

# Clustering and Profiling Students According to their Interactions with an Intelligent Tutoring System Fostering Self-Regulated Learning

François Bouchet, Jason M. Harley, Gregory J. Trevors and Roger Azevedo

Department of Educational and Counselling Psychology (ECP),  
Laboratory for the Study of Metacognition and Advanced Learning Technologies,  
McGill University, Montreal, QC, H3A 1Y2  
{francois.bouchet, roger.azevedo}@mcgill.ca  
{jason.harley, gregory.trevors}@mail.mcgill.ca

---

In this paper, we present the results obtained using a clustering algorithm (Expectation-Maximization) on data collected from 106 college students learning about the circulatory system with MetaTutor, an agent-based Intelligent Tutoring System (ITS) designed to foster self-regulated learning (SRL). The three extracted clusters were validated and analyzed using multivariate statistics (MANOVAs) in order to characterize three distinct profiles of students, displaying statistically significant differences over all 12 variables used for the clusters formation (including performance, use of note-taking and number of sub-goals attempted). We show through additional analyses that variations also exist between the clusters regarding prompts they received by the system to perform SRL processes. We conclude with a discussion of implications for designing a more adaptive ITS based on an identification of learners' profiles.

Key Words and Phrases: Profiling, Cluster, Expectation-Maximization, Intelligent Tutoring System, Agent-Based System, Self-Regulated Learning, Metacognition, Adaptivity

---

## 1. INTRODUCTION

A major challenge for researchers and developers of agent-based ITSs is how best to adapt to learners in order to provide individualized instruction i.e., for pedagogical agents (PAs) to adapt their tutorial strategies to deal with learners' emerging understanding of the topic, model self-regulated learning (SRL) skills in order to support their use during learning, prompt key metacognitive processes related to mental model development, alter the instructional sequence and pace to deal with impasses, engage in dialogue aimed at minimizing negative emotions, etc.; see [Aleven et al. 2010; Shute and Zapato-Rivera 2012; Woolf 2009]. One way to profile learners is to use different measures (e.g., pretests, self-report questionnaires) to assess their knowledge of the topic, cognitive abilities, metacognitive monitoring and control strategies, motivation and emotional traits prior to their learning session. These data are then used to enable the system to build an initial student model and enact particular tutoring strategies to facilitate learning with an ITS (e.g., Shute & Zapato-Rivera 2012). During the session, the system's student model

is updated and its tutoring strategies (e.g., modeling, scaffolding) are modified adaptively according to changes in students' learning, strategy use, performance, emotions, monitoring and regulatory skills, motivation, etc. While there is plethora of research on student models, most of this work has not focused on the complex nature of cognitive, metacognitive, motivational, and emotional processes with ITSs that used non-linear hypermedia learning materials such as MetaTutor [Azevedo et al., 2012]. As such, this paper focuses on trying to establish a posteriori clustering of students' according to their interaction with an ITS scaffolding their use of self-regulated learning processes.

The idea of clustering students according to their behavior in the context of learning systems has been explored in several research works, because of the potential it offers for the system (in an agent-based ITSs) or the human teacher (in a virtual classroom type of environment) to provide more adaptive scaffolding (see section 6.3.2 in Vellido et al. [2011] for a review). Using MetaTutor, we have collected large amounts of data from college students while they were using the system to learn about the human circulatory system. In order to improve the adaptivity of the system, we are therefore interested in studying the relationship on the one hand, between learners' performance and their interactions with MetaTutor, and on the other hand, their deployment of SRL processes. Specifically, we are interested in answering the following questions: (1) can we establish the existence of clusters of students according to their performance and interaction with MetaTutor? And if that's the case, (2) what are the characteristics that distinguish students belonging to those different clusters, and in particular, how do they relate to their use of self-regulated learning processes?

We begin to answer these questions in Section 2, by presenting background information on the SRL model used in this article and the challenges relative to its integration into an ITS. In Section 3, we provide information about the participants in our studies, the particular multi-agent ITS used (MetaTutor), the experimental procedure followed as well as the different types of data collected during the learning session. In Section 4, we describe the analyses performed on the data, first to extract clusters of participants using the Expectation-Maximization algorithm, and then to identify through inferential multivariate statistics the sources and directions of variance between the clusters (therefore validating their extraction). We also consider additional sources of variance (such as system-generated SRL prompts) and compare the distribution with the one obtained when considering another group of students who interacted with a different

version of MetaTutor. Section 5 discusses the significance of those results in the context of MetaTutor and how they can impact future changes to the system. Finally, in Section 6 we discuss related work and limitations, and conclude in Section 7, by presenting several directions for future research.

## 2. SELF-REGULATED LEARNING IN ADAPTIVE INTELLIGENT TUTORING SYSTEMS

### 2.1 Theoretical Framing and Relation to SRL Product and Process Data

This paper is theoretically-guided by contemporary models of SRL that emphasize the temporal deployment of cognitive, metacognitive, and affective (CAM) processes during learning [Azevedo et al., 2005, 2010, 2012]. As such, the goal is to use multiple measures to detect, track, and model learners' use of CAM processes during learning. This led us to use Winne and Hadwin's model [1998, 2008] because it proposes that learning occurs in four basic phases: (1) task definition, (2) goal-setting and planning, (3) studying tactics, and (4) adaptations to metacognition. Their model emphasizes the role of metacognitive monitoring and control as the central aspects of learners' ability to learn complex material across different instructional contexts (e.g., using a multi-agent system, MetaTutor, to track and foster SRL) in that information is processed and analyzed within each phase of the model. Recently, Azevedo and colleagues [2007, 2009, 2012, in press] extended this model and provided extensive evidence regarding the role and function of several dozen CAM processes during learning with student-centered learning environments (e.g., multimedia, hypermedia, simulations, intelligent tutoring systems).

In brief, our model [Azevedo et al. in press] makes the following assumptions: (1) successful learning involves having learners monitor and control (regulate) key CAM processes during learning; (2) SRL is context-specific and therefore successful learning may require a learner to increase/decrease the use of certain key SRL processes at different points in time during learning; (3) a learner's ability to monitor and control both internal (e.g., prior knowledge) and external factors (e.g., changing dynamics of the learning environment; relative utility of an agent's prompt) are crucial in successful learning; (4) a learner's ability to make adaptive, real-time adjustments to internal and external conditions, based on accurate judgments of their use of CAM processes, is fundamental to successful learning; and; (5) certain CAM processes (e.g., interest, self-efficacy, task value) are necessary to motivate a learner to engage and deploy appropriate CAM processes during learning and problem solving.

This model is best suited for this study since it deals specifically with the person-in-context perspective and postulates that CAM processes occur during learning with a multi-agent system, which will be useful in examining when and how learners will regulate their learning about a complex science topic (the human circulatory system in the case of MetaTutor). As such, the macro-level processes used in this paper are reading, metacognitive monitoring, and learning strategies. Reading behavior is critical since it is the most important activity related to acquiring, comprehending, and using content knowledge related to the science topic. During reading, learners need to monitor and regulate several key processes such as: (1) selecting relevant content (i.e., text and diagrams) based on their current sub-goal; (2) spending appropriate amounts of time on each page, depending on their relevance regarding their current sub-goal; (3) deciding when to switch or create a new sub-goal; (4) making accurate assessments of their emerging understanding; (5) conceptually connecting content with prior knowledge; (6) adaptively selecting, using, and assessing the effective use of several learning strategies including re-reading, coordinating informational sources, summarizing, and making inferences, in order to comprehend the material at various levels (i.e., declarative, procedural, and conceptual knowledge); and, (7) making adaptive changes to behavior based on a variety of external (e.g., quiz scores, quality and timing of agents' prompts and feedback) and internal sources (e.g., affective experiences including both positive and negative affective states, perception of task difficulty). In sum, SRL involves the continuous monitoring and regulation of CAM processes during learning with MetaTutor. As such, we have embarked on the analyses of these key CAM processes by specifically integrating several product and process data using mainly log-file data to examine student clusters following their two-hour interaction with an adaptive version of MetaTutor.

## 2.2 Context: Intelligent Tutoring Systems

Learning systems such as advanced agent-based systems are effective to the extent that they can adapt to the needs of individual students by systematically and dynamically providing prompts, scaffolding, and feedback based on their ability to detect, track, and model key SRL processes [Azevedo et al. in press; Biswas et al. 2010; Graesser et al. in press; McQuiggan and Lester 2009; White et al. 2009]. A major challenge in determining how to adapt to students is that these self-regulatory processes are deployed in real-time and fluctuate during learning based on a complex set of interactions between the learner, the agent-based system, and the instructional context that changes dynamically during learning. For example, a learner must set relevant sub-goals for the learning session,

activate relevant prior knowledge in order to anchor new information and determine an optimal instructional sequence, metacognitively monitor and accurately judge their emerging understanding and evaluate the relevancy of multiple sources of information vis-à-vis their current learning sub-goal. Furthermore, they will need to determine and change (at any given point) which learning strategy (e.g., coordination of information sources, summarization) they use in order to facilitate their knowledge acquisition. In addition to these key cognitive and metacognitive processes, the learner must also monitor and control their motivation and emotional processes. For example, a learner may need to determine how best to self-generate interest and find value in a task and topic given parameters like their career choice. They may also need to monitor and control their emotions, such as their level of confusion, to prevent a shift toward emotional states detrimental to their learning performance like frustration and boredom [D'Mello & Graesser in press]. As such, the ability of these environments to provide adaptive, individualized scaffolding is based on an understanding of how learner characteristics, system features, and the mediating contextual learning processes interact during learning [Aleven et al. 2010; Woolf 2009]. A critical aspect of providing individualized instruction is scaffolding, or instructional support in the form of prompts, guidance, and modeling, which are used during learning to support a significantly higher level of understanding than the one students would attain if they learned on their own. While providing adaptive scaffolding to students learning about well-structured tasks with traditional ITSs has been shown to be effective (e.g., see [VanLehn 2011]), providing adaptive scaffolding to students learning about conceptually-challenging domains remains a challenge for agent-based learning systems. We argue that (1) harnessing the full power of agent-based adaptive systems will require empirical research aimed at understanding what kinds of scaffolds are effective in facilitating individualized instruction, and when they are best deployed, and that (2) because of the amount of data collected and of the need to automatically identify students' profiles as they interact with the system, educational data mining and machine learning methods are key to building adaptive multi-agent systems designed to detect, track, model, and foster students' self-regulated learning. As such, the goal of this paper is to use an educational data mining approach on data collected with a multi-agent system, in order to extract different profiles of learners that could be used to improve the adaptivity of the system.

