Can the use of cognitive and metacognitive self-regulated learning strategies be predicted by learners' levels of prior knowledge in hypermedia-learning environments?

Michelle Taub, Roger Azevedo, François Bouchet, Babak Khosravifar

Can the use of cognitive and metacognitive self-regulated learning strategies be predicted by learners' levels of prior knowledge in hypermedia-learning environments?

Michelle Taub [a], Roger Azevedo [a], François Bouchet [b], Babak Khosravifar [c]
[a] Department of Psychology, North Carolina State University, 2310 Stinson Drive, Raleigh, NC 27695-7650, United States
[b] Sorbonne Universités, UPMC Univ. Paris 06, UMR 7606, LIP6, F-75005 Paris, France
[c] Department of Electrical and Computer Engineering, Concordia University, 1616 Ste. Catherine West, EV-3269, Montreal, QC H3G 2W1, Canada

Abstract
Research on self-regulated learning (SRL) in hypermedia-learning environments is a growing area of interest, and prior knowledge can influence how students interact with these systems. One hundred twelve *(*N = 112) undergraduate students' interactions with MetaTutor, a multi-agent, hypermedia-based learning environment, were investigated, including how prior knowledge affected their use of SRL strategies. We expected that students with high prior knowledge would engage in significantly more cognitive and metacognitive SRL strategies, engage in different sequences of SRL strategies, spend more time engaging in SRL processes, and visit more pages that were relevant to their sub-goals than students with low prior knowledge. Results showed significant differences in the total use of SRL strategies between prior knowledge groups, and more specifically, revealed significant differences in the use of each metacognitive strategy (e.g., judgment of learning), but not each cognitive strategy (e.g., taking notes) between prior knowledge groups. Results also revealed different sequences of use of SRL strategies between prior knowledge groups, and that students spent different amounts of time engaging in SRL processes; however, all students visited similar numbers of relevant pages. These results have important implications on designing multi-agent, hypermedia environments; we can design pedagogical agents that adapt to students' learning needs, based on their prior knowledge levels.

Keywords: Metacognition, Prior Knowledge, Self-Regulated Learning, Hypermedia-Learning Environments

1. Introduction

Self-regulated learning, SRL, is an important educational construct that has been shown to be effective for students as they learn and study various subjects (Azevedo, 2005, 2007; Winne & Perry, 2000; Zimmerman & Schunk, 2011). When students self-regulate their learning, they are playing an active role in the learning process by engaging in planning, goal-setting, and other cognitive and metacognitive processes (Azevedo, 2005). Research has found that when students engage in self-regulated learning, they achieve high learning outcomes (Greene & Azevedo, 2010; Azevedo et al., 2013). It is, therefore, important for students to develop and use self-regulated learning skills, such as planning and monitoring, and strategies, such as judgment of learning and summarizing, in order to maximize their learning potential.

Despite increasing evidence for the effectiveness of self-regulated learning on students' learning outcomes (Azevedo et al., 2012a; Azevedo & Feyzi-Behnagh, 2010), research has also revealed that students do not enact these effective SRL strategies during learning (Azevedo, 2005; Azevedo et al., 2012a). Interdisciplinary researchers have been designing and developing computer-based learning environments (CBLEs; e.g., multimedia, hypermedia, intelligent tutoring systems, multi-agent systems) to foster and promote effective self-regulated learning in students as they learn about various topics such as biology, physics, and ecology (Azevedo et al., 2010; Azevedo & Aleven, 2013; Azevedo et al., 2013; Biswas et al., 2010; Graesser et al., 2007; Jonassen & Land, 2012; D'Mello et al., 2013; Lajoie et al., 2013; Lester et al., 2013; Woolf, 2009).

Some CBLEs are agent-based, meaning they are programmed to include one or several pedagogical agents, PAs, who are present to assist students by providing scaffolding and feedback during learning, problem solving, strategy training, and skill acquisition (Azevedo et al., 2012b; Biswas et al., 2010; D'Mello et al., 2013; Graesser & McNamara, 2010; Lester et al., 2013). In addition to content learning, these agents are programmed to assist learning about different aspects of SRL, such as planning, goal-setting, metacognitive monitoring, strategy use, and reflection (see Azevedo & Aleven, 2013; Azevedo et al., 2012a). Research has shown that the use of PAs can be effective for learners because providing students with the appropriate scaffolding can help them to better learn (Kinnebrew et al., 2013) and, more specifically, to self-regulate their learning (Azevedo et al., 2012b Graesser & McNamara, 2010; Winters et al., 2008). The role of prior knowledge is a critical individual differences variable that has not been adequately examined in the context of SRL and learning with multi-agent systems. Results will contribute to theoretical (e.g., understanding the deployment of SRL processes based on prior knowledge) and educational (e.g., providing a dynamic assessment and differential scaffolding based on learners' prior knowledge) implications to SRL, which can assist researchers in designing CBLEs that adapt to student characteristics, such as level of prior knowledge.

The focus of this study is to assess how students' prior knowledge can impact the way they self-regulate their learning in a CBLE, with the assistance of pedagogical agents. Prior knowledge of the domain can greatly affect how students engage in different SRL processes and use learning strategies (Moos & Azevedo, 2008; Shapiro, 2004); thus, when creating these environments, it is important to consider how students' prior knowledge of the domain can potentially influence the SRL skills and strategies (which can be metacognitive or cognitive) that they use. For this study, we acknowledged past findings regarding the importance of prior knowledge, and thus assessed how prior knowledge of the circulatory system influenced how students used cognitive and metacognitive SRL strategies as they learned with MetaTutor, a multi-agent, adaptive hypermedia-learning environment (Azevedo et al., 2010; 2011; 2012a; 2013; Azevedo, 2009).

1.1 Theoretical Framework

In our analyses of self-regulated learning, we view SRL as an event that temporally unfolds in real time (Azevedo et al., 2010; 2013; Winne & Perry, 2000). As our theoretical model of SRL, we used Winne and Hadwin's Information-Processing Model (1998, 2008), according to which learning occurs in four basic phases—definition of the task, setting goals and planning, studying tactics, and adaptations—and information processing occurs within each learning phase (Winne & Hadwin, 1998, 2008). In Phase 1, learners assess the task at hand and determine the environmental factors that are available to help accomplish the task. Phase 2 involves planning and goal-setting, in which learners set goals that are necessary to accomplish the task. In addition, learners plan the appropriate subtasks needed to complete the sub-goals that were set at the beginning of the phase. Phase 3 involves the learners employing the strategies that they planned to engage in during the second phase. In addition, the learners monitor their progress toward achieving the goals that have been set. Lastly, Phase 4 is characterized by a reflection of what was accomplished in Phases 1–3. Learners make appropriate adaptations to plans and goals that were set, which can be based on the learners' modified understanding of the task.

It is expected that students with high prior domain knowledge will progress through each stage differently than students with low prior knowledge of the domain on which the task focuses. In Phase 1 (defining the task), students with high prior knowledge will not differ from students with low prior knowledge in terms of defining what the task is asking of them; however students with high prior knowledge will be more aware of the contextual factors, which can be used as resources in accomplishing the task, compared to students with low prior knowledge, who will have difficulty identifying the appropriate environmental factors. For example, students with high prior knowledge might recall how they approached a similar problem in the past, but students with low prior knowledge will not make this connection between present and past problems. In the second phase (setting goals and planning), students with high prior knowledge will not have difficulties planning or creating the sub-goals necessary to achieve the task; however since students with low prior knowledge are not familiar with the domain, they will experience difficulties in creating the sub-goals needed to accomplish the task. For example, students with high prior knowledge know that if they are learning about the circulatory system, they will need to create a sub-goal that either relates to prior knowledge or goes beyond their prior knowledge (of the circulatory system) and deals specifically with the overall learning goal. Students with low prior knowledge, however, might experience difficulties when creating and prioritizing relevant sub-goals, given their lack of domain knowledge. For Phase 3 (studying tactics), both prior knowledge groups will be able to deploy the strategies that they have set; however students in the high prior knowledge group will differ by deploying more sophisticated and effective use of these strategies. For example, students in both groups might have planned to take notes, but students with high prior knowledge might translate the notes into their own words, whereas students with low prior knowledge might copy the words verbatim from the text. In addition, high prior knowledge students will be able to metacognitively monitor their emerging understanding of the topic more accurately than those with low prior knowledge. Finally, in Phase 4, students with high prior knowledge will be able to reflect on their learning and adjust their understanding of the question, whereas students with low prior knowledge might not be able to make such a reflection. For example, students with high prior knowledge might have planned to spend a particular amount of time achieving a sub-goal; however during reflection, they might realize that they need more time to accomplish another sub-goal. Students with low prior knowledge, however, might not be able to reflect on the time they allotted to completing the sub-goals.

This study focuses on the role of students' prior knowledge on their use of self-regulated learning strategies, and whether we see significant differences between high and low prior knowledge groups. Winne and Hadwin's model (2008) emphasizes the role of prior knowledge as a key factor in self-regulated learning; however, no specific framework or hypotheses have been generated that address the role of prior knowledge in self-regulated learning with CBLEs. Thus, our study provides results that can be used to facilitate the development of a framework for the role of prior knowledge in self-regulated learning with hypermedia (Azevedo et al., 2013).

