmerman, B. (2000) Attaining self regulation: a social cognitive perspective. Handbook of Self Regulation, eds M. Boekaerts, P. R. Pintrich, and M. Zeidner, 13–40.

Zimmerman, B. J. (2002) Becoming a self-regulated learner: An overview. Theory into practice, 41, 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, 51.

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, 614–628.

Zimmerman, B. J. and Schunk, D. H. (2011) Handbook of Self Regulation of Learning and Performance. Routledge.

Tables

Table 1: Student Activities and Cognitive Processes Associated with Learning in Betty's Brain

Activity	Description	Cognitive Process
Read	Student reads resource pages (to learn about domain) or teacher's guide (to get suggestions for teaching Betty)	IA
Make Notes	Student takes/edits notes generated from reading resources for organizing information and for future reference	IA
Causal Map Edits	Student adds/deletes concepts or adds/deletes/modifies links to build/refine their causal map	SC
Take Quiz	Student asks Betty to take a quiz on a topic and reviews quiz results	SA
Quiz Expl	Student probes deeper into quiz results by checking the causal links Betty used to answer specific quiz questions	SA

Table 2: Inflection point triggers and their corresponding scaffolds

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

<table>
<tr><th colspan="2">Inflection Point Trigger</th><th colspan="2">Corresponding 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 Solution construction (Edit-Ineff)</td><td>Strategic hint: Assess by Quiz Hint2</td><td>Betty suggests taking a quiz, as a good 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>Ineffective Solution construction (Edit-Ineff) → Solution 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>
</table>

	Case 2: WHERE The Edit-Ineff was a *shortcut link* addition (e.g., an A→C link instead of an A→B→C link)	Strategic hint: Shortcut Link **Hint4**	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? ..."*
	Case 3 (No additional contexts)	Strategic hint: Debug from Map **Hint5**	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."*
	Case 4 (No additional contexts)	Encouragement: Reassure **Enc3**	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 this a little tricky. But with you to teach me, I'm sure we can do it."*
Solution assessment (Quiz) → Information acquisition (Read-Long)		Strategic hint: Debug from Read **Hint6**	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."*

(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 Solution 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"*
		Strategic hint:	Mr. Davis suggests marking the possibly correct links on the map as "correct", as an efficient map organization strategy.

Efficient Solution construction (Edit-Eff) → Solution assessment (Quiz)	Case 1	Mark Correct	*"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 …"*
		Hint1	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." 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 3: Pre-post learning outcomes: All students (n=98)

Pre/post question type	**Pre-test score**	**Post-test score**	**Pre to post learning gains (NLG)**	**Pre to post 1-way ANOVA**	**Effect size**
	mean (sd)	***mean (sd)***	***mean (sd)***	***F-ratio (p-value)***	***Cohen's d***
Multiple Choice (Max=8)	2.73 (1.3)	4.7 (1.92)	0.35 (0.41)	66 (< 0.05)	1.2
Short Answer (Max=15)	0.86 (1.03)	2.82 (2.33)	0.14 (0.15)	56 (< 0.05)	1.09
Overall (Max=23)	3.59 (1.9)	7.52 (3.9)	0.2 (0.19)	80 (< 0.05)	1.28

Table 4: Number of adaptive scaffolds for students in the High (n=44) and Low (n=45) groups

Adaptive Scaffold	**Category**	**No. of times a student got the scaffold**		**No. of students (% of category) who got the scaffold**				
		Range	**Mean (SD)**	**never**	**1 time**	**2 times**	**3 times**	**4+ times**
Hint1 Mark Correct	Hi	0-2	1.2 (0.4)	25 (57%)	15 (34%)	4 (9%)	0	0
	Lo	0-2	1.4 (0.5)	34 (75%)	7 (15%)	4 (9%)	0	0
Hint2 Assess by Quiz	Hi	0-7	2.3 (1.6)	14 (32%)	12 (27%)	6 (14%)	6 (14%)	6 (14%)
	Lo	0-4	1.8 (1.4)	27 (60%)	11 (24%)	3 (7%)	2 (4%)	2 (4%)
Hint5	Hi	1-35	13 (8.2)	0	2 (4%)	0	1 (2%)	41 (93%)