### 2.3 Challenges

Using multi-agent systems such as MetaTutor to learn about a complex and challenging topic, such as the human circulatory system, requires a student to regulate their CAM processes throughout the task while non-linearly navigating the system by hyper-linking to different pages of content and managing the various informational sources (i.e., text and diagrams). Learners also need to monitor: how much they already know about the topic (and still need to learn during the predetermined time set for the learning session); their emerging understanding as they progress through pages and diagrams of the circulatory system; how content presented in the system relates to their prior knowledge (which may require deciding on an optimal instructional sequence of the content and of the information sources); the relevancy of content given their current sub-goal; and their progress toward completing their goals. Each of these monitoring processes leads to metacognitive judgments, varying in accuracy, and impacting both the students' decision to adapt and their selection of the learning strategy to use in order to rectify the judgment. For example, a judgment of learning (e.g., "I do not understand this paragraph on the role of the bicuspid valve") may lead a student to re-read the sentences over again to see if they can improve their comprehension of the role of the valve. However, the choice of learning strategies is based on students' making accurate metacognitive judgments, and having metacognitive knowledge and the regulatory skills needed to continuously regulate during learning with the system. They also need to monitor their understanding and modify their plans, goals, strategies, and effort in relation to both internal (e.g., cognitive, metacognitive, motivational, and affective) and contextual conditions (e.g., changing task conditions, scaffolding from the pedagogical agents, perceived utility of an agent's advice, prompts, guidance, and modeling), and, depending on the learning task, reflect on the learning episode [Azevedo et al. 2010; Winne and Hadwin 1998; 2008]. The complexity of these processes and their dynamics and fluctuations during learning pose several problems for agent-based systems, including overwhelming empirical evidence that most students do not regulate these processes during learning (see [Azevedo and Aleven in press; Graesser and McNamara 2010; Winne and Nesbit 2009]). This calls into question key issues related to agent-based systems' ability to assess and deliver adaptive scaffolding through the use of their agents.

Recently, some researchers have focused on providing adaptive scaffolding via their PAs. For example, MetaTutor agents have been used to prompt metacognitive judgments by asking questions to students that trigger a judgment, such as determining whether content

is relevant to one's current sub-goal [Azevedo et al. in press]. In AutoTutor, agents have been used to engage in a dialogue (with students) aimed at eliciting and rectifying misconceptions in physics [Graesser et al. in press]. Students using Betty's Brain are guided through a complex sequence of metacognitive and cognitive activities designed to develop their conceptual understanding of ecosystems [Leelawong and Biswas 2008]. Each of these agent-based systems provide some level of adaptivity and scaffolding, based on their student models of the dozens of learner, system, and contextual factors. Moreover, the aforementioned complex nature of SRL adds to the challenge of providing accurate and timely scaffolding to each individual learner. As such, researchers have recently turned to machine learning and educational data mining techniques to augment their system's effectiveness (e.g., [Baker et al. in press; Baker and Yacef 2009; Kinnebrew and Biswas 2011]).

### 3. DATA COLLECTION WITH METATUTOR

#### 3.1 Participants

One hundred and six ( $N = 106$ ) undergraduate students from two large, public universities in North America participated in this study. The mean age of the sample was 20.9 years ( $SD = 2.85$  years) and the mean self-reported GPA was 3.05 ( $SD = 0.45$ ). Participants were randomly assigned to two different conditions (cf. Section 3.4 for more details on those). More than half of the participants were female (69%) and approximately half of the sample (47%) was Caucasian, followed by African American (40%), while the remainder identified themselves as belonging to some other ethnic group. Participants represented several academic majors including social sciences (32%), humanities (21%), science, math, and engineering (15%), and management and business (30%). Less than half of the sample (40%) reported taking biology courses at the undergraduate level prior to their learning session with MetaTutor. Of this 40%, only half had taken more than one undergraduate-level biology course<sup>1</sup>.

#### 3.2 MetaTutor: A Multi-Agent Learning Environment

MetaTutor [Azevedo et al. 2010; 2011; 2012; in press] is a multi-agent intelligent hypermedia tutoring environment which contains 41 pages of text and static diagrams

---

<sup>1</sup> With the exception of ethnicity and gender, all other demographic information was based on data collected from 76 participants (71.7% of sample), as the remaining participants did not provide this information.

about the human circulatory system organized by a table of contents (see Figure 1). The underlying assumption of MetaTutor is that students should regulate key CAM processes in order to learn about complex and challenging science topics. This non-linear, self-paced environment allows learners to access content and to navigate to new pages by selecting a subtopic from headings located in the table of contents (cf. B in Figure 1). A timer, located at the top left-hand corner of the environment, displays the amount of time remaining in the session (cf. A in Figure 1). The experimenter's overall learning goal and participants' self-set relevant and topic-specific sub-goals are displayed at the top of the interface, which they can select to manage or prioritize their sub-goals, and to track the percentage of relevant content already learnt based on progression bars (cf. C in Figure 1). One of the four PAs (Gavin, Pam, Mary, or Sam) is always displayed in the upper right-hand corner of the environment (cf. D in Figure 1). These agents provide varying degrees of prompting and feedback throughout the learning session to scaffold students' SRL skills such as summarizing and making judgments of learning and content understanding (see [Azevedo et al. 2010] for details). Each agent serves a different purpose: (1) Gavin the Guide helps students to navigate through the system, (2) Pam the Planner guides students in setting appropriate sub-goals, (3) Mary the Monitor helps students to monitor their progress toward achieving their sub-goals, and (4) Sam the Strategizer helps students to deploy SRL learning strategies, such as summarizing and note-taking (see Appendices A and B for more details). Learners can interact with these agents and enact specific SRL learning processes by selecting any feature of the SRL palette displayed at the right-hand side of the interface during the learning session (cf. E in Figure 1). Students can use this palette to indicate their intention to deploy planning, monitoring, or learning strategies. For instance, by clicking the "Take Notes" button on the SRL palette, participants can take notes of the content. Similarly, they can click other buttons to test their understanding of the content by assessing their understanding and completing a quiz, activate their prior knowledge of the content, evaluate the relevancy of the content, make an inference, or summarize (see Appendix A for a complete list of SRL learning strategies and cognitive and metacognitive processes activated during MetaTutor and how students trigger them through the palette). Learners can access text entered on the keyboard and their interaction history with agents by clicking a button at the bottom right-hand corner of the environment to view their interaction log (cf. F in Figure 1). MetaTutor tracks all learner interactions and logs every action taken by the learner in a log-file. These log-files are uploaded to a database, which is then mined for information about participant interactions.

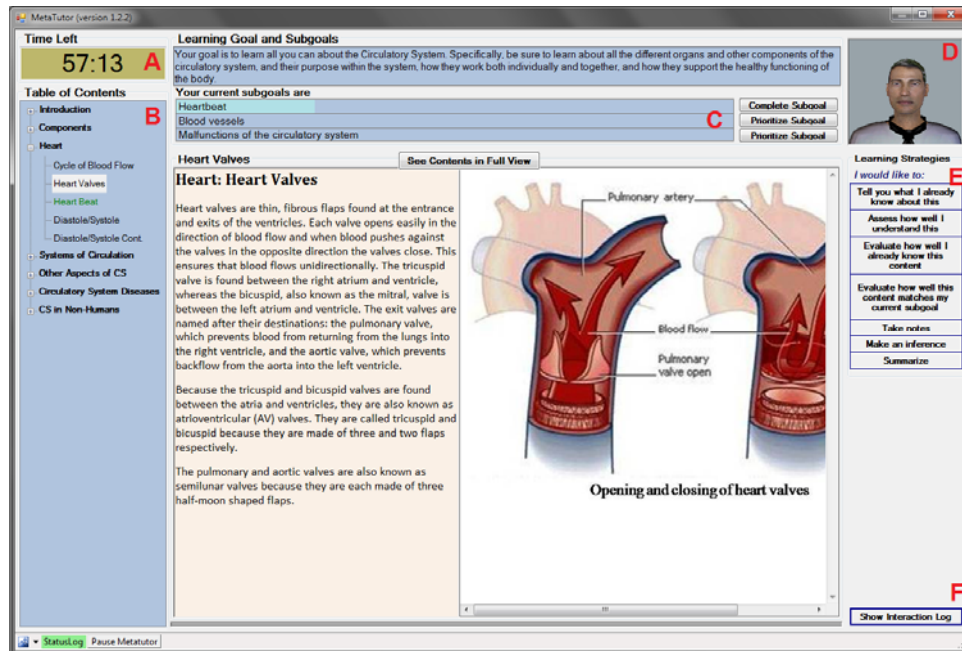


Figure 1. Annotated screenshot of MetaTutor interface

### 3.3 Measures & Materials

**3.3.1 Process measures.** Process measures collected during the two-hour learning session with MetaTutor included: log-files, facial expressions, diagrams drawn and notes taken on paper and eye-tracking data. These streams of process data provided information about learners' cognitive, metacognitive, and affective SRL processes during the learning session. In this article, we focus exclusively on log-file data to mine and to analyze the SRL processes (see Section 3.6 for further details). Those log-files collected learners' interactions (i.e., mouse-clicks and keyboard entries) within the system, including number of times visiting each page, time spent on each page, and time spent taking notes using the embedded note-taking interface. The content of hand-written notes and diagrams taken on paper, captured on an ACECAD DigiMemo L2 digital notepad, is not considered in this paper, but since the device was connected to MetaTutor, it allowed adding events into the log-file when the participant was starting or finishing taking notes or drawing on paper, and those events are therefore included.