Based on the existing literature on prior knowledge and Winne and Hadwin's SRL model, we make two assumptions. First, learners with high prior domain knowledge (HPK) will be more effective at self-regulating their learning, compared to students with low prior domain knowledge (LPK), because they have more relevant domain knowledge that allows them to anchor new knowledge to existing knowledge (Mayer, 2004). Second, learners with high prior domain knowledge will be more effective at self-regulating their learning, compared to students with low prior domain knowledge, because they have more working memory (WM) capacity to devote to metacognitive monitoring and the selection of sophisticated cognitive strategies (Paas et al., 2003).

In MetaTutor, several cognitive and metacognitive processes are available for students to engage in as they learn about the circulatory system (Azevedo et al., 2012a). Based on the above assumptions, for our study we have created several new assumptions regarding the role of SRL processes in students with high, compared to low, prior knowledge when using MetaTutor. First, according to cognitive load theory (Paas et al., 2003), students with high prior knowledge will engage in more cognitive and metacognitive processes (combined) during learning. Second, students with high prior domain knowledge of the human circulatory system will engage in more cognitive processes (e.g., take notes [TN], make inferences [INF], create summaries [SUMM], and activate prior knowledge [PKA]). Third, students with high prior knowledge will engage in more metacognitive processes, such as judgment of learning (JOL), feeling of knowing (FOK), content evaluation (CE), and monitoring progress toward goals (MPTG). Based on data mining hypotheses, fourth, we expect that students with high prior knowledge will engage in more effective sequences of use of SRL strategies, which implies that students will engage in metacognitive strategies prior to engaging in cognitive strategies. For example, it is more effective for learners to judge whether they understand the content (i.e., do a JOL) before they take notes (TN) on such content (Azevedo et al., 2011). Students in the low prior knowledge group, however, might not make this differentiation, and will engage in both cognitive and metacognitive strategies when they start a new page or sub-goal. Fifth, students with high prior knowledge will engage in a greater number of SRL strategies and spend more time engaging in these strategies as they work on a sub-goal, compared to students with low prior knowledge, who will spend more time reading the content, engage in fewer SRL strategies, and thus spend less time engaging in SRL strategies. Finally, students with high prior knowledge will visit more pages that are relevant to the sub-goal they are working on, compared to students with low prior knowledge, who will visit more pages, but not just those that are relevant to their sub-goal.

1.2 Literature Review: Prior Knowledge, SRL, and CBLEs

Prior knowledge is an important factor in learning, and thus it is important for researchers to investigate and understand its role in students' self-regulated learning. When students self-regulate, they engage in cognitive, affective, metacognitive, and motivational (CAMM) processes, which can all be influenced by learners' prior knowledge levels, and therefore their individual needs. Consequently, we must adapt our instructional methods in designing hypermedia-learning environments to cater to students' individual learning needs in order to promote effective use of CAMM processes to self-regulate their learning (Azevedo et al., 2012a). Previous research has examined the effect of prior knowledge on learning (Ericsson et al., 2006;

Zimmerman & Schunk, 2011); however there is limited research investigating the role of prior knowledge on self-regulated learning of complex topics with hypermedia environments.

Although there is limited research on the role of prior knowledge in assessing students' self-regulated learning with hypermedia, some researchers have conducted studies that have assessed prior knowledge in multimedia learning environments. Winters and Azevedo (2005) used GenScope, a computer-based learning environment, which teaches students about genetics. Students were paired into dyads, based on their prior knowledge as determined by pretest data. Their results indicated that based on the pre- to post-test data shift, students with low prior knowledge displayed an increase in their understanding of the material, whereas students with high prior knowledge did not present any changes in their understanding of the content. Their analyses of students' verbalizations revealed that students with low prior knowledge relied on others to help them engage in self-regulated learning, compared to students with high prior knowledge, who, during the study, regulated their own learning and provided support for the low prior knowledge students who sought assistance. These results were important in demonstrating that appropriate scaffolds can be provided to foster the use of effective SRL strategies.

Moos and Azevedo (2008) conducted a study where they examined the relationship between prior knowledge and self-regulated learning with hypermedia. They implemented a 40-minute learning session, where the results demonstrated that prior knowledge is significantly related to students' use of self-regulated learning during learning with hypermedia. More specifically, the results revealed that prior knowledge was positively related to monitoring and planning, and was negatively related to strategies. In a subsequent study, Moos and Azevedo (2009) implemented a 30-minute hypermedia task and used a series of science questions to foster conceptual understanding about the human circulatory system. First, they found that self-efficacy and monitoring processes were significantly related. Second, they found that prior domain knowledge and monitoring understanding were significantly related. Finally, through regression analysis they found that the relationship between self-efficacy and hypermedia learning was mediated by students' monitoring levels and by the environment.

Several studies have failed to find a significant effect of prior knowledge on learning outcomes. For example, Shapiro (1999) used interactive overviews (IOs) on ecology to help students meet their learning goals. She based the study on Kintsch's construction integration model (1988), which explains that in order for deep learning to occur, students must integrate this newly learned information with their prior knowledge of the content. However, results demonstrated that IOs were helpful tools for students to achieve their learning goal, even when they did not have any prior knowledge on the subject (Shapiro, 1999). Van Seters et al. (2012) used e-learning materials to demonstrate how students work differently, based on their own characteristics. To determine these characteristics, the authors collected participants' demographic information, and they analyzed participants' motivation, prior knowledge levels, use of learning strategies, and learning paths. Results showed that students did follow different learning paths. Furthermore, there were significant differences among Dutch BSc students and international MSc students in terms of their intrinsic motivation, the learning paths they followed, and the learning strategies they used. However, they found that prior knowledge did not have an effect on students' learning paths.

All of the obtained results provide important implications for designing and testing the effect of prior knowledge on learning outcomes, and it is important for future studies to consider even more dependent variables that can be influenced by prior knowledge in a range of learning environments. In the current study, we examine the specific processes of self-regulated learning, and how students' prior knowledge levels can influence the use of these cognitive and metacognitive strategies.

1.3 Current Study: Overview and Research Questions

In this study, we examined participants' pretest scores (i.e., answers to a 25-item multiple-choice test) to determine whether their prior knowledge of the human circulatory system influenced how they deployed cognitive and metacognitive SRL strategies as they interacted with MetaTutor, a multi-agent hypermedia learning environment, where the goal is to learn as much as possible about the human circulatory system in a 60-minute session. In addition, we used data mining techniques to examine (1) sequential patterns of the use of SRL strategies, and (2) whether there were differences between prior knowledge groups in their use of combined sets of two product variables. We plotted students' use of SRL strategies, the time they spent engaging in SRL strategies, and the number of relevant pages they visited as they worked on various sub-goals. We posed the following research questions: (1) Is there a significant difference in the frequency distribution of learners' use of total, cognitive, or metacognitive self-regulated learning strategies; (2) What are the most frequent sequences of SRL strategies that differentiate between prior knowledge; and (3) are there differences between learners' time and use of SRL strategies, and between learners' use of SRL strategies and visits to relevant sub-goal pages across individual and combined MetaTutor learning sub-goals, based on knowledge groups? Overall, we expected that students in the high prior knowledge group would engage in more cognitive and metacognitive SRL strategies than students in the low prior knowledge group, and would, furthermore, engage in more adaptive uses of SRL strategies (e.g., a metacognitive strategy, followed by a cognitive strategy).

2. Methods

2.1 Participants

One hundred twelve (*N* = 112) participants were selected from a study with one hundred twenty-three undergraduate students at three large universities in North America. Participants' ages ranged from 18–31 years old, with a mean age of 21.08, *SD* = 2.85; 65% were female. Participants were given a monetary compensation of up to $40 dollars for completing the study.

2.2 Research Design

In this study, we used a quasi-experimental design, as for the analyses, participants were assigned to either the low prior knowledge group (LPK, *n* = 56), or the high prior knowledge group (HPK, *n* = 56), based on their pre-test scores they completed on the human circulatory system, during Day 1 of the study. The median pre-test score was 20 out of 25; *M* = 18.87, *SD* = 4.18, and students were assigned to a prior knowledge group based on if their score was higher (i.e., HPK group) or lower (i.e., LPK group) than the median score of 20. For more details on the selection of the groups, see section 2.6: *Coding and Scoring: Product and Process Data*

2.3 Materials

Two equivalent 25-item multiple-choice pretests and posttests developed by Azevedo and colleagues (Azevedo et al., 2012a) were used to assess participants' learning during the 1-hour session with MetaTutor.

We extracted data from the log-files, which captured the students' interactions with the MetaTutor environment. The extracted data consisted of the SRL strategies students engaged in, as well as the time these strategies were initiated and ended, based on their use of the SRL palette. Additionally, we extracted sequences of SRL processes students engaged in from the same log-files. The extracted log-file data were then analyzed with respect to HPK and LPK groups.

2.4 MetaTutor: An Intelligent, Hypermedia Multi-Agent System

MetaTutor is an intelligent, hypermedia-learning environment that engages students in learning about a complex science topic, the circulatory system (Azevedo et al., 2010; 2012a; 2013). The learning environment contains 38 pages of text and diagrams. MetaTutor allows us to collect a wide array of data, including log-file data, eye-tracking data, think-aloud data, electrodermal activity, screen recordings of learner-system interactions, and facial expressions of participants' emotions. We collected these multichannel data from students while they navigated the MetaTutor system and learned about the circulatory system. While participants interact with MetaTutor, four pedagogical agents assist the students in learning by providing the appropriate scaffolding for each participant. Each agent specializes in one particular area of self-regulated learning. Gavin the Guide focuses on directing the participants through the environment. Pam the Planner's role is to aid and emphasize planning, creating relevant sub-goals, and activating prior knowledge. Mary the Monitor specializes in helping the participants to monitor what occurs throughout the learning session by emphasizing the use of monitoring progress toward goals (MPTG), content evaluation (CE), feeling of knowing (FOK), and judgment of learning (JOL). Sam the Strategizer assists in learners' use of effective strategies (e.g., creating good summaries) as they learn in the environment.