Debug from Map	Lo	0-37	11 (7.6)	2 (4%)	3 (6%)	4 (9%)	2 (4%)	34 (75%)
Hint6	Hi	3-45	23 (9.6)	0	0	0	1 (2%)	43 (97.7%)
Debug from Read	Lo	1-43	17 (10.6)	0	1 (2%)	2 (4%)	1 (2%)	41 (91%)
Enc1	Hi	0-4	1.5 (0.8)	25 (57%)	12 (27%)	5 (11%)	1 (2%)	1 (2%)
Praise	Lo	0-4	1.4 (0.8)	31 (69%)	11 (24%)	2 (4%)	0	1 (2%)
Enc2	Hi	0-4	1.5 (0.9)	23 (52%)	15 (34%)	2 (4%)	3 (7%)	1 (2%)
Praise & Quiz	Lo	0-2	1.3 (0.4)	30 (67%)	11 (24%)	4 (9%)	0	0
Enc3	Hi	0-3	1.3 (0.6)	31 (70%)	10 (23%)	2 (4%)	1 (2%)	0
Reassure	Lo	0-3	1.4 (0.5)	35 (77%)	8 (18%)	1 (2%)	1 (2%)	0

Figure Legends *(Figures submitted in a separate .pdf file)*

Figure 1: System interfaces for the Betty's Brain learning environment
(a) The 'science book' view, (b) The 'causal map' view, (c) The 'quiz results' view

Figure 2: The Hierarchical Task Model for the Betty's Brain Environment (Modified from Kinnebrew et al., 2017)

Figure 3: Implementation of the adaptive scaffolding framework in Betty's Brain

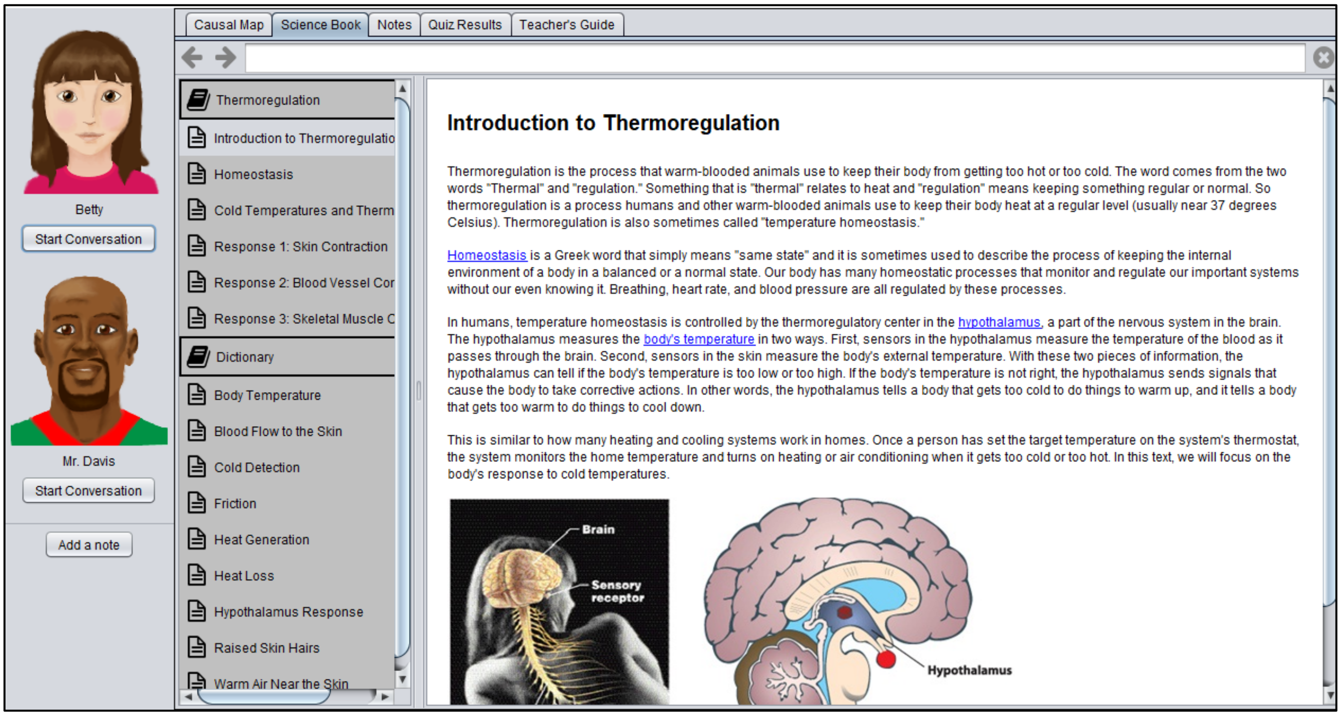
Figure 4: Conversation tree representations of two scaffolds from our framework
(a) Progression levels of a conversation tree for a map-debugging scaffold by Mr. Davis
(b) Conversation tree for Hint2, a map-assessment scaffold initiated by Betty