**3.3.2 Product measures.** Product measures included: a demographics questionnaire, an SRL quiz, as well as pretest and posttest on the human circulatory system. A demographic questionnaire was administered to collect participants' demographic information such as age, gender, academic major, and ethnicity. A 13-item SRL quiz was

also administered to assess participants' existing knowledge of self-regulated learning processes. A pretest and posttest assessed students' understanding of the human circulatory system. Each was comprised of 25 multiple-choice items with three foils for every question (i.e., near miss, thematic, and unrelated to the target answer). Items on the pretest and posttest included text-based items (which could be answered by directly referring to one sentence within the content) and inferential items (which required integrating information from at least two sentences within the content). Two equivalent versions of the test were created for the pretest and posttest and were counterbalanced across participants. Participants' hand-written notes and drawings taken while learning about the circulatory system were collected at the end of the session (before the administration of the posttest).

### 3.4 Research Design & Learning Conditions

Two versions of the MetaTutor environment were designed in this study to examine the effectiveness of pedagogical agents' scaffolding on participants' use of SRL processes and learning outcomes. Participants were randomly assigned to either a prompt and feedback (PF) condition or a control (C) condition and asked to learn about the circulatory system using MetaTutor. In the PF condition, participants were prompted by PAs to use specific planning, metacognitive monitoring, and learning strategies and were given immediate feedback about the quality and accuracy of these processes. For example, after completing a quiz, participants in the PF condition were given information about their performance on the quiz and, depending on their knowledge acquisition, were prompted to either continue reviewing the multimedia content or progress to another sub-goal. The timing of agent-generated prompts was adaptive to each learner and was determined using various interaction factors, such as time on page, time on current sub-goal, number of pages visited, relevancy of current page for the sub-goal (see Appendix B for a complete description of system-generated rules – note that those rules did not apply to participants in condition C). Participants randomly assigned to the control condition did not receive prompts or feedback from the PAs. However, in both versions of MetaTutor, all other features of the environment were available and all participants were able to interact with the PAs during the learning session by clicking on one of the SRL palette buttons.

### 3.5 Experimental Procedure

The MetaTutor experiment was conducted across two sessions for each participant. Session 1 of the experiment took one hour (maximum) and Session 2 of the experiment

took three hours to complete (amount of time was fixed to facilitate comparison between students). In a few cases, both sessions took place on the same day, so long as they occurred one hour apart (to avoid participant fatigue) and always occurred no more than three days apart. During the first session, participants filled out the consent form and were given as much time as they needed to complete the demographics questionnaire and the SRL quiz (designed to gauge participants' existing declarative knowledge of SRL processes). Subsequently, they were administered and asked to complete the circulatory system pretest within 20 minutes. All participants used MetaTutor on a desktop computer with a Core 2 Duo 2.80GHz processor, 2GB of RAM and Windows XP, using a 17" monitor with a 1024x768 resolution (to have MetaTutor running in full screen). Agents' verbalizations, generated through Text-To-Speech engines from Nuance and Cepstral, were presented through speakers hooked up to the desktop. At the end of the session, each participant was paid \$10 for the 60-minute session.

During the second session, on the same computer, the eye-tracking device was calibrated for each learner individually. Next, each learner was shown a short video (50s) briefly presenting the learning environment and providing the learners with their overall learning goal. Following the introductory video, the learners were instructed by one of the PAs to set their sub-goals for their learning session by typing freely while the system matched (when relevant) their proposed sub-goal to one of the seven ideal sub-goals (i.e., not too broad or specific) associated to the studied topic. If the proposed sub-goal was related to one of the appropriate sub-goals, the PA guided them towards it. Once learners had set three appropriate sub-goals, they were shown another video (3m20s) explaining and demonstrating the various functionalities of MetaTutor, including the use of the electronic note-taking feature (accessible through the SRL palette) and of the peripheral drawing pad (ACECAD DigiMemo L2) if they chose to draw or take notes on paper. Finally, participants were given 120 minutes to learn about the human circulatory system using MetaTutor. All participants were provided the opportunity to take a five-minute break during the two hours, although not all chose to do so. During the learning session, all participants were permitted to take notes or draw (although they were instructed that they could not access these notes or drawings during the posttest). Immediately after the learning session, participants were given up to 20 minutes to complete the posttest. Finally, all participants were paid \$40 for completion of the 2-session, 4-hour experiment and debriefed before leaving the lab.

### 3.6 Coding and Scoring

In this section, we present some of the log-file and learning outcomes data measured with MetaTutor, which included 26 variables (see Table I for definitions and the coding and scoring procedure used for each). These variables can be organized into four groups<sup>2</sup> (cf. categories titles in bold in Table I), including those which measured learners' knowledge, learning goal management, reading time, and self-regulated learning strategies that learners' could engage in.

Table I. Definitions of variables, organized by thematic groups.

| Variable                           | Definition                                                                                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Knowledge</b>                   |                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <u>ScorePre</u>                    | Number of correct answers scored by a participant on the Pretest. The range of scores was between 0-25.                                                                                                                                                                                                                                                                                                                            |
| <u>PostScore</u>                   | Number of correct answers scored by a participant on the Posttest. The range of scores was between 0-25.                                                                                                                                                                                                                                                                                                                           |
| <u>NumSGQuiz</u>                   | Number of times a participant took a sub-goal quiz (in order to change sub-goal), per period of 10 minutes (normalized over the session time).                                                                                                                                                                                                                                                                                     |
| <u>NumPageQuiz</u>                 | Number of quizzes a participant answered on individual pages, per period of 10 minutes, (normalized over the session time).                                                                                                                                                                                                                                                                                                        |
| <u>ScoreSGQuiz1stMean</u>          | Average proportion of correct answers scored by a participant on the first time they took a quiz associated to a sub-goal.                                                                                                                                                                                                                                                                                                         |
| <u>ScorePageQuiz1stMean</u>        | Average proportion of correct answers scored by a participant on the first time they took a quiz associated to a page.                                                                                                                                                                                                                                                                                                             |
| <b>Learning goal management</b>    |                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <u>PropSGattempted<sup>a</sup></u> | Proportion of the 7 sub-goals attempted by a participant.                                                                                                                                                                                                                                                                                                                                                                          |
| <u>NumSGChanges</u>                | Number of times a participant set a new current sub-goal, per period of 10 minutes, (normalized over the session time).                                                                                                                                                                                                                                                                                                            |
| <b>Session duration</b>            |                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <u>DurSession</u>                  | Time in seconds spent in the learning session: between the moment when the initial sub-goals had been set (i.e., Pam saying "you have set up 3 good sub-goals, let's move on to the next phase") and the moment when Gavin introduced the Posttest to a participant. Varies depending on the time spent on the initial sub-goal setting and on if the user is in the middle of an activity (summary, quiz...) when the time is up. |

<sup>2</sup> Note: Thematic groups appear in bold and variables used for cluster analysis and MANOVAs are underlined

DurReading Time in seconds a participant spent reading the multimedia content on the human circulatory system, including text and diagrams. Evaluated by subtracting from the session duration the moments when the participant was viewing a video, typing or drawing notes and deploying any SRL process.

---

**Self-regulated learning behaviors**

NumNote Taking<sup>b</sup> Number of times a participant opened the embedded note-taking interface and added content.

NumNote Checking<sup>b</sup> Number of times a participant opened the embedded note-taking interface without adding content.

DurNote Taking Time hours, minutes, and seconds during which a participant was typing notes using the note-taking embedded interface.

NumPLAN<sup>b</sup> Number of times a participant managed their sub-goals, which includes: 1) setting up a sub-goal (mandatory at the beginning of the session, but more can be set up later too) 2) postponing the current sub-goal, 3) receiving a suggestion to set up a new sub-goal.

NumSUMM<sup>b,c</sup> Number of times a participant was prompted to type (without the possibility to refuse) or initiated (under their own volition) a summary of the current page's content.

NumMPTG<sup>b,c</sup> The number of times a participant was prompted to manage, or decided to manage on their own initiative, their progression toward the current sub-goal by assessing their current understanding (therefore leading to a sub-goal quiz).

NumRR<sup>b</sup> The number of times a participant was prompted by the system to reread a page they were currently on.

NumCOIS<sup>b</sup> The number of times a participant was prompted by the system to coordinate different information sources (i.e., text and corresponding image).

NumPKA<sup>b,c</sup> The number of times a participant was prompted to activate, or activated on their own initiative, prior knowledge about their current sub-goal.

NumJOL<sup>b,c</sup> The number of times a participant was prompted to judge, or judged on their own initiative, how well they had been learning from the page they were currently viewing. A page quiz always followed. Those events happened after spending some time on a page or if they left from a relevant page after having spent between 5 and 14 seconds on it – i.e., long enough to have done more than skimming through the content, but enough to have read a substantial amount.

NumFOK<sup>b,c</sup> The number of times a participant was prompted to express, or expressed on their own initiative, their feelings about their knowledge regarding the page they were viewing.

|                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| NumCE <sup>b,c</sup>  | The number of times a participant was prompted to evaluate, or evaluated on their own initiative, the content of the page they were viewing regarding the sub-goal they were working on. It happened after spending some time on a page or if they left from a relevant page after having spent between 5 and 14 seconds on it – i.e., long enough to have done more than skimming through the article, but enough to have read a substantial amount. |
| NumINF <sup>b,c</sup> | The number of times a participant took the initiative to make an inference about the content of the page they were viewing.                                                                                                                                                                                                                                                                                                                           |

---

<sup>a</sup> Note: when a participant was not actively working on any sub-goal, it was counted as a virtual undefined eighth sub-goal, therefore a participant with 4 sub-goals might have done the 3 original sub-goals + 1 extra one, or the 3 original sub-goals and kept working without setting any new sub-goals. It is hence also possible to reach a maximum number of 8 sub-goals. Less than 3 sub-goals means a participant never completed all the sub-goals set with Pam at the beginning of Session 2.

<sup>b</sup> Note: Number of events or processes, per period of 10 minutes, normalized over the session time. Prior to analysis, raw frequency counts of each of the SRL behaviors was divided by each participant's time on task in 10-minute intervals (i.e., total session time minus time spent viewing videos, listening to agents, typing, and completing quizzes). This was done to control for the variation in the time participants spent with the material. The resulting rates thus represent the average frequency of a specific system-initiated SRL triggered during a 10-minute interval.