This study has two conditions: a *prompt and feedback condition* and a *control condition*. In the prompt and feedback condition, participants were provided with scaffolding from the pedagogical agents, and thus were not required to work independently. In this condition, the pedagogical agents prompted the learners to engage in learning and SRL strategies, such as JOL and summarizing, in addition to learners using these strategies themselves. Furthermore, when participants in the prompt and feedback condition set their sub-goals, Pam provided feedback on their proposed sub-goals, including whether the sub-goal was too broad or too general, and continued to assist the participants in setting the appropriate sub-goals. In the control condition, participants were free to navigate the system without any scaffolding or feedback from any of the pedagogical agents. Learners were not prompted to use any of the learning or SRL strategies that are part of the SRL palette; however they were still able to engage in these strategies if they chose to on their own. In addition, during the sub-goal setting phase, Pam did not provide any feedback, but simply suggested the sub-goal that participants should choose, which they could decide whether or not to accept; and if they rejected Pam's suggestion, Pam asked them to repeat the sub-goal, and would suggest another sub-goal for them to set. Therefore, in this condition, Pam simply generated a sub-goal for the participant, which required less effort and scaffolding. The participants navigated through the same environment, they were provided with the same instructions, they were presented with the same instructional videos, and they used the same multimedia learning content. The only difference between the two conditions was the scaffolding and feedback from the pedagogical agents, such that there was no assistance given from the agents to the participants in the control group. These two conditions were created to examine the usefulness and effectiveness of pedagogical agents in scaffolding participants as they learned about a complex science topic.

The system interface includes several elements designed to detect, track, model, support, and foster self-regulated learning (see Figure 1). On the left-hand side, we find the table of contents, which displays the title of each of the 38 pages organized by sections. A clock located above the table of contents informs participants of the time remaining in the learning session.

This allows for participants to monitor their time; monitoring progress toward goals (MPTG) is an important SRL strategy, which keeps learners metacognitively aware of where and how they are allotting their learning time. On the right pane of the interface, there is the SRL palette. This tool allows participants to engage in self-regulatory learning strategies, such as taking notes, JOL, FOK, CE, creating summaries, and making inferences. Participants can use any of these SRL or learning strategies at any point throughout the session, and they can be either self- or agent-initiated, such that participants can choose to engage in the learning or SRL strategies located on the SRL palette by clicking on it, or the agent can prompt the participants to engage in such strategies (based on a set of preset rules). The pedagogical agent is placed just above the SRL palette, in the top right corner of the interface. Only one agent is displayed at a time, and thus where the learner is in the session, the learner's previous actions (e.g., metacognitive judgments regarding the relevancy of a particular page and corresponding diagram), and which learning strategies the system will provide scaffolding for will determine which agent is present on the screen, and this changes throughout the session. The sub-goals set by the learner are located on the top center of the interface. The current sub-goal is the one on the top, to remind the learners what they are working on at that moment. As learners progress through a sub-goal, a colored bar will display the amount of progress made toward completing the sub-goal. Furthermore, participants can choose to complete a sub-goal at any time; learners must then complete a quiz on the content to ensure that they have learned an adequate amount of information to, in fact, complete this sub-goal and move on to another one. At the bottom of the screen is a textbox where participants enter their sub-goals, and write down their prior knowledge of the circulatory system and of each particular sub-goal. This box is also used when participants choose to engage in certain SRL processes, such as JOL, FOK, and CE, where they can select a response to questions such as “how well do you think you understand the material presented on this page,” and where multiple options are given to respond with. Above that box, another larger non-editable textbox (not visible on Figure 1) is used to display all the previous interactions between the participant and all the pedagogical agents, to allow him/her to re-read any previous advice or comment from the agents. Finally, the middle of the interface contains the text and diagrams, which are the materials needed to accomplish the overall goal of learning everything about the circulatory system. Each page from the table of contents is associated with a text and a single diagram**.**

2.5 Experimental Procedure

Participants had to be available for two sessions, which had to occur within three days of each other. The first session took approximately 1 hour, and the second session lasted up to 3 hours. During the first session, participants completed a consent form and were then given an explanation of the study. Participants began the experiment by completing a series of self-report questionnaires, which measured demographic information and their emotions (e.g., AEQ; Pekrun et al., 2011). Participants then completed a 25-item multiple-choice pretest to assess their prior knowledge of the human circulatory system. At the end of the session they were paid $5 for completing Session 1.

In the second session, participants began by creating sub-goals with the assistance of Pam. In MetaTutor, there are seven predetermined sub-goals based on different aspects of the circulatory system, which the pedagogical agents are programmed to recognize and lead the participants to set. The seven sub-goals are: (1) Path of Blood Flow, (2) Heartbeat, (3) Heart Components, (4) Blood Vessels, (5) Blood Components, (6) Purposes of the Circulatory System, and (7) Malfunctions of the Circulatory System. Once two sub-goals had been set, the participants were presented with multiple videos that introduced the system, including all of the interface elements and how to engage in self-regulated learning strategies, which are crucial in helping a student learn about the circulatory system (Azevedo et al., 2010; 2012a). Following the introduction to the system, Pam, the pedagogical agent, asked participants to recall all they knew about the circulatory system by writing it all down in the textbox. Participants could write as much or as little as they knew, and thus we expected students with a higher prior knowledge of the circulatory system to have more to present, which could then be referred back to in the log-file data, which captured everything that was written into the system. Next, participants chose the sub-goal to begin working on, and Pam asked them again to mark down everything they already knew about the given sub-goal. Finally, participants began learning with the system by freely navigating to the pages they wished to; they were able to engage in the SRL strategies at any point during the session by selecting the strategy they felt would be most useful from the SRL palette. Throughout the 1-hour session, four more self-report questionnaires measuring students' emotions were presented every 14 minutes, for a total of four times during the session, and had to be completed by the participants before they could continue learning. The repeated administration of questionnaires was used to assess fluctuations in emotions as they progressed during the learning session with MetaTutor. When participants completed all of their original sub-goals, they were given the possibility to keep learning without a sub-goal or to set up a new one. The same thing happened if they finished the additional sub-goals. Once the 1-hour learning session was up, they were then presented with a 25-item post-test on the human circulatory system.

During learning with MetaTutor, several multichannel data were collected including log-files, concurrent think-aloud protocols, electrodermal activity (EDA), facial expressions, eye-tracking, and audio recordings about each participant's self-regulated behaviors. Participants were then debriefed and paid $40 for completing the study. For this study, we only extracted and analyzed log-file data.

2.6 Coding and Scoring: Product and Process Data

In order to determine learners' prior knowledge levels, we conducted a median split on pretest scores, such that scores that fell below the median were categorized as low prior knowledge (LPK) and those who scored above the median were classified as high prior knowledge (HPK). For this data set, the median score was 20 (out of 25, or 80%). Therefore, scores in the HPK group ranged from 21–25 (M = 22.48, SD = 1.22), and scores in the LPK group ranged from 9–19 (M = 15.25, SD = 2.68). Time-stamped, log-file data were used to extract the frequency of use of SRL strategies and the time spent engaging in these SRL processes during the experimental session.

We used data mining techniques to determine the sequences of SRL strategies that were deployed by a subset of the participants (n = 52) during the session and compared strategies between prior knowledge groups. A smaller subset was used to make the data more comprehensive for analysis, and to limit the amount of data points obtained, for we extracted 100,000 sequences of SRL strategies from this smaller subset. More specifically, we used the following data mining techniques (Hegland, 2001; Pujari, 2001) to seek patterns in the data set: (1) anomaly detection, to identify the unusual data; (2) clustering, to structure the data; (3) regression, to determine the functions to model the data; and (4) summarization, to provide comprehensive representations of the results. The data were mapped, based on the generated hypotheses, to determine the influence of the SRL activities with their sequences on HPK and LPK groups.

Furthermore, we extracted data points from three variables from the same subset of participants—number of SRL processes deployed, time spent engaging in SRL processes, and page relevancy—which were used as features in our data set. Each data set, therefore, contained three features: (1) number of SRL processes, (2) time spent on SRL processes, and (3) a binary variable for page relevancy, which were extracted from the log-files, divided, and plotted by prior knowledge group. We were then able to define the decision boundary that could formally classify these data points with respect to its prior knowledge level. Once more, we only used a subset of the data because analyses would have been difficult if it involved over one hundred data plots.

3. Results

To investigate the differences between prior knowledge groups and their use of SRL processes during learning with MetaTutor, based on clicks made on the SRL palette, we conducted several analyses to test for differences between each group and among each SRL strategy. The strategies were categorized as cognitive or metacognitive, which yielded two sets of multiple and parallel analyses. We performed chi-squares to examine the differences in frequency distributions of these strategies among high and low prior knowledge groups.

First, we conducted a chi-square analysis to determine the difference in total SRL strategy-use between prior knowledge groups. Next, we performed an additional chi-square analysis to examine the differences between each cognitive strategy among prior knowledge groups. We then performed the same chi-square analysis for the metacognitive strategies, by comparing the frequencies of use of each metacognitive strategy among prior knowledge groups. The following section describes all of the analyses conducted, along with their results and illustrative representations of the results obtained.