Figure 5: Causal map of the human thermoregulation process in Betty's Brain

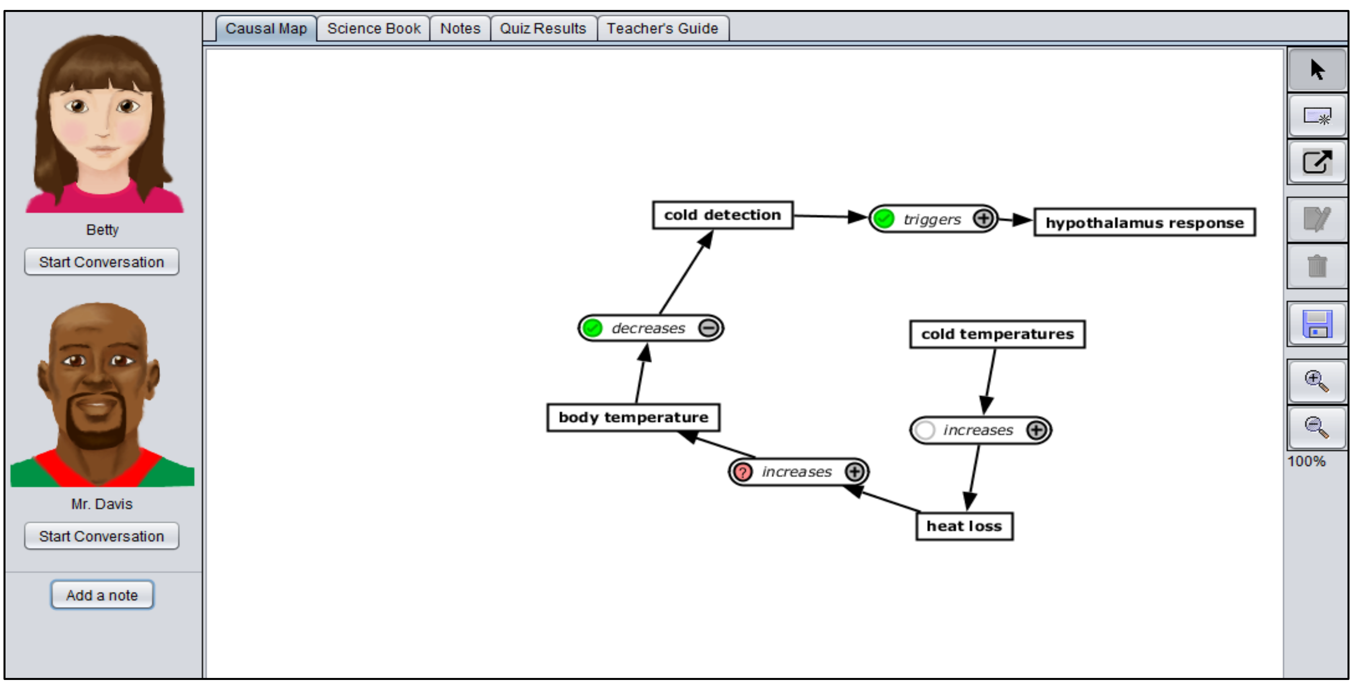
Figure 6: Distribution of the pre-to-post learning gain (NLG) scores (x-axis), by number of participants who achieved the score (y-axis), in the empirical study