<sup>c</sup> Note: SRL processes could be either user- or system-initiated. In our analyses, we only used system-initiated SRL processes, (e.g., we excluded SRL processes initiated by students' clicks on the palette).

## 4. CLUSTER ANALYSIS

### 4.1 Clusters Extraction

To assess the existence of different categories of students, we ran a cluster analysis over a subset of 12 of the variables (cf. the underlined ones in Table I) that did not include the ones related to system-initiated SRL processes (since we wanted to check afterward if there was a correlation between the value of those and the clusters extracted, whenever they weren't used for their formation – cf. Section 4.3.1). The posttest score was also excluded from the considered variables as it is one of the only variables of the list for which no value can be available before the very end of the session<sup>3</sup>: including it would therefore limit the potential use of those clusters for an online dynamic adaptation of the system. We decided to use the Expectation-Maximization (EM) algorithm, as implemented in Weka 3.6.5 [Hall et al. 2009], over the sub-sample of students in the PF condition, since they were the ones who interacted with a version of the system in which pedagogical agents provided them with the most adaptive and complex scaffolding of

---

<sup>3</sup> DurSession being mainly influenced by the time used by the student to set up initial sub-goals at the beginning of the learning session, a value for it can be available even before the session ends.

their SRL processes (we will come back to the case of participants from the control condition in Section 4.3.2). As we did not know a priori the number of categories of students to find, we used a 10-fold cross-validation technique with an increment of the number of clusters (starting with 1) as long as the log-likelihood averaged over the 10 folds was increasing (i.e. we stopped as soon as we got a lower log-likelihood with  $N+1$  clusters than with  $N$  clusters). To compensate for the sensitivity of EM to the choice of seed (i.e. the cluster initiator) for the algorithm, linked to its tendency to get stuck into local optima, we ran it with 2000 different seeds to initialize it, which yielded the results presented in Table II.

Overall, we see that the most frequent partitions of the subjects are the ones with 2 and 3 clusters, and we therefore focused on these ones. As expected, some of the partitions using different seeds were identical, which allowed us to associate a weight to the different partitions obtained with 2 and 3 clusters. In both cases, we observed that the partitions obtained were not very different from each other, with only a few subjects switching from one cluster to another. We therefore calculated the dominant partition with 2 and 3 clusters by selecting the cluster associated to each student according to the number of times he/she was classified in it (e.g., if student 1 was classified in cluster 0 in 84.6% of the 627 partitions with 3 clusters, and in cluster 1 in 15.4% of the other partitions, we considered it belonged to cluster 0 in the dominant partition). Using this method, only one of the 51 students was associated to two different clusters with a margin inferior to 60% (i.e., 80% of classification in one cluster and 20% in another), which means that only one dominant partition exists with 2 clusters (with 27 students in one cluster and 24 in the second one) as well as with 3 clusters (with 14, 9 and 28 students in each).

At this point of the analysis, both partitions could be acceptable: a traditional way to evaluate the additional value of each cluster is to examine their associated log-likelihood value and look for a scree-plot pattern (or “elbow”). Figure 2 reveals that such a pattern happens when increasing the number of clusters from three to four, therefore indicating that three seems to be an appropriate value for the number of clusters to consider. It is, however, essential to evaluate the coherence of the clusters from a statistical analysis of the different variables involved in their extraction, i.e. to identify which variables contribute to the clusters distinction. For this reason, we will consider in the next section the partition made of three clusters and check that it is indeed a logical choice from a statistical perspective.

Table II. Number of clusters obtained applying EM algorithm with 2000 different seeds

| Number of clusters                          | 1   | 2   | 3   | 4   | 5  | 6 and + |
|---------------------------------------------|-----|-----|-----|-----|----|---------|
| Number of times EM found that many clusters | 340 | 800 | 627 | 209 | 23 | 1       |

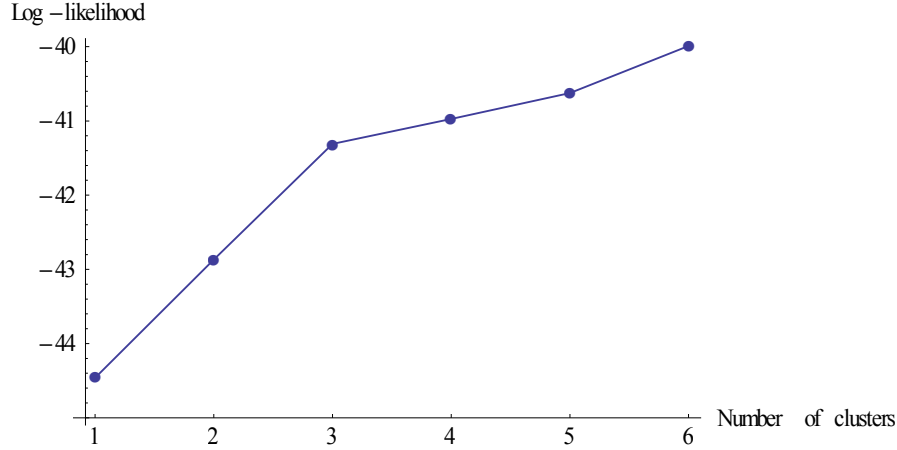


Figure 2. Mean log-likelihood associated to each clustering when applying EM algorithm with 2000 different initial seeds.

## 4.2 Clusters Characterization and Profiles

*4.2.1 Statistical evaluation of the clusters.* In order to characterize each cluster from the dominant partition with three clusters, we ran a MANOVA in which we tested whether the three clusters significantly differed on the 12 variables (treated as dependent variables in the MANOVA) that were entered during the clusters' formation. The results of an omnibus MANOVA, used to examine the relevancy of the clustering, indicated a significant, multivariate difference between the three clusters, Pillai's Trace = 1.48  $F(24, 62) = 7.33$ ,  $p < .01$ ,  $\eta_p^2 = .74$ , and therefore supported their formation and ability to organize participants' based on differences in their scores on 12 variables related to their learning with MetaTutor. Pillai's Trace criterion was used because Box's Test of Equality of Covariance Matrices was violated, Box's  $M$ ,  $F(78, 1984.66) = 1.88$ ,  $p < .01$ .

We looked at the follow-up ANOVAs to identify significant differences in participants' scores on the 12 variables between the clusters. Results presented in Table III indicated that significant differences existed between clusters for all of the 12 variables at the  $p < .05$  level and for 11 of them at the  $p < .01$  level (alpha levels presented below are rounded up to the second decimal point).

Table III. Summary of follow-up ANOVA results for the 12 dependent variables used in the cluster formation and pairwise difference for the three clusters

| Variables            | df    | F     | p      | $\eta_p^2$ | Pairwise difference<br>( $p < .05$ )? |     |     |
|----------------------|-------|-------|--------|------------|---------------------------------------|-----|-----|
|                      |       |       |        |            | 0&1                                   | 0&2 | 1&2 |
| ScorePre             | 2, 41 | 22.00 | 0.00** | 0.52       | X                                     | X   | X   |
| DurReading           | 2, 41 | 8.13  | 0.00** | 0.28       | X                                     | X   | X   |
| PropSGattempted      | 2, 41 | 10.07 | 0.00** | 0.33       | X                                     | X   | X   |
| NumSGChanges         | 2, 41 | 17.23 | 0.00** | 0.46       | X                                     | X   | X   |
| ScoreSGQuiz1stMean   | 2, 41 | 17.41 | 0.00** | 0.46       | X                                     | -   | X   |
| NumSGQuiz            | 2, 41 | 5.04  | 0.01** | 0.20       | -                                     | X   | X   |
| NumPageQuiz          | 2, 41 | 5.52  | 0.00** | 0.21       | -                                     | X   | X   |
| NumNoteTaking        | 2, 41 | 37.84 | 0.00** | 0.65       | X                                     | X   | -   |
| DurNoteTaking        | 2, 41 | 20.56 | 0.00** | 0.50       | X                                     | X   | -   |
| ScorePageQuiz1stMean | 2, 41 | 14.36 | 0.00** | 0.41       | X                                     | -   | X   |
| NumNoteChecking      | 2, 41 | 5.00  | 0.01** | 0.20       | X                                     | -   | X   |
| DurSession           | 2, 41 | 3.33  | 0.046* | 0.14       | -                                     | -   | X   |

\*  $p < 0.05$ , \*\*  $p < 0.01$

Levene's Test of Equality of Error Variances was violated for DurReading,  $F(2, 41) = 9.77$ ,  $p < .01$ , NumSGChanges,  $F(2, 41) = 3.68$ ,  $p < .05$ , NumNoteTaking,  $F(2, 41) = 8.82$ ,  $p < .01$ , NumNoteChecking,  $F(2, 41) = 9.15$ ,  $p < .01$ , DurNoteTaking  $F(2, 41) = 8.61$ ,  $p < .01$ , and DurSession,  $F(2, 41) = 6.98$ ,  $p < .01$ , therefore, a more stringent alpha level ( $p < .01$ ) was used in order to identify significant differences for these variables [Tabachnick and Fidell p.86 2007]. Several scores produced solution outliers and were therefore deleted from the analyses (e.g., possessed standardized residual scores exceeding +/- 3.29). These included one score for NumSGQuiz, ScoreSGQuiz1stMean, NumPageQuiz, and DurSession and three scores for DurNoteTaking.

Finally, we used Fisher's least significant difference (LSD) test to make pairwise comparisons between the different clusters for each of the dependent variables in order to determine between which of the three clusters the previously reported significant differences existed (cf. last 3 columns of Table III). These results tell us that generally, different dependent variables were useful in partitioning the clusters. More specifically, we can see that significant differences were identified for nine variables when clusters 0 and 1 were compared, eight variables when clusters 0 and 2 were compared, and ten when clusters 1 and 2 were compared.

Overall, as shown in Table IV, it appears that learners classified in cluster 2 had the highest values across the variables with the exception of durations (session, reading and note-taking) and number of notes taken, while those in cluster 1 generally scored the lowest and learners' scores in cluster 0 were more distributed, and sometimes represented a middle ground between clusters 1 and 2. Considering the clear distinction existing between the three clusters according to each variable used for their formation, and that a partition with three clusters provides more details over one with two, we will not report the analysis of the two clusters version. For the same reason, in the following section, we will exclusively consider the dominant partition in three clusters which has been analyzed in this section.