3.1 Research Question 1: Is There a Significant Difference in the Frequency Distribution of Learners' Use of Total, Cognitive, or Metacognitive Self-Regulated Learning Strategies?

A chi-square analysis was performed to determine whether there was a significant difference between HPK and LPK learners' use of the total (i.e., cognitive and metacognitive) SRL strategies. We extracted and summed the frequencies of use of SRL strategies from the learning session log-files.

A chi-square test of independence revealed significant differences in the frequency distribution of learners' use of SRL strategies across prior knowledge groups: $\chi^2(1) = 4.66$, $p = .03$. See Figure 2 for an illustration of the frequency distributions of SRL strategies between prior knowledge groups. Based on these results, further analyses were performed, which differentiated between cognitive and metacognitive SRL strategies, and are addressed in the following research questions.

Chi-square analyses were performed to determine whether there were significant differences in the distribution of learners' use of cognitive and metacognitive strategies across prior knowledge groups. We extracted the frequencies from the log-files by analyzing and enumerating the uses of *cognitive* strategies, TN, INF, and SUMM, and PKA, and *metacognitive* strategies, JOL, FOK, CE, and MPTG.

In order to eliminate potential for type I error, we performed two subsequent chi-square analyses, which investigated each SRL processes, as opposed to comparing the total uses of these strategies. The 2×4 and 2×5 chi-square analyses revealed that there were no significant differences in the distributions of different cognitive SRL strategies; $\chi^2(3) = 2.24$, $p = .52$, however there results revealed that there were significant differences in the distributions of metacognitive SRL strategies across prior knowledge groups: $\chi^2(3) = 10.19$, $p = .02$ for metacognitive SRL strategies. The frequencies for each cognitive and metacognitive strategy, based on prior knowledge group, are illustrated in Figures 3 and 4, respectively.

3.2 Research Question 2: What Are the Most Frequent Sequences of SRL Strategies that Differentiate between Prior Knowledge?

To answer this research question, we used data mining techniques, with a subset of the total sample, including anomaly detection to identify unusual data, clustering into two groups, HPK and LPK, to structure the data, regression to find functions that can model the data, and summarization to provide compact representation of the results to seek patterns in the data set (Berkhin, 2006; Hegland, 2001). We extracted the most commonly used quintet sequences of cognitive and metacognitive SRL strategies throughout the learning session, and grouped them based on levels of prior knowledge. We extracted sets of five sequences because they represent an adequate time-scale of learner-system interactions from which to interpret SRL behaviors during learning with MetaTutor. The goal for a pedagogical agent is to scaffold students to learn in the most efficient way, in order to accomplish sub-goals and engage in effective cognitive and metacognitive processes. The agent does this by assessing students' performance as they interact with the system. If the agents are programmed to assess how the students engage in sequences of SRL processes, it might not be beneficial to assess a large sequence because students could benefit from agent scaffolds and feedback sooner in the session. Therefore, we used five sequences of SRL strategies to improve the agents' planning capabilities, so they can investigate the most effective patterns to guide students of different prior knowledge levels.

3.2.1 Most frequently deployed SRL sequences.

We used data mining techniques to examine over 100,000 quintet sequences of SRL strategies collected during the entire learning session from each of the 52 HPK and LPK participants. Results demonstrated that some sequences were engaged in 44 times, while others were engaged in one or zero times. For example, HPK learners engaged in the sequence of: PLAN-PKA-JOL-MPTG-SUMM 44 times, which was the most frequent sequence across the HPK sample. Next, HPK learners engaged in PLAN-PKA-JOL-SUMM-TN 43 times, followed by PLAN-PKA-JOL-MPTG-CE and PLAN-PKA-SUMM-TN-MPTG, which had both been engaged in 42 times. It is evident, therefore, that HPK students most frequently engaged in PLAN, then PKA, as the first two SRL strategies. In the four most frequent quintet sequences of SRL strategies used by HPK students, they began with these two SRL strategies. The most frequent sequence, PLAN-PKA-JOL-MPTG-SUMM, involved students creating summaries after monitoring their progress toward goals, which is an effective sequence of strategies, since they decided to create summaries after they planned, activated their prior knowledge of the content, judged whether or not they understood what they were learning, and monitored how far along they were to completing their sub-goals. In the second most common sequence, PLAN-PKA-JOL-SUMM-TN, learners created summaries and took notes after they planned, activated their prior knowledge, and judged whether or not the content was relevant, which is also an effective sequence since they metacognitively planned and monitored their progress, and then engaged in the appropriate cognitive learning strategies. The students in the LPK group differed in the sequences of SRL strategies they used; however, they also used some similar sequences. Like those in the HPK group, the four most frequent quintet sequences of SRL strategies used by LPK students began with planning, followed by PKA. The two most common sequences were PLAN-PKA-SUMM-TN-FOK and PLAN-PKA-SUMM-TN-JOL, which were both used 42 times. The next two most frequent quintet sequences used by LPK students were PLAN-PKA-JOL-SUMM-TN and PLAN-PKA-SUMM-TN-MPTG, which were both used 41 times. These sequences were the same as those used by HPK participants, which demonstrates that although we do see differences in the sequences of use of SRL strategies between prior knowledge groups, LPK participants are not all necessarily ineffective in using SRL strategies. Previous results indicated no significant differences among the frequencies of use of some SRL strategies, and so it is not surprising to find results that demonstrate that students in both prior knowledge groups engage in similar sequences of SRL strategies.

Overall, the data mining sequences have shown that HPK students engage in different sequences of SRL strategies than LPK students. However, we also saw some similarities in the sequences used by both prior knowledge groups, which is appropriate since our previous results demonstrated some significant and non-significant differences in the frequencies of use of SRL strategies between prior knowledge groups.

3.3 Research Question 3: Are There Differences between Learners' Time and Use of SRL Strategies, and between Learners' Use of SRL Strategies and Visits to Relevant Sub-Goal Pages across Individual and Combined MetaTutor Learning Sub-Goals, Based on Knowledge Groups?

To address this research question, we used the following data mining techniques (Berkhin, 2006; Hegland, 2001; Pujari, 2001): clustering, regression, and summarization, which combined sets of product variables, such as time engaging in SRL processes, in order to plot individual data sets from a subset of the total sample (n = 52), which were categorized by sub-goal. The values for these variables were obtained by extracting the information from the log-files. In MetaTutor, an SRL activity is considered as an event, and thus we extracted the time students spent on each event. The variables used for analyses in this research question were the number of SRL processes engaged in and time spent engaging in SRL processes. HPK students are shown as blue dots, and LPK students are shown as red dots. Figure 5 displays the plots of different prior knowledge students' data points with respect to the number of SRL processes used and the time spent engaging in SRL processes for each of the seven MetaTutor sub-goals (see section 2.5), and for all of the seven sub-goals combined, for a total of eight data plots. The blue and red dots are scaled normalized values of the parameters, which were obtained from the formula, $x_n = (x - M)/SD$, where x_n represents the normalized score, x represents the data plot score, *M* is the mean for all the data plots, and *SD* is the standard deviation. To apply this equation, we normalized the data points, and we then mapped them into an interval of [0,1].

For the *Path of Blood Flow* MetaTutor sub-goal, HPK students engaged in a greater number of SRL processes and spent more time engaging in SRL processes compared to LPK students. For the *Heartbeat* sub-goal, in general, neither group engaged in as many SRL processes as used in other sub-goals, and therefore did not spend much time engaging in SRL processes. For *Heart Components,* students in the HPK group spent more time engaging in SRL processes and they engaged in more SRL processes. According to Figure 6, few students in the LPK group selected *Heart Components* as the sub-goal, or they did not engage in any SRL processes while completing this sub-goal. For the *Blood Vessels* sub-goal, LPK students engaged in more SRL processes, and the students who spent more time engaging in SRL processes also engaged in a larger number of SRL processes. For *Blood Components*, HPK students engaged in more SRL strategies, and thus spent more time engaging in SRL strategies than LPK students, which implies that more HPK students set *Blood Components* as a sub-goal and engaged in SRL strategies compared to LPK students, who rarely set *Blood Components* as a sub-goal, and those who did used fewer SRL strategies than HPK students. We see a similar pattern with *Purposes of the Circulatory System*, which had more HPK students, and those that engaged in more SRL strategies also spent more time engaging in SRL processes throughout the session. Many LPK students engage in only a few SRL processes and thus did not spend much time engaging in SRL processes, with the exception of a select few, who overall did spend more time engaging in SRL processes if they engaged in more SRL processes.

Finally, more HPK students set the sub-goal of *Malfunctions of the Circulatory System* during the second session of the experiment than LPK students. Furthermore, the HPK students who spent more time engaging in SRL processes were the ones who also engaged in more SRL processes for the sub-goal. The few LPK students who used SRL strategies used a small number of SRL processes, and thus did not spend much time engaging in SRL processes. From these results, we can infer that LPK students either did not engage in SRL processes as they attempted to complete this sub-goal, or did not set *Malfunctions of the Circulatory System* as one of their sub-goals in the learning session. Overall, the results from this analysis revealed apparent differences in the sub-goals that were set by HPK students, compared to LPK students, and that we see different uses and durations of uses in engaging in SRL processes depending on the sub-goal that is being attended to. *Heart Components*, *Purposes of the Circulatory System*, and *Malfunctions of the Circulatory System* are the sub-goals where HPK students engaged in a larger number of SRL processes compared to the other four sub-goals, *Path of Blood Flow*, *Heartbeat*, *Blood Vessels*, and *Blood Components*. Overall, LPK students use fewer SRL processes during the learning session than HPK students.