Figures

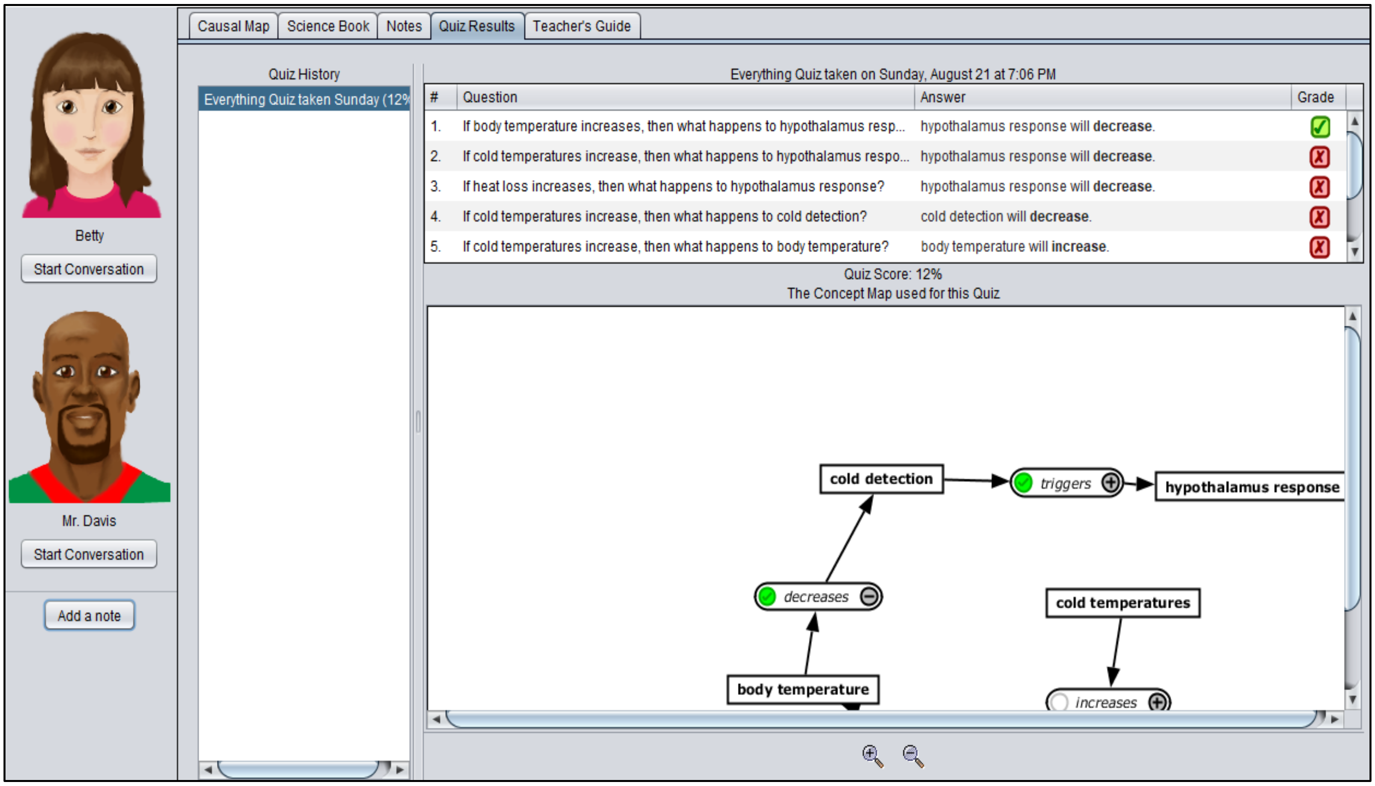
Figure 1



(a) The ‘science book’ view