*4.2.2 Cluster Profiles.* The next step of our analyses was to look at the clusters' means and standard deviations for each of the dependent variables in order to create profiles for each cluster (see Table IV). The means allowed us to determine the direction of the previously reported significant pairwise comparisons. In this table, the mean scores have also been dummy-coded into high, medium, and low based on significant differences between clusters in order to heuristically characterize the differences between clusters. Therefore, a variable with significant pairwise comparison differences between all three clusters would have a low (L: lowest value), medium (M: middle value) and high (H: highest value) dummy code. Figure 3 provides a more graphical version of those results to facilitate visual comparison of the features distinguishing the 3 clusters.

The cluster profiles provide us with an understanding of three different 'types' of learners, based on twelve learner-driven variables. These three clusters also provide us with insight on how these variables varied between groups. In general, we saw that one 'type' (Cluster 2) of learner was characterized as scoring high on the pretest as well as on the first sub-goal quiz and first page quiz. Cluster 2 learners also spent relatively less time than others reading and taking notes (they also took few notes), though they did dedicate more time checking the notes they did take and to their sub-goals, attempting the greatest number, changing the sub-goals they were working on the most often and taking the greatest number of quizzes regarding their sub-goals. Relatedly, these learners also took the greatest number of page quizzes. Cluster 2 and Cluster 0 learners took less time to complete their learning session than Cluster 1. Given learners' high scores on quizzes and tests in this cluster and focus on their sub-goals we can think of them as the high performance monitoring group.

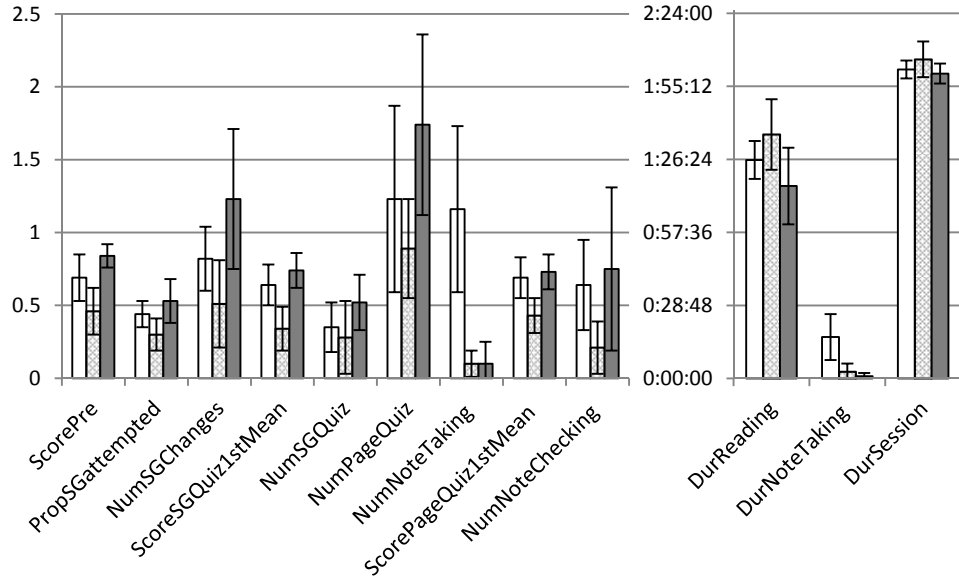


Figure 3. Means and standard deviations on dependent variables (per 10 minutes) for each cluster (white: cluster 0, dotted light grey: cluster 1, dark solid grey: cluster 2). Time-based variables use a different y axis.

Table IV. Summary of means and standard deviations on variables for each cluster as well as their dummy coded value (DC)

| Variables             | Clusters |           |           |          |           |           |          |           |           |
|-----------------------|----------|-----------|-----------|----------|-----------|-----------|----------|-----------|-----------|
|                       | 0        |           |           | 1        |           |           | 2        |           |           |
|                       | <i>M</i> | <i>SD</i> | <i>DC</i> | <i>M</i> | <i>SD</i> | <i>DC</i> | <i>M</i> | <i>SD</i> | <i>DC</i> |
| ScorePre              | 0.70     | 0.14      | M         | 0.43     | 0.15      | L         | 0.84     | 0.09      | H         |
| DurSession            | 2:01:51  | 0:03:30   | L         | 2:05:42  | 0:08:07   | H         | 2:00:97  | 0:04:03   | L         |
| DurReading            | 1:25:12  | 0:06:56   | M         | 1:35:07  | 0:15:53   | H         | 1:14:27  | 0:14:46   | L         |
| PropSGAttempted       | 0.44     | 0.10      | M         | 0.29     | 0.12      | L         | 0.54     | 0.16      | H         |
| NumSGChanges          | 0.83     | 0.22      | M         | 0.46     | 0.23      | L         | 1.28     | 0.46      | H         |
| NumSGQuiz             | 0.38     | 0.16      | L         | 0.29     | 0.27      | L         | 0.54     | 0.17      | H         |
| ScoreSGQuiz1st Mean   | 0.64     | 0.15      | H         | 0.34     | 0.17      | L         | 0.74     | 0.13      | H         |
| NumPageQuiz           | 1.29     | 0.66      | L         | 0.80     | 0.33      | L         | 1.75     | 0.65      | H         |
| ScorePageQuiz1st Mean | 0.67     | 0.13      | H         | 0.42     | 0.14      | L         | 0.74     | 0.12      | H         |
| NumNoteTaking         | 1.13     | 0.50      | H         | 0.12     | 0.09      | L         | 0.10     | 0.16      | L         |
| NumNoteChecking       | 0.65     | 0.28      | H         | 0.20     | 0.15      | L         | 0.76     | 0.59      | H         |
| DurNoteTaking         | 0:16:05  | 0:09:44   | H         | 0:03:22  | 0:03:19   | L         | 0:00:50  | 0:01:16   | L         |

Learners in Cluster 1 differed substantially from learners in Cluster 2 in that they performed significantly less well on the pretest as well as on first sub-goal quiz and first page quiz. This group also spent the longest reading and on the learning session. Also, these learners spent less time on their sub-goals, attempting the fewest, changing sub-goals the least number of times, and taking fewer sub-goal quizzes than Cluster 2. Similarly to Cluster 2 learners, Cluster 1 learners spent relatively little time taking notes (and also took few of them), but unlike Cluster 2, checked those notes less often. They also took fewer page quizzes than Cluster 2. Given these learners' low scores and high devotion of time to reading we can call these the low performance reading group.

Learners in Cluster 0 represented a third distinct profile, with mid-point (relative to Clusters 1 and 2) performances on the pretest, but high performance on the first sub-goal quiz and first page quiz, similar to Cluster 2. These learners also occupied a middle ground in terms of the time they spent reading, the proportion of sub-goals they attempted and the number of times they changed the sub-goal they were working on. Similar to Cluster 1, they took fewer sub-goal and page quizzes than Cluster 2 learners. Unlike either Cluster 1 or 2, these students spent a lot of time taking notes and took many of them. Similar, to Cluster 2, they spent more time checking the notes they took and less time completing their session. Following the performance and learning-behavior labels of clusters 1 and 2 we can call Cluster 0 the middle-point note-taking group.

#### *4.2.3 Online model evaluation*

In order to be able to apply the clustering proposed above within a future version of MetaTutor, it is necessary to evaluate the accuracy of prediction of the obtained clusterer for new instances of students. In order to do so, we performed a new EM clustering analysis (where the number of clusters was forced to 3 since it's the version we have chosen to focus on) using a 50-fold leave one out cross validation, i.e. we rebuilt the clusterer 51 times using 50 of the 51 participants and evaluated (with 1000 seeds) if the obtained clusterer was able to classify the remaining participant correctly. Overall, the clusterers classified correctly 78.8% of the instances, with however some important differences depending on the classes: 92.8%, 70.4% and 56.3% of students belonging to clusters 0, 1 and 2 respectively were classified correctly. Those values are to be compared to a baseline of 54.9%, if every participant was to be classified in the most-likely cluster (Cluster 0).

The main limits for an online implementation lie in the nature of the variables used in the clusters formation. The first one is the need for pretest score, which makes it mandatory to be kept, i.e. we cannot consider relying only on the quizzes given during the session to track the students' progress. Another one is the fact that some variables (such as the ones relative to sub-goals and page quizzes) are not available immediately as the students start their learning session. The classification of students would therefore be only possible after the students have used the system enough for those variables to all have a value (as for instance, after 5 minutes, it is likely that the number of sub-goals changes will be 0 for everyone). The adaptation of the system could potentially start half-way through the session (after approximately 1 hour) and be dynamic from there on.

#### 4.3 Application of the Clusters to Other SRL Processes and Students

##### *4.3.1 System-initiated SRL processes across user-derived clusters.*

Given that clusters were formed on the basis of user-initiated behaviors, we sought to investigate whether the system differentially prompted users to engage in SRL behaviors according to their cluster membership. To test if system-initiated SRL prompts differed between clusters, a MANOVA was conducted. Differences between three levels of the independent variable (user clusters) were tested on seven dependent variables<sup>4</sup> that comprised of the system-initiated prompts for specific SRL behaviors<sup>5</sup> (see Table I for system-initiated rules): Summarizing (SUMM), Re-Reading (RR), Coordinating Informational Sources (COIS), Prior Knowledge Activation (PKA), Judgment of Learning (JOL), Feeling of Knowing (FOK), and Content Evaluation (CE). For each SRL prompt that was found to statistically differ between user clusters, descriptive statistics are provided in Table V.

The omnibus MANOVA statistic was significant, Wilks  $\lambda = .51$ ,  $F(14, 84) = 2.38$ ,  $p < .01$ ,  $\eta^2 = .284$ , which indicated a multivariate difference between clusters on the seven system-initiated SRL prompts. A review of follow-up ANOVA tests showed group differences on three SRL prompts that were statistically significant: PKA,  $F(2, 48) =$

---

<sup>4</sup> Although nine variables of system-initiated SRL prompts were available for analysis, Planning (PLAN) and Monitoring Progress Toward Goals (MPTG) prompts were excluded. Data on participants' learning goal management were used in the formation of clusters, which are related to system rules to trigger PLAN and MPTG prompts, making tests for group differences tautological.