Another pair of product variables that we assessed by individual data plots was the number of SRL processes with page relevancy. Refer to section 3.3 for the details in extracting the number of processes and time spent on SRL variables from the log-files. To extract the page relevancy data, the log-file records the pages that the participants read, as well as the sub-goals that they set during learning. Moreover, we predetermined which pages were relevant to each sub-goal, and thus calculated the ratio using the current sub-goal and the relevant pages read when accomplishing the sub-goal. See Figure 6 for plots of SRL processes with page relevancy for individual sub-goals and all sub-goals collectively.

Overall, students engaged in a varying number of SRL processes while visiting relatively similar numbers of relevant pages. More specifically, regardless of what sub-goal the students were working on, they navigated to the same number of pages that were relevant to their current sub-goal. However, students tended to use differing numbers of SRL processes depending on the current sub-goal. Both HPK and LPK students used lower numbers of SRL processes for *Path of Blood Flow* and *Heartbeat*, and most students used more SRL processes for *Heart Components*, *Blood Vessels*, *Blood Components*, *Purposes of the Circulatory System*, and *Malfunctions of the Circulatory System*. *Path of Blood Flow* and *Heartbeat* can be seen as easier sub-goals, and so we might assume that fewer SRL processes would be needed to complete them.

4. Discussion

The results from this study demonstrated how low– and high–prior knowledge students used cognitive and metacognitive SRL strategies as they learned about the human circulatory system in a 60-minute session with MetaTutor, a multi-agent intelligent hypermedia system. More specifically, we investigated how students' prior knowledge of the circulatory system affected how they used different learning strategies, such as taking notes, prior knowledge activation, judgment of learning, feeling of knowing, and others. Results indicated that prior knowledge groups significantly differed in their use of total cognitive and metacognitive SRL processes; and more specifically, results revealed significant differences in each metacognitive SRL strategy; however prior knowledge groups did not significantly differ in their frequencies of use of each cognitive SRL strategy. Furthermore, prior knowledge groups differed in their sequences of use of SRL strategies and their engagement in SRL strategies as they worked on particular MetaTutor sub-goals. Students did not differ in the amount of pages they visited that were relevant to the sub-goals they were working on. The following sections will address the specific results obtained, based on each research question.

Research Question 1: Is there a difference in the frequency distribution of learners' use of total, cognitive, or metacognitive self-regulated learning strategies? This question addressed the overall frequencies of use of self-regulated learning strategies, such that it determined if students with different levels of prior knowledge differed in their use of all self-regulated learning strategies as they learned with the MetaTutor environment. Results demonstrated that the prior knowledge groups differed significantly in the total use of SRL strategies. Furthermore, these results support the majority of findings in other prior knowledge research, such as that by Winters and Azevedo (2005) and Moos and Azevedo (2008, 2009), who all found a significant effect of prior knowledge on student learning. Thus, these findings further emphasize the importance of prior knowledge in learning, and how we should consider students' prior knowledge levels when designing hypermedia-learning environments.

This question also addressed the use of cognitive and metacognitive SRL strategies among prior knowledge groups by using chi-square analyses to determine if there were significant differences between HPK and LPK groups in their frequency distributions of taking notes, making inferences, creating summaries, and activating prior knowledge, and if there were significant differences in the use of the individual metacognitive strategies: judgment of learning, feeling of knowing, content evaluation, and monitoring progress toward goals among prior knowledge groups as they engaged in learning with MetaTutor.

Results indicated no significant differences for cognitive strategies, such that prior knowledge groups did not differ based on their use of the specific cognitive strategies (e.g., TN, INF, SUMM and PKA), which can be used during the MetaTutor learning session. These results, therefore, do not support the hypothesis that HPK students would engage in more cognitive SRL strategies than LPK students, and thus support Shapiro (1999) and van Seters et al.'s (2012) findings that prior knowledge did not influence students' learning goals and did not affect students' learning paths. Thus, when creating environments that adapt to prior knowledge levels, researchers should focus on aspects of SRL other than cognitive strategy use.

Further results demonstrated that there were significant differences in the use of each metacognitive strategy (e.g., JOL, FOK, CE, and MPTG). This partially supports the initial hypotheses, which stated that HPK students would engage in more metacognitive processes than LPK students, because although HPK students engaged in significantly more JOL, CE, and MPTG than LPK students, they engaged in significantly fewer FOK than LPK students. These findings contribute to research in SRL because they emphasize the level of granularity at which SRL processes are coded and analyzed.

Research Question 2: What are the most frequent sequences of SRL strategies that differentiate between prior knowledge groups? This research question assessed the most commonly used quintet sequences of SRL strategies, and whether HPK and LPK students differed in their most frequent use of these strategies. The most frequent sequence for HPK students was PLAN-PKA-JOL-SUMM-TN, while LPK students most frequently engaged in PLAN-PKA-SUMM-TN-FOK. Thus, HPK and LPK students most often engaged in different sequences of SRL processes. HPK students engaged in metacognitive strategies before cognitive strategies because they were more focused on monitoring what they knew from what they did not know, and this requires metacognitive knowledge and skills; they also had more working memory capacity to allocate to metacognitive monitoring processes. LPK students engaged in cognitive strategies before metacognitive strategies because they were focused on learning the material, therefore using more cognitive strategies, which supports the hypothesis that HPK students would be able to engage in more effective uses of SRL processes (i.e., metacognitive prior to cognitive) than LPK students. According to previously mentioned analyses (Research Question 1), there were some significant differences among prior knowledge groups in the frequency of use of metacognitive SRL processes, and so these results were expected. We did, however, discover sequences that were similar among both knowledge groups, and this can be attributed to the results from Research Question 1, which indicated that there were no significant differences in the uses of cognitive strategies among prior knowledge groups. These results support the findings made by Winters and Azevedo (2005) and Moos and Azevedo (2008, 2009), who found a significant effect of prior knowledge on learning; however the results obtained by Shapiro (1999) and van Seters et al. (2012) were also supported, since we sometimes did not find differences in learning among prior knowledge groups. These results should encourage researchers to seek where students with varying prior knowledge levels differ when interacting with hypermedia-learning environments, in order to design the most effective CBLEs to promote learning in students with all levels of prior knowledge.

Research Question 3: Are there differences between learners' time and use of SRL strategies, and between learners' use of SRL strategies and visits to relevant sub-goal pages across individual and combined MetaTutor learning sub-goals, based on knowledge groups? Results indicated that duration of and use of SRL processes were lower for LPK students compared to HPK students; however, the results also depended on the sub-goal that was being worked on. HPK students set *Heart Components*, *Purposes of the Circulatory System*, and *Malfunctions of the Circulatory System* as their sub-goals more frequently than the other sub-goals. It was determined, therefore, that few students with high prior knowledge set *Path of Blood Flow*, *Heartbeat*, *Blood Vessels,* or *Blood Components* as a sub-goal, and those who did set those sub-goals did not engage in many SRL processes or spend much time engaging in SRL processes. These results support the hypothesis that HPK students would engage in and spend more time engaging in more SRL processes than LPK students. It should be noted, however, that HPK students and LPK students at times engaged in different sub-goals, and so these results can influence the way we program multi-agent systems to adapt to students' prior knowledge levels by providing different sub-goals for them to work on.

Results also indicated that regardless of the sub-goal students were working on, all students visited the same number of relevant pages. We did see differences, however, in the number of SRL processes students used for the different sub-goals. Both HPK and LPK students engaged in fewer SRL strategies for *Path of Blood Flow* and *Heartbeat*, and engaged in more SRL processes for the other sub-goals. Thus, we can assume that these are easier sub-goals that require fewer SRL strategies. We did not see many differences between knowledge groups, which is appropriate based on previous findings that did not find significant differences in the frequency distributions of cognitive strategies between prior knowledge groups. These results emphasize the importance of setting sub-goals when working in a hypermedia-learning environment, and that we need to adapt learning environments to contain the appropriate sub-goals for students based on their individual learning needs.

4.1 Limitations of the Study

There were several limitations to this study. We determined prior knowledge based on pretest scores, which were measured by students' performance on a multiple-choice test. In addition, we assumed that HPK students were better self-regulators than LPK students because of their familiarity with the content; however we did not measure students' self-regulatory knowledge and skills. Moreover, for the data mining analyses, we used a subset of participants from a larger study based on a median-split, and therefore the results are sample-specific to the subset included in this study. In addition, the participants in the study were in both experimental conditions (i.e., prompt and feedback, and control), which differ based on the prompts given by the pedagogical agents. Participants in the prompt and feedback condition received prompts from the agents to engage in a number of SRL strategies, whereas the agents did not prompt students in the control condition. Therefore, students in the prompt and feedback condition, regardless of their prior knowledge, might have used higher frequencies of SRL strategies throughout the learning session because they were instructed to do so. Finally, this study only included data obtained from participants' log-files, which limited the data mining analyses that generated the qualitative results. Our findings are also limited because participants might have engaged in SRL processes that were covert, and thus were not captured from the log-files. For example, HPK students might have engaged in more PKAs and FOKs, but they might have done so out loud, and did not select the option to do so on the SRL palette. The log-file data, therefore, would not have captured the use of these strategies, and we would have to converge other trace data (e.g., screen recording of learner-system interaction) to observe this. It is important, therefore, to use multichannel data for analyses, which would give us greater understanding of what the students were doing during the learning session.