(b) The ‘causal map’ view



(c) The 'quiz results' view

Figure 2

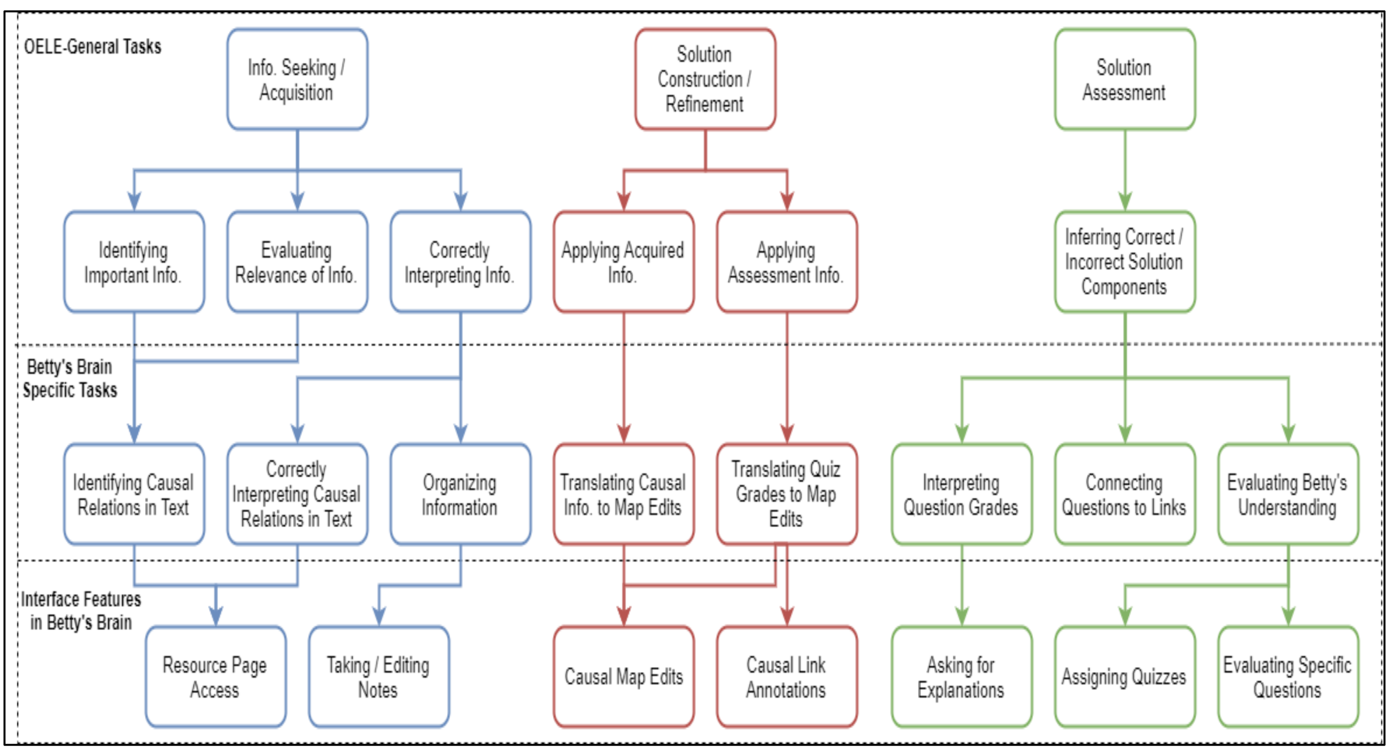


Figure 3

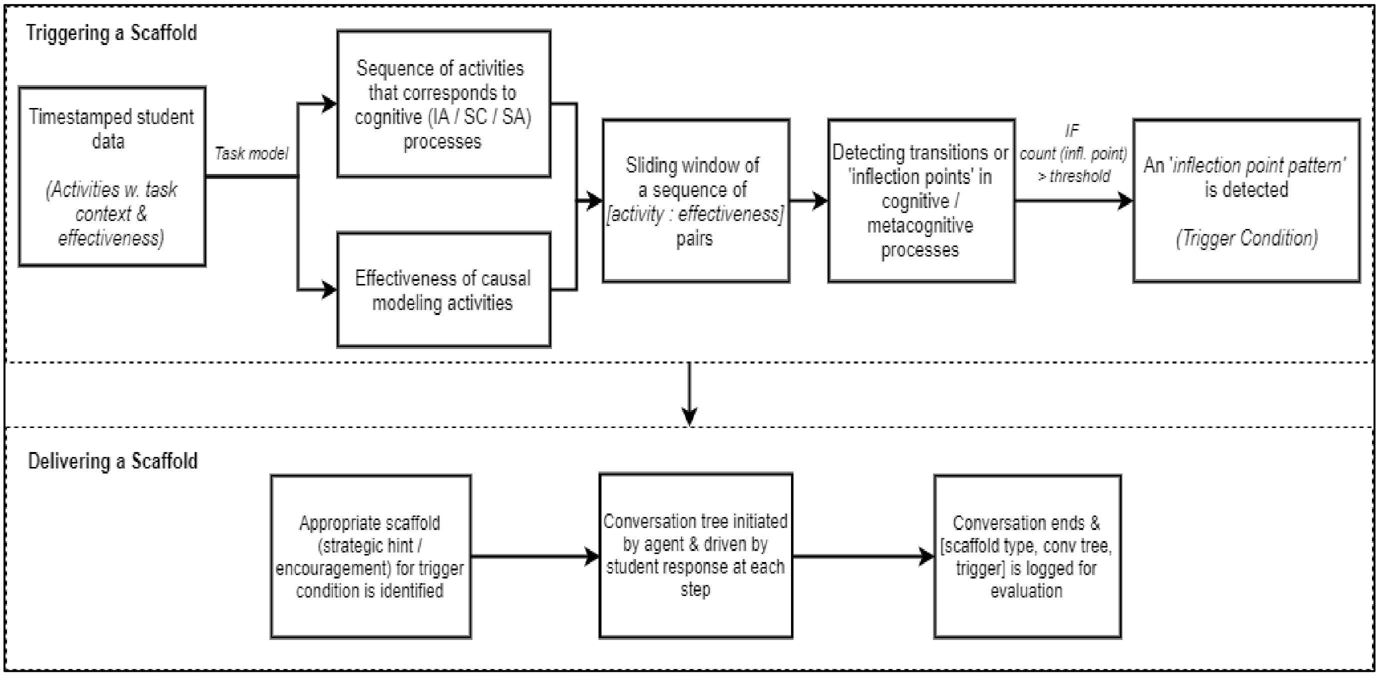