<sup>5</sup> As noted below Table I, SRL processes could be user- or system-initiated, however, for our analyses, only data on the frequency of system-initiated SRL prompts were included.

10.02,  $p < .001$ ,  $\eta^2 = .295$ ; SUMM,  $F(2, 48) = 6.52$ ,  $p < .005$ ,  $\eta^2 = .214$ ; and CE,  $F(2, 48) = 3.84$ ,  $p < .05$ ,  $\eta^2 = .138$ .

Descriptive statistics are reported in Table V and means are plotted in Figure 4. Post-hoc Bonferroni comparisons showed that for PKA and SUMM, differences were found between cluster 2 and clusters 0 and 1 ( $p < .05$ ), but no significant differences were observed between clusters 0 and 1 for these prompts ( $p > .05$ ). For CE, a statistical difference was only found between cluster 2 and cluster 0 ( $p > .05$ ).

In sum, participants in cluster 2 were prompted to a greater extent by MetaTutor to activate prior knowledge, summarize and evaluate the relevancy of instructional content compared to participants in cluster 0 or 1, similar to the direction of mean differences between clusters on user-initiated behaviors (cf. Table IV).

Table V. Means and standard errors for 10 minute rates of system-initiated SRL prompts by user clusters

| System-initiated SRL | Cluster           |           |                   |           |                   |           |
|----------------------|-------------------|-----------|-------------------|-----------|-------------------|-----------|
|                      | 0                 |           | 1                 |           | 2                 |           |
|                      | <i>M</i>          | <i>SD</i> | <i>M</i>          | <i>SD</i> | <i>M</i>          | <i>SD</i> |
| PKA                  | 0.65 <sup>a</sup> | 0.20      | 0.43 <sup>a</sup> | 0.12      | 0.87 <sup>a</sup> | 0.32      |
| SUMM                 | 0.49 <sup>a</sup> | 0.24      | 0.35 <sup>a</sup> | 0.15      | 0.75 <sup>a</sup> | 0.39      |
| CE                   | 0.44 <sup>b</sup> | 0.28      | 0.43              | 0.18      | 0.72 <sup>b</sup> | 0.34      |

<sup>a</sup> Cluster 2 > 0,1,  $p < .05$ ; <sup>b</sup> Cluster 2 > 0,  $p < .05$

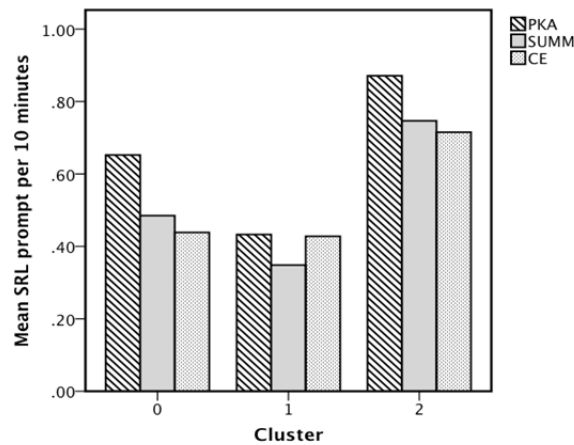


Figure 4. Means for each user cluster of the number of system-initiated SRL rules triggered during a normalized 10-minute interval of reading time within MetaTutor.

#### 4.3.2 Application of clusters to students in the Control (C) condition.

As explained in Section 3.4, the specificities of condition PF were the SRL processes prompts and the feedback on their performance that learners were receiving while learning with MetaTutor. Therefore, and in order to better characterize the differences between students belonging to each of the three clusters, we applied the partitions obtained in Section 4.1 with the EM algorithm (using a seed that provided the three clusters studied so far), to participants in the C condition. The goal was therefore not to compare our previous clusters with those obtained by an application of the EM algorithm to students in the C condition (since that clustering would be based on different criteria and not directly comparable), but to compare the distribution of students into the three clusters in the PF and C conditions. In particular, if one of the clusters ended up having a higher proportion of learners in the C condition (as opposed to the PF one), it would tend to provide support for the argument that the prompt and feedback condition “pushes” students toward the other two clusters. The results of this application are given in Table VI, where the repartition of students from the PF condition (mentioned in Section 4.1) is also given as a reminder.

It appears that students in condition C have a distribution very similar to the one of participants in the PF condition when classified according to the same classifier. There are slightly more students classified in cluster 0 and slightly less in cluster 1, but as there are no significant differences in terms of prompts to perform SRL processes received by students in those two clusters, we cannot draw any conclusion. Since we know from previous studies [Azevedo et al. 2012] that there are significant differences between participants in those two conditions, it only means that they are not distinguishable according to the set of variables considered here. This is confirmed by the fact that the log-likelihood value is much lower for participants in condition C (-56.02) than for those in condition PF (-41.31), which indicates that the clusters do not match very well the distribution of participants in condition C.

Table VI. Repartition of students from C and PF conditions within the 3 clusters

| Condition | Clusters |      |   |      |    |      |
|-----------|----------|------|---|------|----|------|
|           | 0        |      | 1 |      | 2  |      |
|           | N        | %    | N | %    | N  | %    |
| PF        | 28       | 54.9 | 9 | 17.6 | 14 | 27.5 |
| C         | 33       | 60   | 7 | 12.7 | 15 | 27.3 |

## 5. TOWARD ENHANCED ADAPTIVE RESPONSES BASED ON CLUSTERS

Our analyses suggest that the current design of MetaTutor responded differently to each cluster of participants. Specifically, when compared to clusters 0 and 1, participants in cluster 2 were prompted by MetaTutor to a greater extent to activate their prior knowledge (PKA), summarize (SUMM) and evaluate the relevancy of the content (CE) over the course of an average ten-minute period. This is a noteworthy finding given that the intended design of the experimental condition of MetaTutor was not to differentially scaffold participants, but instead to provide consistent scaffolding across individuals. In other words, although all participants had equal potential to receive the same number of system prompts at the start of each session, by the end of the session, based on the clusters of learner characteristics and behaviors, the system prompts they received ultimately differed. Why these differences occurred can be determined in light of the parameters governing how these SRL prompts are triggered by MetaTutor.

Specifically, the cause of these differences seems to be related to learners' behaviors, in particular the number of content pages they visit. Learners who visit more pages will receive more prompts to engage in PKAs, CEs, and SUMMs when they enter or exit the page (see Appendix B), whereas prompts for the other four SRL processes analyzed (RR, COIS, JOL, and FOK, see section 4.3.1) are triggered only after a certain amount of time has been spent reading the page. Therefore, we understand this difference in prompts as a difference in overall frequency of page visitations between clusters, but not necessarily as a bias in MetaTutor design favoring one type of cluster over another in SRL scaffolding. Such a difference in hypermedia navigation was perhaps due to participants in cluster 2, with higher content knowledge (see Table IV), being able to visit a page, quickly skim its contents and navigate away to another, thereby accessing more pages during their learning session. With this caveat in mind, the original design and intention of MetaTutor to provide learners (in the same experimental condition) with an equivalent, though adaptive, set of prompts is maintained.

Indeed, any differences in SRL prompts and associated benefits for learning that occur within the experimental condition are relatively minor compared to the differences between the experimental and control conditions. Considered within the larger context of research on self-regulated learning with MetaTutor, previous studies have consistently demonstrated higher posttest learning efficiency scores of content knowledge for students in the experimental condition who receive system prompts and feedback on SRL processes compared to students in the control condition who receive none [Azevedo et al.

2010b; 2012b]. Thus, regardless of the differences in system prompts between clusters of users within the experimental condition, overall the SRL scaffolds MetaTutor provides have a positive impact on content learning on average.

What the current findings reveal are insights into interactions between learner characteristics, system features, and the mediating SRL processes, which provide specific targets for future system improvements for increasingly adaptive, individualized support [Azevedo et al. 2010a; Aleven et al. 2010; VanLehn, 2011; Wolff 2009]. The success of such adaptive learning environments are predicated on the extent to which they can systematically and dynamically adapt their scaffolding of key SRL processes to the diverse needs of learners [Azevedo et al. 2012; Biswas et al. 2010; Graesser et al. in press; McGuiggan and Lester 2009; White et al. 2009]. In the current study, given how many rules for SRL prompts are organized around page visitations (see Appendix B), in future design iterations we can modify the probabilities of firing SRL prompts to better target and adapt scaffolding across diverse learners with complex profiles. We believe this insight highlights the use of educational data mining and machine learning methods to build agent-based system designed to detect, track, model, and foster students' self-regulated learning.

It must be noted that causal inferences cannot be drawn from these findings. First, the cluster formation was not experimentally manipulated, but instead was data driven. This method is highly informative of participant behavior with an agent-based system, but does not allow for experimenter control of many relevant variables, and therefore limits the causal conclusions that can be made. Second, given that MetaTutor is an adaptive environment, the directionality of causation is currently unknown. The systems' prompting will necessarily be triggered in response to participants' actions, yet once prompted, participants' subsequent actions are expected to be altered. Thus, rather than a direct line of cause-and-effect, the pattern of user behavior and system responses is better conceived as mutually reinforcing cycles.

A more general limitation of the current analysis is the fact that the system does not have a sophisticated student model, therefore most scaffolding and feedback mechanisms are only based on localized student behaviors. For instance, the probability to be prompted to perform a content evaluation when leaving a page quickly is the same throughout the learning session, regardless of how well the student has been applying this particular strategy while visiting previous pages.