Furthermore, we did not measure how the sequences were influential in learning. For example, data mining analysis informed us of the most commonly used quintet sequences for HPK and LPK groups; however we did not measure the impact these sequences made on students' post-test scores. Thus, we cannot determine the influence of the sequences on learning. Moreover, when we analyzed the sequences of SRL strategies that students used during learning, we did so on a global level (i.e., the entire learning session), and did not limit the analysis of sequences to a particular page. Lastly, the data mining analyses were descriptive and qualitative; therefore we did not quantitatively compare differences in engagement of SRL sequences between prior knowledge groups.

4.2 Future Directions and Educational Implications

The results from this study stimulate many future directions for analyses and design of hypermedia environments, stressing the importance of assessing and accommodating prior knowledge groups. Future analyses on prior knowledge will account for the limitations in this study. Future studies will improve our assessment of prior knowledge by determining prior knowledge levels with more reliable methods, such as evaluating previous school test scores or assessing the students during a period prior to the learning session. Future studies might also include assessing self-regulatory knowledge and skills and their impact of SRL behaviors, performance, and learning. Methodologically, studies on prior knowledge should include larger samples, which will be better obtained if a median split is not performed, and which will allow for better generalization of obtained results. Additionally, future studies will include participants in the same experimental condition. The pedagogical agents do not prompt participants in the control condition, and so a measure of students' use of SRL processes in the control condition (and not the prompt and feedback condition) will be solely based on what the students initiate during learning. It will be beneficial, therefore, to continue to collect data and to analyze the results both within and by condition, such that we can compare results among participants in the feedback condition only (or the control condition only), and we can also compare results between the feedback condition and the control condition. Analytically, future studies will incorporate analyses of multichannel data including eye-tracking, physiological, audio, and video data, which will allow us to gain a greater understanding of what students are doing at each moment of the learning session (Azevedo et al., 2013). Furthermore, future studies will expand on the data mining analyses that were performed in this study, incorporate the use of quantitative data, and assess student performance throughout and following the learning session.

The data mining analysis was an insightful preliminary analysis of how we can detect patterns of use of cognitive and metacognitive SRL strategies, and how these patterns might be different between prior knowledge groups (Bouchet, Harley, Trevors, & Azevedo, 2013; Bouchet, Kinnebrew, Biswas, & Azevedo, 2012). Future studies can further assess these patterns, while including additional variables to be measured during the learning session. More specifically, it can be beneficial to use additional data mining techniques to examine the influence of the most and least commonly used strategies on learning, such that we can include post-test scores in our analysis (e.g., Kinnebrew et al., 2013). Such analyses can be conducted at more local levels, such as the page level, in order to determine the sequences of SRL strategies used on particular pages, and whether or not this differs by prior knowledge group. Finally, future studies should expand on the qualitative analysis to include quantitative data to compare the sequences in order to investigate for significant differences between sequences of SRL strategies used between prior knowledge groups.

Additionally, this study mentioned the concept of cognitive load in generating hypotheses, although this construct was not measured in this study. It can be proposed that LPK students might use fewer SRL strategies based on cognitive overload. Possible reasons for this issue could be that these students feel they have to compensate for their lack of knowledge on the content by engaging in many SRL cognitive strategies. Future studies could assess cognitive load in students during learning, including assessing the correlation between evidence of cognitive load (e.g., pupil dilation) in students and the number of SRL processes these students engage in, in order to test the proposed hypothesis along with many others. We would address such issues as: (1) how would we define cognitive load; (2) how would we measure cognitive load; (3) when does cognitive overload appear in students; and (4) how can we design agents to determine how and when to help students who are experiencing cognitive overload during learning. Such analyses will require real-time analysis of student performance, which we hope to make available in newer versions of MetaTutor.

The findings from this study will help us design multi-agent systems with pedagogical agents that can adapt their decision making for students based on the students' levels of prior knowledge. For example, the results demonstrated that HPK students engaged in some sub-goals more frequently than LPK students, and LPK students engaged in other sub-goals more frequently than HPK students. Pedagogical agents can be designed to assign sub-goals to students based on their prior knowledge of the content. Furthermore, agents can be designed to monitor student performance in real time, which will allow agents to provide scaffolding to students at times when they appear to be having difficulties, or if students might be engaging in maladaptive SRL strategies (Azevedo & Feyzi-Behnagh, 2010). It can be beneficial to design pedagogical agents that can adapt to students' individual differences, which will allow for the most optimal learning environment and can cater to each student's individual learning needs.

It was noted in the limitations section that SRL strategies, which are considered events in MetaTutor, might overlap with other events (i.e., might occur in parallel), or might be overestimated in the time of use (e.g., if students take notes and then do an MPTG, we are only informed of the time when the students start taking notes and start engaging in the MPTG; and so it seems as though the participants took notes from the start time until they started the MPTG, even if they completed taking notes prior to engaging in the MPTG). Future systems could work to more accurately determine the time of each SRL strategy, which will give a better measure of the time students spend engaging in SRL processes and will help us to better differentiate between prior knowledge groups. More enhanced multi-agent hypermedia-learning environments can cater to each student's needs, such as considering levels of prior knowledge, and thus be more effective in teaching students to become better self-regulators of their learning.

References

Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2010). Automated, unobtrusive, action-by-action assessment of self-regulated learning with an intelligent tutoring system. *Educational Psychologist*, *45*(4), 224–233.

Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist (Special Issue on Computers as Metacognitive Tools for Enhancing Student Learning)*, *40*(4), 199–209.

Azevedo, R. (2007). Understanding the complex nature of self-regulatory processes in learning with computer-based learning environments: An introduction. *Metacognition and Learning*, *2*, 57–65.

Azevedo, R. (2009). Theoretical, methodological, and analytical challenges in the research on metacognition and self-regulation: A commentary. *Metacognition & Learning, 4*(1), 87–95.

Azevedo, R., & Aleven, V. (2013). Metacognition and learning technologies: An overview of the current interdisciplinary research. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 1–16). Amsterdam, the Netherlands: Springer.

Azevedo, R., Cromley, J. G., Moos, D. C., Greene, J. A., & Winters, F. I. (2011). Adaptive content and process scaffolding: A key to facilitating students' self-regulated learning with hypermedia. *Psychological Testing and Assessment Modeling, 53*, 106-140.

Azevedo, R., & Feyzi-Behnagh, R. (2010). Dysregulated learning with advanced learning technologies. In R. Pirrone, R. Azevedo, & G. Biswas (Eds.), *Proceedings of the AAAI Fall Symposium on Cognitive and Metacognitive Educational Systems* (pp. 5–10). Menlo Park, CA: AAAI Press.

Azevedo, R., Feyzi-Behnagh, R., Duffy, M., Harley, J., & Trevors, G. (2012a). Metacognition and self-regulated learning in student-centered learning environments. In D. Jonassen & S. Land (Eds.), *Theoretical foundations of learning environments*. New York, NY: Routledge.

Azevedo, R., Harley, J., Trevors, G., Duffy, M., Feyzi-Behnagh, R., Bouchet, F., & Landis, R. (2013) Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 427-449). Amsterdam, The Netherlands: Springer.

Azevedo, R., Landis, R.S., Feyzi-Behnagh, R., Duffy, M., Trevors, G., Harley, J., Bouchet, F., Burlison, J., Taub, M., Pacampara, N., Yeasin, M., Rahman, A.K.M.M., Tanveer, M.I., & Hossain, G. (2012b). The effectiveness of pedagogical agents' prompting and feedback in facilitating co-adapted learning with MetaTutor. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Proceedings of the 11th International Conference of Intelligent Tutoring Systems* (pp. 212-221). Amsterdam, The Netherlands: Amsterdam.

Azevedo, R., Moos, D., Johnson, A., & Chauncey, A. (2010). Measuring cognitive and metacognitive regulatory processes used during hypermedia learning: Issues and challenges. *Educational Psychologist*, *45*(4), 210-223.

Berkhin, P. (2006). A survey of clustering data mining techniques. In J. Kogan, C. Nicholas, & M. Teboulle (Eds.), *Grouping multidimensional data: Recent advances in clustering* (pp. 25–71). Heidelberg, Germany: Springer.

Biswas, G., Jeong, H., Kinnebrew, J., Sulcer, B., & Roscoe, R. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology-Enhanced Learning*, *5*(2), 123–152.

Bouchet, F., Harley, J., Trevors, G., & Azevedo, R. (2013). Clustering and profiling students according to their interactions with an intelligent tutoring system fostering self-regulated learning. *Journal of Educational Data Mining*, *5*(1), 104–146.

Bouchet, F., Kinnebrew, J., Biswas, G., & Azevedo, R. (2012). Identifying students' characteristic learning behaviors in an intelligent tutoring system fostering self-regulated learning. In K. Yacef, O. Zaiane, A. Hershkovitz, & J. Stamper (Eds.), *Proceedings of the 5th International Conference of Educational Data Mining* (pp. 65–72). Amsterdam, the Netherlands.

D'Mello, S. K., Chauncey Strain, A., Olney, A., & Graesser, A. (2013). Affect, meta-affect, and affect regulation during complex learning. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 669–681). Amsterdam, the Netherlands: Springer.