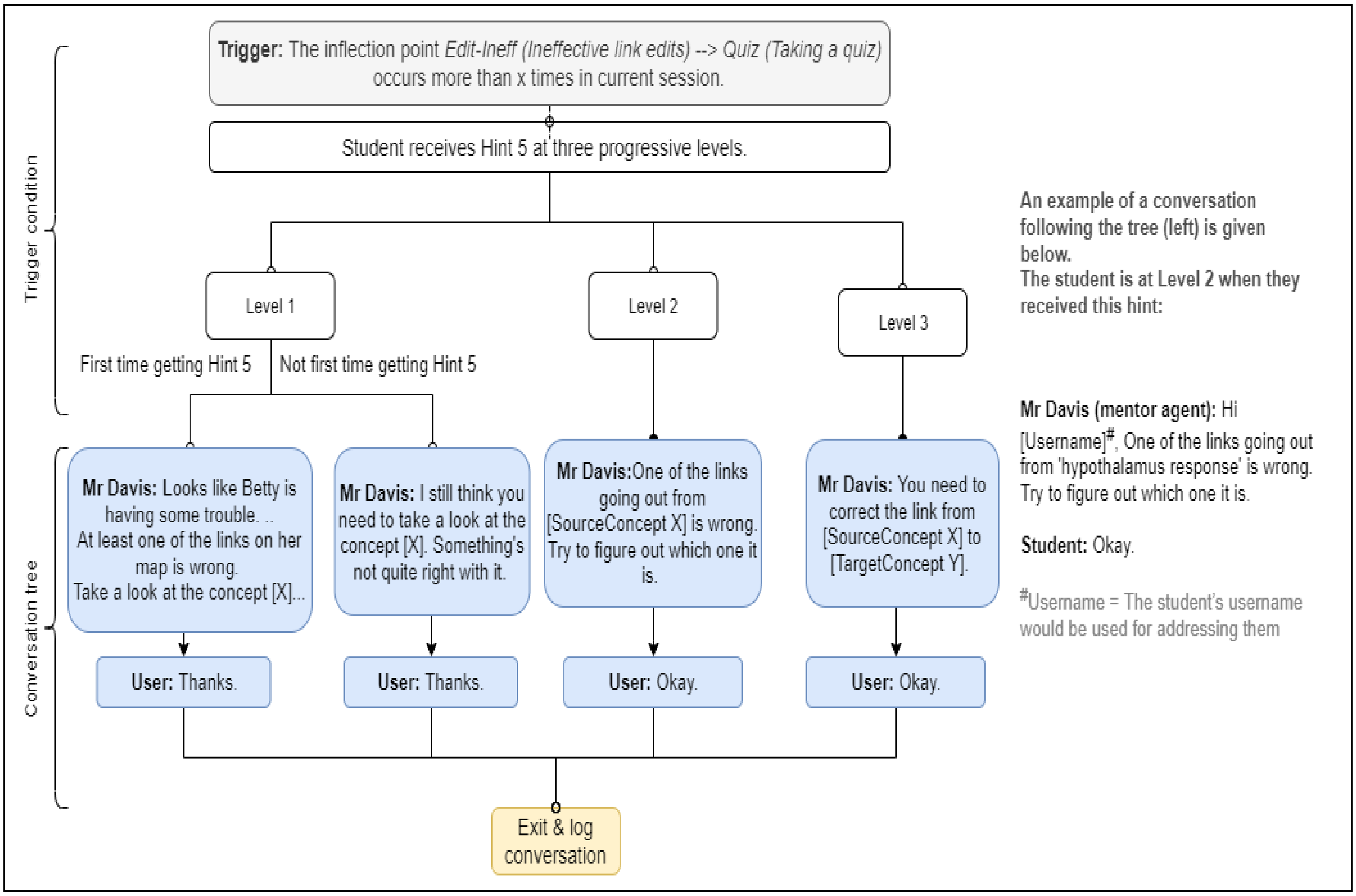
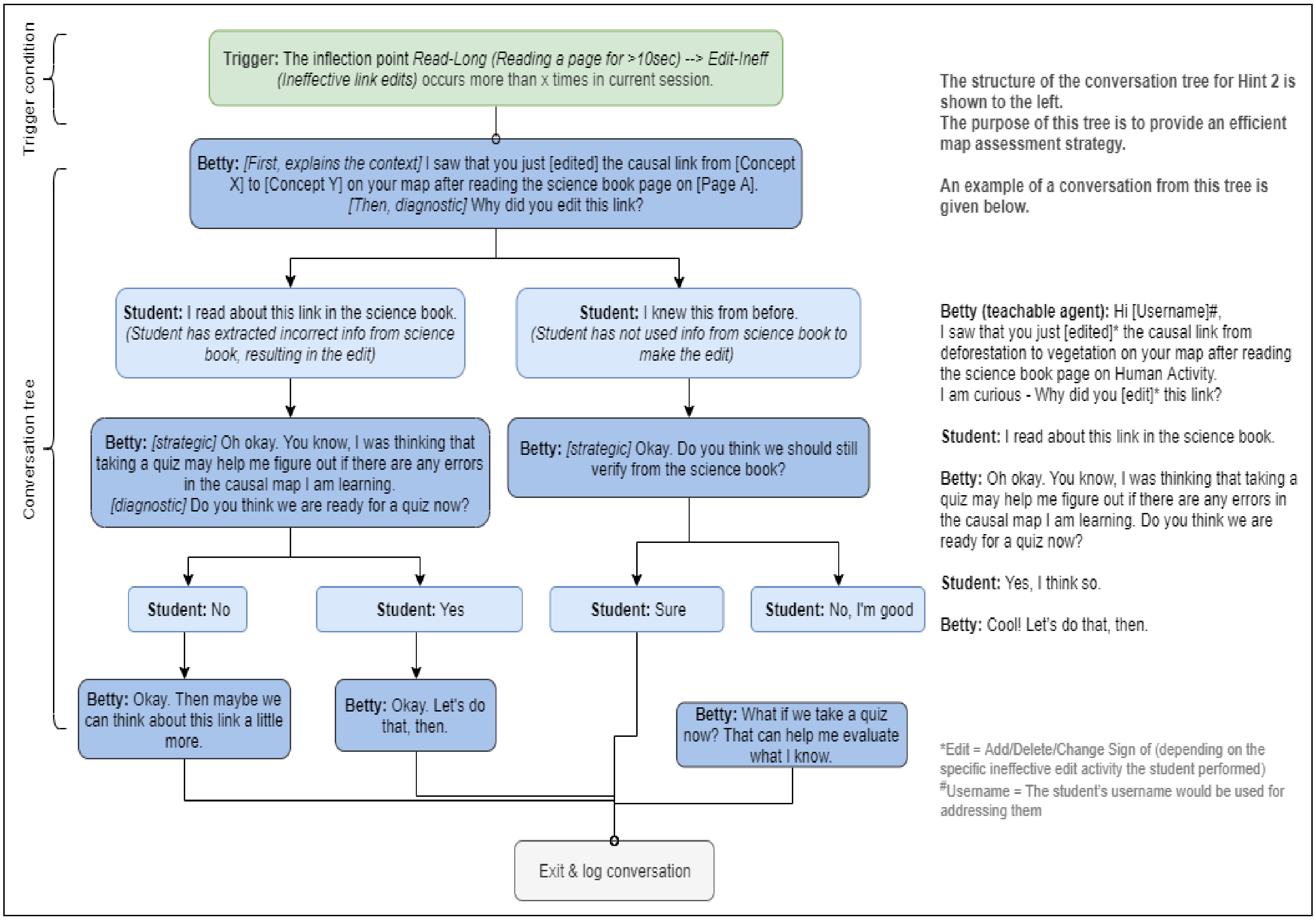