## 6. RELATED WORK

Vellido et al. [2006] used clustering of multivariate data regarding students' behaviors in a virtual course in order to identify and characterize atypical students (outliers) and to estimate the relevance of available data features. Their general approach was therefore quite similar to ours, since here we have first worked on identifying clusters of similar students (and not particularly outliers) and then estimated the most relevant features of these clusters (using statistical analyses), but our context of an agent-based ITS is different from theirs (a virtual campus for students to learn online). In this work, Vellido et al. also demonstrated that the knowledge obtained from the cluster analysis could be fed back into their system to provide adapted guidance to their students, while the use of the clusters as an input of MetaTutor remains to be done. Tian et al. [2008] used both the learning strategies employed by the students as well as information regarding their personality to cluster them (an element of information that wasn't available to us here). Their methodology is also in two steps, since they validate their clusters definition through an analysis of frequent patterns. Similarly, works by Zakrzewska [2008], where students using a virtual classroom environment were clustered with a two-phase clustering algorithm using their individual characteristics and usability preferences. Manikandan et al. [2006] provided an interesting example of a virtual classroom system grouping students by performance, which is similar to our objectives here. Among the six variables they use for this purpose, half of them (memory retention ability, interestedness, prerequisite knowledge) require the existence of a glossary and of sequences of pages to be read in a particular order. The three other ones are comparable to some of our variables: marks in previous exams is a variable similar to ScorePre (although only one previous exam is considered in our case), read amount is based on the number of pages visited which is a parameter not directly taken into account here but which is correlated to NumPageQuiz and DurReading, and reading speed would correspond to a ratio of DurReading with the number of pages read (which we did not consider here).

In terms of clustering algorithms used, we can cite Teng et al. [2004], who grouped students according to their browsing behaviors using the EM algorithm, similarly to us. Their context was however different: as data had been collected in a virtual classroom environment, the information obtained from the clusters was directly provided to human teachers and it was up to them to empirically adapt their scaffolding to each group of students. On the contrary, in our case, it is mandatory to profile the clusters beforehand, as the adaptive scaffolding needs to be provided by agents from MetaTutor. In a similar

way, Talavera and Gaudioso [2004] also used EM to analyze students' behaviors, in the context of a collaborative virtual classroom environment.

Wayang Outpost [Arroyo et al. 2004; Ferguson et al. 2006] is an example of agent-based ITS with which researchers have used a Bayesian Network in order to infer positive or negative attitudes of students (collected through self-report measures), and evaluated the relationships between those attitudes and students' performance. We have collected, as mentioned in Section 3.3, information about emotions experienced by learners using MetaTutor, but this data wasn't used in the study presented here [Harley et al. 2011, 2012]. With Reading Tutor, Chang et al. [2006] used Dynamic Bayes Networks with parameters estimated with EM to model the students' knowledge and predict their performance. The context of Reading Tutor is however very different from ours, since it simply presents sentences to children who should read them, and the prediction was therefore about knowing if a child's word would or not be rejected, in a binary way. We, on the other hand, are more interested in the general performance of students and on the way they use learning strategies (such as SRL behaviors) than in predicting accurately if they are going to fail or succeed on the next quiz taken in MetaTutor. More similar to what we did here and related to our future directions (cf. Section 7), Amershi and Conati (2007) used both interface features and eye-tracking data to cluster learners using the k-means algorithm, and then built a classifier based on those clusters to perform an online supervised classification. The methodology followed in [Amershi and Conati 2009] to evaluate the potential practical use of the obtained clustering for an online classification to be used by a modified version of the learning environment would be particularly well-suited for our future needs: they consider the parameters used for the classification at different moments during the learning session, and check what percentage of students can be classified accurately (i.e., as labeled by the cluster algorithm). In our case, among the variables used for the clustering, only the learning session time wouldn't be available at any moment of the session for an online classification, and should therefore be dropped.

## 7. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we presented an analysis of data from college students learning about the human circulatory system with MetaTutor, in order to distinguish different classes of learners. Using data from participants in the Prompt and Feedback (PF) condition, we have shown (using the Expectation-Maximization algorithm) they could be classified into 3 different clusters, which could be organized by performance and learning behaviors.

Statistical analyses revealed that these profiles mainly differed in terms of performance, but also in terms of the amount of SRL processes they were prompted to engage in (although these parameters were not considered for the formation of the clusters).

Our results regarding system-initiated processes provided us with an opportunity to explore some possible differences regarding interactions and system-directed behaviors of learners who were in the PF condition. We found that learners in cluster 2 received the most prompts to engage in SRL processes, while learners in cluster 1 received the fewest and learners in cluster 0 were generally a middle point, though closer to cluster 1 than 2. Given the characteristics of Cluster 2, these results will allow us to modify certain rules so that we can minimize the number of prompts students in this cluster receive since we argue that students who are characterized as regulating their learning effectively should not receive such a greater number of SRL prompts. Such changes to the current system architecture will be necessary in order to enhance the learning and deployment of SRL processes on learners in the other two clusters.

One possible future direction is to use the clusters that have been defined and characterized in this article as input for a classifier to be used on-line (as opposed to the a posteriori only analysis done here), i.e. to be able to predict at any moment during the students' learning session with MetaTutor, the probability that they will be sorted into each cluster. Similarly to [Amershi and Conati 2009], it will also have to be considered at different moments during the learning session (e.g. after 10 minutes, after half of the session, etc.), to evaluate the possibility of profiling students as they learn with MetaTutor, in order to adapt the scaffolding to use SRL processes and the types of feedback provided by the agents. A major issue with such an approach is determining the ideal time episode since most models of SRL assume that SRL processes dynamically-unfold in real-time and that there are feedback cycles that impact SRL behaviors (see [Winne and Hadwin 2008]). Furthermore, such experimentation would augment current models of SRL by stipulating which and how SRL processes unfold in real-time and then be subsequently used to make instructional prescriptions to develop more sophisticated students models capable of providing more accurate and detailed individual instructional feedback and scaffolding.

In our current research, we have included collected data from several other channels not included in the data presented in this paper. For example, we collected additional process data including learners' basic and learning-centered emotions (from an analysis of video

recordings of their faces while they learned with MetaTutor) and gaze behavior (using an eye-tracker to examine learners' selection, organization and integration of multiple representations of information). We have also included self-report measures (i.e., questionnaires to measure learners' self-perception regarding their motivation and emotions) which may provide some useful additional information in order to either distinguish the clusters of learners defined here according to parameters of different nature, or to reapply the cluster extraction process described here using a different set of features, not unlike what is done in [Rodrigo et al. 2008], where emotion data comes from observation in the classroom.

These data hold great promise for helping us to improve our accuracy in detecting, modeling, and tracking CAM processes and have great predictive potential in terms of building a more sophisticated student model. However, despite the potential, several conceptual, theoretical, and instructional issues still need to be addressed. For example, current models and theories of SRL cannot predict how a multi-agent system should intervene if a student repeatedly inaccurately misjudges his understanding of the content (based on JOL prompts and log-file analyses), expresses frequent and prolonged bouts of frustration and confusion (based on facial expressions and behavior signatures from GSR data), tends to fixate on irrelevant text but fixates on relevant areas in diagrams (based on eye-tracking data), creates accurate and relevant sub-goals, and performs poorly on embedded quizzes.

Lastly, the unit of analysis for educational data mining and machine learning research that aims to improve the adaptive nature of ITSs still remains an issue. More specifically, the focus has been exclusive to analyzing learner behaviors within the learning environment (e.g., pretest scores, learners' frequency of self-regulatory behaviors, etc.) instead of adopting learner-system interaction cycles as the unit of analyses. The validity of EDM analyses can be compromised if the manner in which the system interacted with the learner is not taken into account. As such, current methods need to be augmented and emphasize learner-system interaction cycles as the unit of analysis. Similar issues and debates are currently being discussed in the educational and learning sciences literatures (e.g., see [Hadwin et al. 2011; Johnson et al. 2011]). As such, there is a need for researchers to be explicit when characterizing the type(s) of regulatory processes they study in their particular contexts. One future direction is for us and others to extend current conceptions of SRL to externally-regulated learning (ERL) within the context of students learning about complex science topics with MetaTutor which provides ERL

through its four pedagogical agents. While these characterizations are necessary for the field to advance they will be challenged by the contextual nature of learning systems (e.g., MetaTutor), contexts (e.g., solo lab studies vs. classroom peer learning vs. human tutoring sessions), and a myriad of other key issues (e.g., individual differences, internal standards, monitoring skills, emerging task understanding, etc.) that interact and change during learning.

## ACKNOWLEDGMENTS

This research was supported by funding from the National Science Foundation (0633918 and 1008282), the Social Sciences and Humanities Research Council of Canada (413-2011-0170), and funding from the Canada Research Chair program awarded to the last author. The authors would like to thank Reza Feyzi Behnagh, Melissa Duffy, Jonathan Burlison, Melissa Stern, Maroof Moral, Candice Burkett, Amber Strain, Amy Johnson, Ashley Fike, Erik Brooks, Timothy Crenshaw and Lea Hay for assisting in data collection, and Robert Bracewell for his feedback regarding dependent variable parameters.

## APPENDICES

### Appendix A: Macro- and Micro-Level SRL Processes and Associated Interface

#### Actions

| Macro-Level SRL Processes  | Micro-Level SRL Processes                 | Operational Definitions                                                                                     | Interface action to be performed by the learner                                                                |
|----------------------------|-------------------------------------------|-------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|
| <b>Planning</b>            | Planning (PLAN)                           | A plan involves coordinating multiple sub-goals.                                                            | Use of the sub-goals management buttons (cf. C in Figure 1)                                                    |
|                            | Prior Knowledge Activation (PKA)          | Searching long-term memory (LTM) for relevant prior knowledge.                                              | Use of the “Tell you what I already know about this” button in the palette                                     |
| <b>Monitoring</b>          | Monitoring Process Toward Goal (MPTG)     | Assessing whether previously set goal has been met.                                                         | Use of the sub-goals validation button (cf. C in Figure 1)                                                     |
|                            | Judgment of Learning (JOL)                | Student becomes aware that they do not know or understand everything they read.                             | Use of the “Assess how well I understand this” button in the palette                                           |
|                            | Feeling of Knowing (FOK)                  | Student is aware of having read something in the past and having some understanding.                        | Use of the “Evaluate how well I already know this content” button in the palette                               |
|                            | Content Evaluation (CE)                   | Monitoring content relative to goals.                                                                       | Use of the “Evaluate how well this content matches my current subgoal” button in the palette                   |
| <b>Learning Strategies</b> | Coordinating Informational Sources (COIS) | Coordinating multiple representations of information (e.g., drawing and notes, or text and diagrams).       | Click on the thumbnail image associated to each page of content to make it larger                              |
|                            | Inference (INF)                           | Making inferences based on what was read, seen or heard in the hypermedia environment.                      | Use of the “Make an inference” button in the palette                                                           |
|                            | Summarizing (SUMM)                        | Summarizing what was just read, inspected, or heard in the hypermedia environment.                          | Use of the “Summarize” button in the palette                                                                   |
|                            | Taking Notes (TN)                         | Copying text or elaborating on the text from the hypermedia environment.                                    | Use of the “Take Notes” button in the palette                                                                  |
|                            | Re-reading (RR)                           | Re-reading (text or diagram) or re-visiting a section (e.g., sub-topic page) of the hypermedia environment. | Use of the table of contents (cf. B in Figure 1) to visit a page already visited (which name appears in green) |