Ericsson, K., Charness, N., Feltovish, P. J., & Hoffman, R. (2006). *The Cambridge handbook of expertise and expert performance*. Cambridge, England: Cambridge University Press.

Graesser, A., Chipman, P., King, B., McDaniel, B., & D'Mello, S. (2007). Emotions and learning with AutoTutor. In R. Luckin et al. (Eds.), *Proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED 2007), Frontiers in Artificial Intelligence and Applications, Vol. 158* (pp. 569–571). IOS Press.

Graesser, A., & McNamara, D. (2010). Self-regulated learning in learning environments with pedagogical agents that interact in natural language. *Educational Psychologist*, *45*(4), 234–244.

Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist*, *45*(4), 203–209.

Hegland, M. (2001). Data mining techniques. *Acta Numerica*, *10*, 313–355. doi:10.1017/S0962492901000058

Jonassen, D., & Land, S. (Eds.). (2012). *Theoretical foundations of learning environments*. New York, NY: Routledge.

Kinnebrew, J., Biswas, G., Sulcer, B., & Taylor, R. (2013). Investigating self-regulated learning in teachable agent environments. In R. Azevedo & V. Aleven (Eds.), *International* *handbook of metacognition and learning technologies* (pp. 451–470). Amsterdam, the Netherlands: Springer.

Kintsch, W. (1988). The role of knowledge discourse comprehension: A construction-integration model. *Psychological Review 95*, 163–182.

Lajoie, S. P., Naismith, L., Poitras, E., Hong, Y.-J., Cruz-Panesso, I., Ranellucci, J., Mamane, S., & Wiseman, J. (2013). Technology-rich tools to support self-regulated learning and performance in medicine. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 427–449). Amsterdam, the Netherlands: Springer.

Lester, J. C., Mott, B. W., Robinson, J. L., & Rowe, J. P. (2013). Supporting self-regulated science learning in narrative-centered learning environments. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 471–483). Amsterdam, the Netherlands: Springer.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*(1), 14–19. doi:10.1037/0003-066X.59.1.14.

Moos, D. C., & Azevedo, R. (2008). Self-regulated learning with hypermedia. The role of prior domain knowledge. *Contemporary Educational Psychology*, *33*, 270–298.

Moos, D. C., & Azevedo, R. (2009). Self-efficacy and prior domain knowledge: To what extent does monitoring mediate their relationship with hypermedia learning? *Metacognition and Learning*, *4*, 197–216.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychology, 38*(1), 1–4.

Pekrun, R., Goetz, T., Frenzel, A., Barchfeld, P., & Perry, R. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*, 36–48.

Pujari, A. K. (2001). *Data mining techniques*. Hyderguda, India: Universities Press Private Limited.

Shapiro, A. (1999). The relationship between prior knowledge and interactive organizers during hypermedia-aided learning. *Journal of Educational Computing Research*, *20*, 143–163.

Shapiro, A. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal*, *41*, 159-189.

Van Seters, J. R., Ossevoort, M. A., Tramper, J., & Goedhart, M. J. (2012). The influence of student characteristics on the use of adaptive e-learning material. *Computers & Education*, *58*, 942–952.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Hillsdale, NJ: Erlbaum.

Winne, P. H., & Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and Self-regulated learning: Theory, Research and Applications* (pp. 298–314). New York: Lawrence Erlbaun.

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). Orlando, FL: Academic Press.

Winters, F. I., & Azevedo, R. (2005). High-school students' regulation of learning during computer-based science inquiry. *Journal of Educational Computing Research*, *33*(2), 189–217.

Winters, F., Greene, J. A., & Costich, C. M. (2008). Self-regulation of learning within computer-based learning environments: A critical analysis. *Educational Psychology Review*, *20*(4), 429–444.

Woolf, B. (2009). *Building intelligent interactive tutors*. Amsterdam, the Netherlands: Morgan Kaufmann.

Zimmerman, B., & Schunk, D. (Eds.). (2011). *Handbook of self-regulation of learning and performance*. New York, NY: Routledge.

Figure Legend

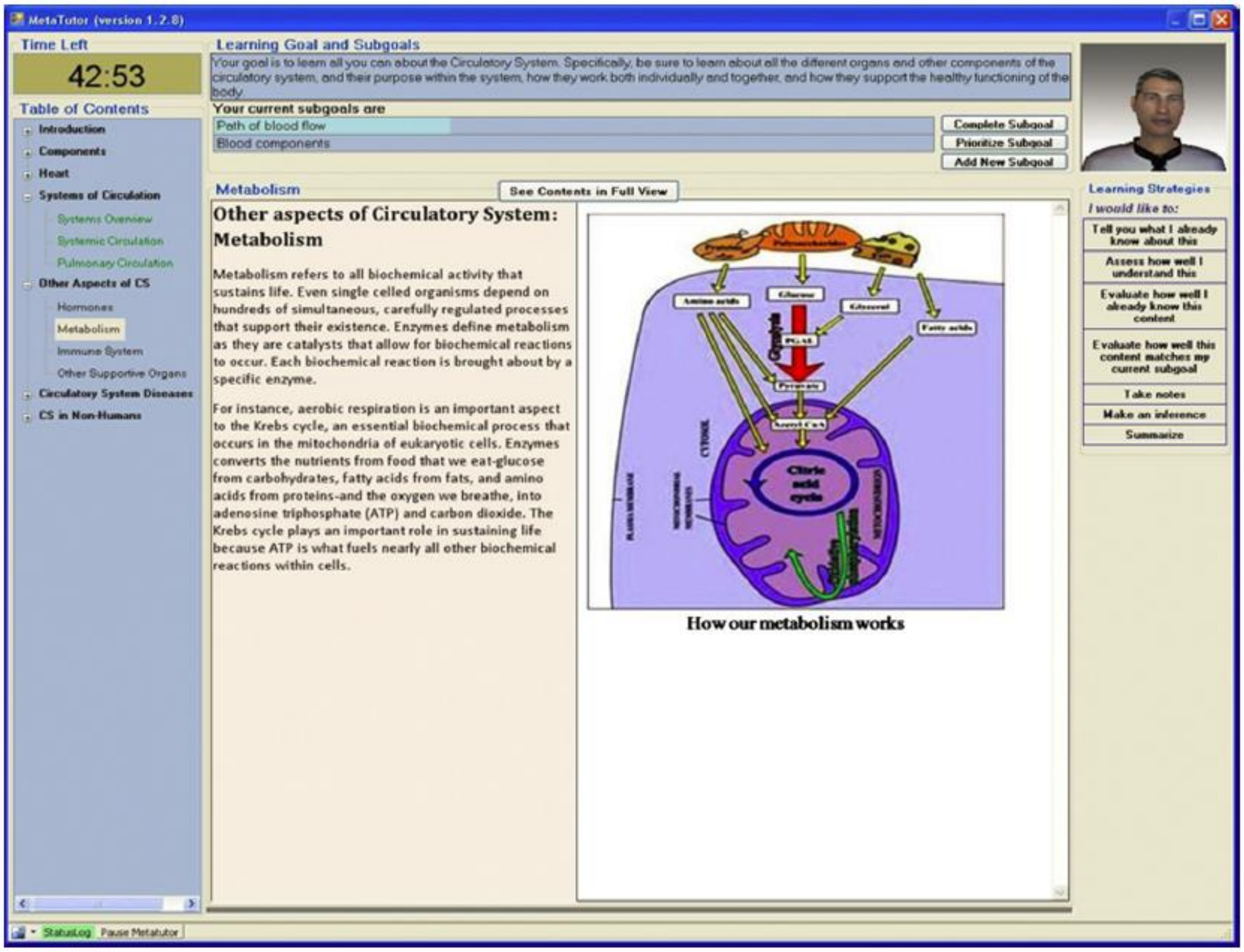

Figure 1. Screenshot of the MetaTutor interface.

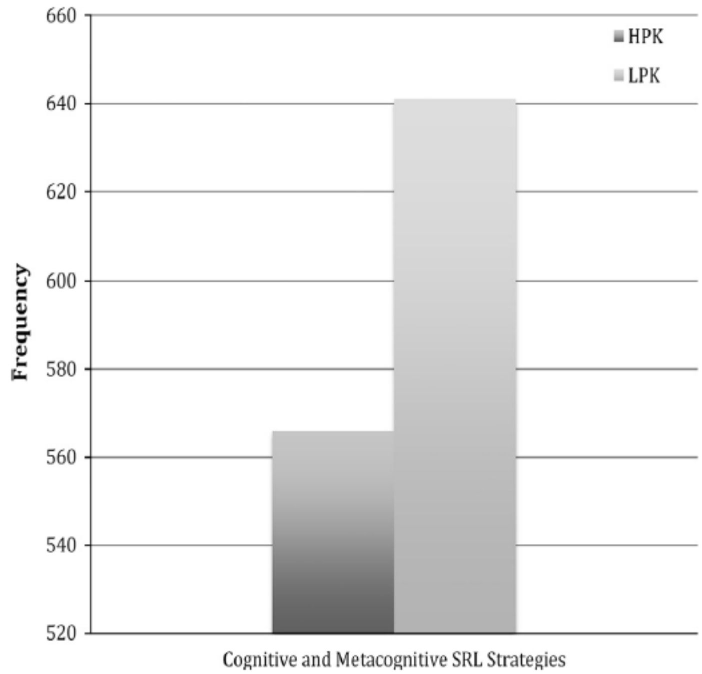

Figure 2. Frequencies of total self-regulated learning strategies.