Triggering a Scaffold
Timestamped student data
(Activities w. task context & effectiveness)
Task model
Sequence of activities that corresponds to cognitive (IA / SC / SA) processes
Effectiveness of causal modeling activities
Sliding window of a sequence of [activity : effectiveness] pairs
Detecting transitions or 'inflection points' in cognitive / metacognitive processes
IF count (infl. point) > threshold
An 'inflection point pattern' is detected
(Trigger Condition)
Delivering a Scaffold
Appropriate scaffold (strategic hint / encouragement) for trigger condition is identified
Conversation tree initiated by agent & driven by student response at each step
Conversation ends & [scaffold type, conv tree, trigger] is logged for evaluation

Figure 4



(a) Progression levels of a conversation tree for a map-debugging scaffold by Mr. Davis



(b) Conversation tree for Hint2, a map-assessment scaffold initiated by Betty

Figure 5

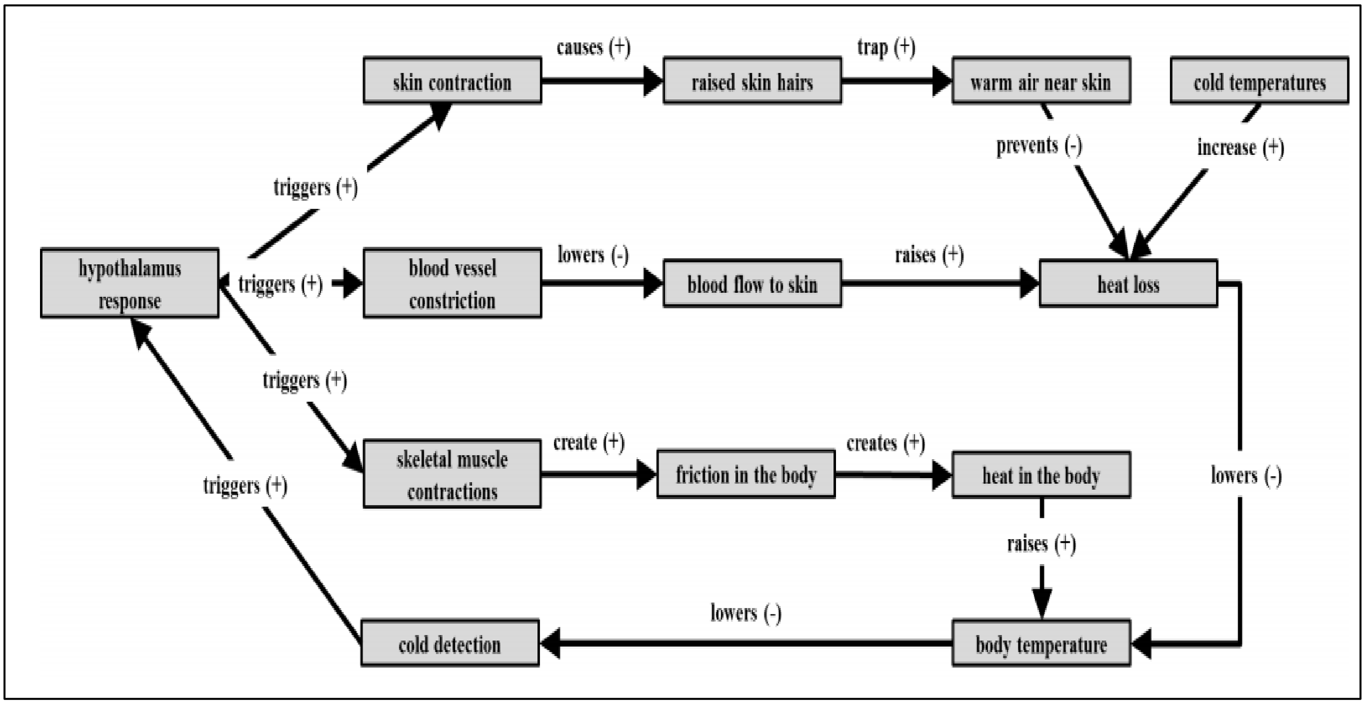
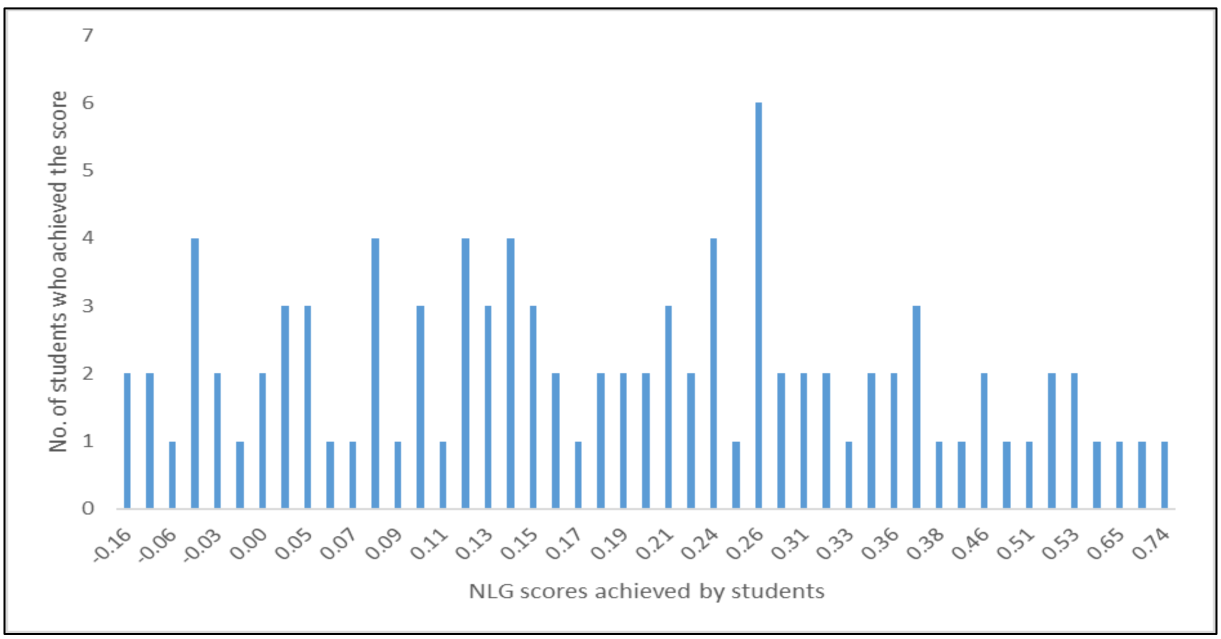


Figure 6



Appendix A

Example questions from the paper-based pre- and post-tests used in the study:

Example of a *Multiple Choice (MC)* question:

Choose the best answer:
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.

Example of a *Short Answer (SA)* question:

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: ________________________

Step 3: ________________________

Step 4: ________________________

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