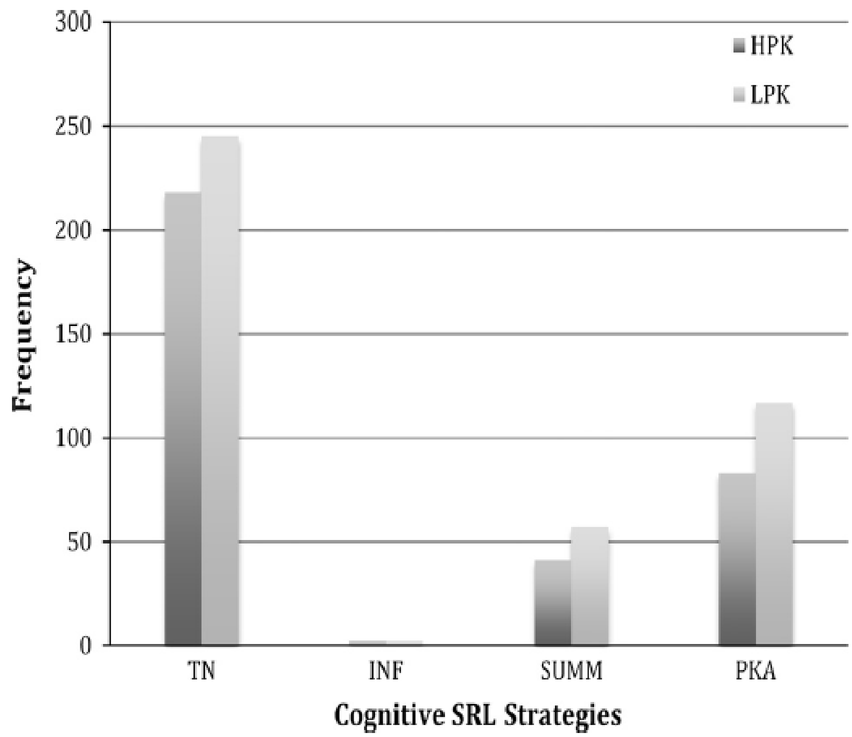

Figure 3. Frequencies of each cognitive SRL strategy.

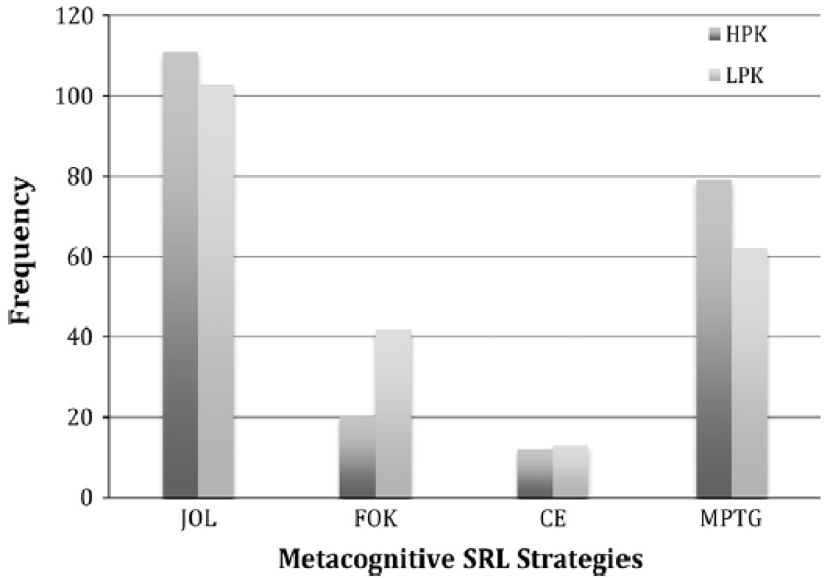

Figure 4. Frequencies of each metacognitive SRL strategy.

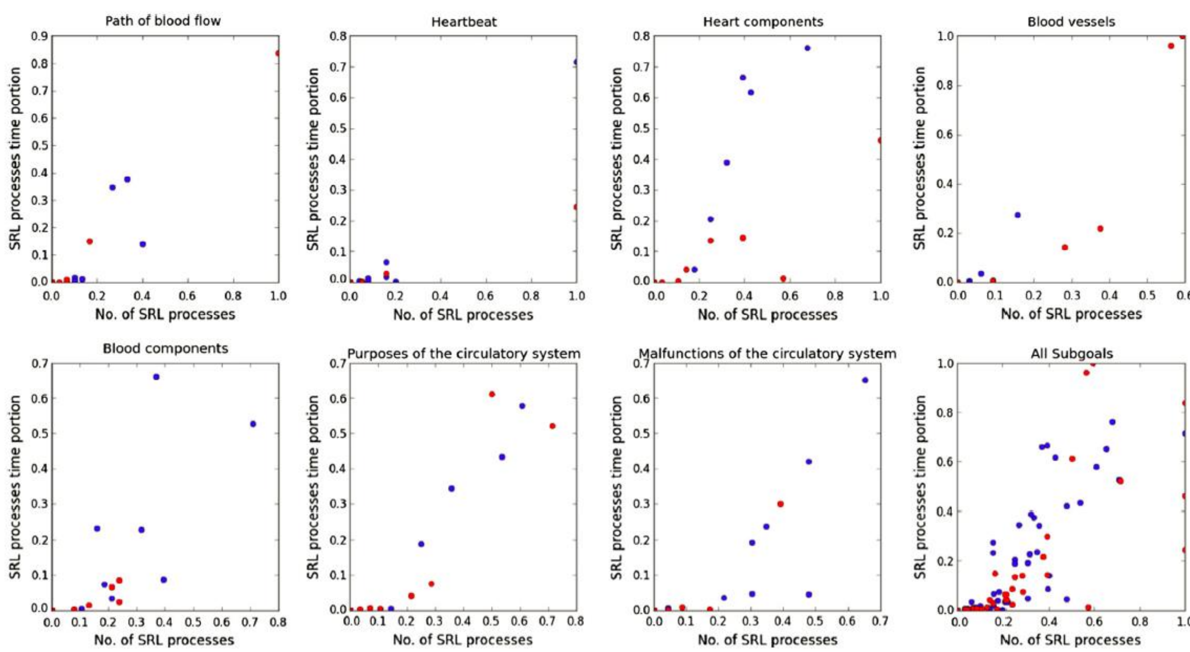

Figure 5. Data plots for each of the seven MetaTutor sub-goals and all seven sub-goals based on the number of SRL processes and time spent engaging in SRL processes. Note: HPK students are shown as blue dots, and LPK students are shown as red dots

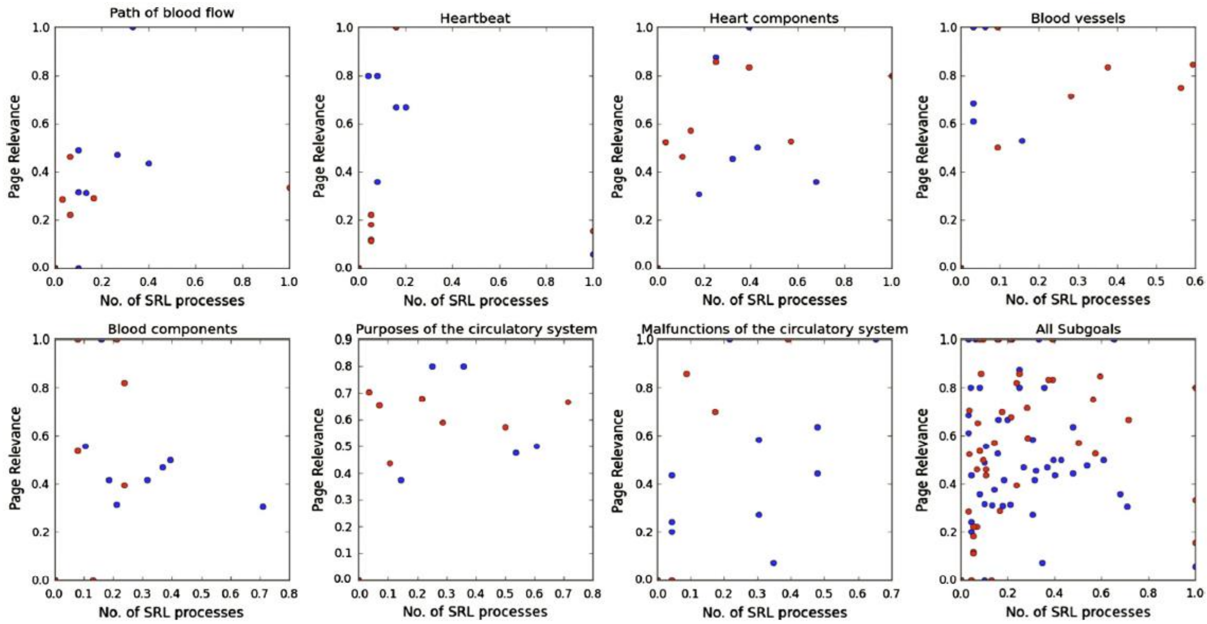

Figure 6. Data plots for each of the seven MetaTutor sub-goals and all seven sub-goals based on the number of SRL processes and page relevancy. Each plot represents a data point for each participant: the HPK students are represented in blue, and the LPK students are represented in red.