## Appendix B: System-generated rules

| Self-Regulated Learning Processes | Rules initiated by MetaTutor                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Action Sequence                                                                                                                                                                                                                                                                                                                  |
|-----------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| PLAN                              | <ul style="list-style-type: none"> <li>— Prompt the learner to add a new sub-goal when the session starts or after the three sub-goals set at the beginning of the learning session.</li> <li>— Ask the learner if they want to postpone their current sub-goal and move to a new sub-goal</li> <li>— Ask the learner to take the posttest at the end of the session</li> </ul>                                                                                                                                                                                                                                                                                                          | <p>Gavin: Greet and introduces Pam</p> <p>Pam: Asks learner to activate their prior knowledge (PKA)</p> <p>Pam: Asks learner to define 3 sub-goals</p> <p>Gavin: Leads learner into the learning session</p> <p>Pam: Informs learner that there are no more sub-goals and that they should add a new sub-goal to learn about</p> |
| PKA                               | <ul style="list-style-type: none"> <li>— When the learner starts their learning session, before setting sub-goals, they are asked to activate any previous knowledge about the circulatory system.</li> <li>— When the learner begins a new sub-goal, they will be prompted to provide any prior knowledge relevant to the sub-goal before starting to read the content.</li> <li>— When the learner enters a new page, they are prompted to provide as much information as they can about the current topic. This action occurs when the learner has encountered a relevant page within the sub-goal for the first time. This rule fires randomly for one out of four pages.</li> </ul> | <p>Pam: prompts the learner to activate the PKA for the new sub-goal</p> <p>Pam: give proper feedback based on sub-goal match</p> <p>Pam: ask learner to activate their PKA for the current page</p> <p>Pam: evaluate the input and gives feedback</p>                                                                           |

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| MPTG | <ul style="list-style-type: none"> <li>– The system asks the learner if they have adequately completed the given sub-goal, if they have spent too much time on a single sub-goal. This rule fires after the learner stays more than 20 minutes on a sub-goal.</li> <li>– When the learner has visited 100% of the pages related to the current sub-goal, the system asks them if they feel they have adequately completed the given sub-goal.</li> </ul>                                                                                                                                                                                              | <p>Mary: Tells learner believes enough time has passed for the current sub-goal and ask learner if they feel that they know enough about the sub-goal in order to complete it</p> <p>Mary: Gives a quiz (10 question) to evaluate student's knowledge</p> <p>Mary: If 60% of quiz is correct then mark the sub-goal as complete</p> <p>Mary: Tells student she believes enough of the current sub-goal has been covered and asks student if they wish to complete the sub-goal</p> <p>Mary: Give a quiz (10 question) to evaluate learner's knowledge</p> <p>Mary: If 60% of quiz is correct then mark the sub-goal as complete</p> |
| JOL  | <ul style="list-style-type: none"> <li>– The system prompts the learner to make a judgment about how well they understood the content after an appropriate amount of time. This rule fires if the student stays on a sub-goal-relevant page longer than the average reading time for that page.</li> <li>– When the current page is relevant and the learner changes the page after spending enough time to process some of the information from the page. This rule fires if the learner changes a sub-goal-relevant page after reading the page for at least 14 seconds, or after a minimum reading time relative to that specific page.</li> </ul> | <p>Mary: ask student to rate their level of learning</p> <p>Mary: gives a 3 question quiz on the current page</p> <p>Mary: gives feedback after quiz depending on student's self-rating and quiz results</p>                                                                                                                                                                                                                                                                                                                                                                                                                        |
| FOK  | <ul style="list-style-type: none"> <li>– The system asks the learner how well they already know the content they are reading. This rule fires when the page is relevant to the current sub-goal, and the student has read the page longer than a minimum of 57 seconds. Also, the probability for the firing of this rule is one out of three relevant pages.</li> </ul>                                                                                                                                                                                                                                                                              | <p>Mary: asks learner to rate their knowledge</p> <p>Mary: gives a 3 question quiz on the current page</p> <p>Mary: gives feedback after quiz depending on student self-rating and quiz results</p>                                                                                                                                                                                                                                                                                                                                                                                                                                 |

|      |                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                              |
|------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| CE   | — The agent prompts the learner to evaluate the appropriateness of the content on the page they are currently on after a sufficient amount of time to make the judgment. This rule is for when the page is relevant to the current sub-goal, and the student has read the page for longer than 14 seconds. Also, this rule fires for one out of five relevant pages. | Mary: asks learner if the page and image are relevant or not<br><br>Mary: gives proper feedback after that                                                   |
|      | — The learner is prompted after a sufficient amount of time reading the page to make a judgment. This rule fires when the page is irrelevant to the current sub-goal, and the learner has read the page for more than 14 seconds. The probability for initiation of this rule is one out of five irrelevant pages.                                                   | Mary: ask learner why they decided to change page so soon<br><br>Mary: reply to the response given or give short quiz (3 questions) if the case calls for it |
|      | — When the learner navigates away from a relevant page after reading the page for less than 14 seconds, the system asks them why they navigated away from the page after a short period of time, not having spent enough time to really process information of the page.                                                                                             |                                                                                                                                                              |
| COIS | — MetaTutor prompts opening the image for a learner who has been on a relevant page for longer than 45 seconds, and they has not opened the image associated with the current page.                                                                                                                                                                                  | Sam: suggests that the learner open the image that is associated with current page                                                                           |
| INF  | <i>Only triggered by an action from the learner</i>                                                                                                                                                                                                                                                                                                                  | N/A                                                                                                                                                          |
| SUMM | — MetaTutor prompts the learner who has read a relevant page for some time (proportional to page length), and has not opened the image associated with the page and is now navigating away to summarize the content on the page.                                                                                                                                     |                                                                                                                                                              |
| TN   | <i>Only triggered by an action from the learner</i>                                                                                                                                                                                                                                                                                                                  | N/A                                                                                                                                                          |
| RR   | — MetaTutor prompts the learner to re-read the contents of a relevant page, after they have spent enough time, and when the image is already opened. This rule fires for one out of four relevant pages, if the other conditions are also met.                                                                                                                       | Sam: suggests to re-read the page                                                                                                                            |

## REFERENCES

- ALEVEN, V., ROLL, I., MCLAREN, B.M., AND KOEDINGER, K.R. 2010. Automated, unobtrusive, action-by-action assessment of self-regulation during learning with an intelligent tutoring system. *Educational Psychologist*, 45 (4) 224-233.
- AMERSHI, S., AND CONATI, C. 2007. Unsupervised and supervised machine learning in user modeling for intelligent learning environments. *Proceedings of the 12th international conference on Intelligent user interfaces*, IUI '07 (pp. 72–81). New York, NY, USA: ACM.
- ARROYO, I., MURRAY, T., WOOLF, B., AND BEAL, C. 2004. Inferring Unobservable Learning Variables from Students' Help Seeking Behavior. *Lecture Notes in Computer Science*, Vol. 3220. Springer, Berlin Heidelberg New York, 782–784.

- AZEVEDO, R., AND ALEVEN, V, Eds. in press. *International handbook of metacognition and learning technologies*. Amsterdam, The Netherlands: Springer.
- AZEVEDO, R., BEHNAGH, R., DUFFY, M., HARLEY, J., AND TREVORS, G. 2012. Metacognition and self-regulated learning in student-centered learning environments. D. JONASSEN AND S. LAND, Eds. *Theoretical foundations of student-center learning environments* (2<sup>nd</sup> ed., pp. 171-197). New York: Routledge.
- AZEVEDO, R., CROMLEY, J.G., MOOS, D.C., GREENE, J.A., AND 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., JOHNSON, A., BURKETT, C., FIKE, A., LINTEAN, M., CAI, Z., AND RUS, V. 2010b. The role of prompting and feedback in facilitating students' learning about science with MetaTutor. R. PIRRONE, R. AZEVEDO, AND G. BISWAS, Eds. *Proceedings of the AAAI Fall Symposium on Cognitive and Metacognitive Educational Systems* (pp. 11-16). Menlo Park, CA: Association for the Advancement of Artificial Intelligence (AAAI) Press.
- 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., AND HOSSAIN, G. 2012b. The effectiveness of pedagogical agents' prompting and feedback in facilitating co-adapted learning with MetaTutor. S. A. CERRI ET AL., Eds. 11th International Conference on Intelligent Tutoring Systems, ITS2011. LNCS Vol. 7315. Springer, Berlin Heidelberg, 212-221.
- AZEVEDO, R., MOOS, D., JOHNSON, A., AND CHAUNCEY, A. 2010. Measuring cognitive and metacognitive regulatory processes used during hypermedia learning: Issues and challenges. *Educational Psychologist*, 45 (4), 210-223.
- BAKER, R.S.J.D., GOLDSTEIN, A.B., AND HEFFERNAN, N.T. in press. Detecting Learning Moment-by-Moment. To appear in *International Journal of Artificial Intelligence in Education*.
- BAKER, R.S.J.d., and YACEF, K. 2009. The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, 1 (1), 3-17.
- BISWAS, G., JEONG, H., KINNEBREW, J., SULCER, B., AND 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.
- CHANG, K., BECK, J., MOSTOW, J., CORBETT, A., AND A BAYES. 2006. Net Toolkit for Student Modeling in Intelligent Tutoring Systems. IKEDA, M., ET AL., Eds. 8<sup>th</sup> International Conference on Intelligent Tutoring Systems, ITS2006. LNCS Vol. 4053. Springer, Berlin Heidelberg New York, 104-113
- D'MELLO, S., AND GRASSER, A. in press. Dynamics of affective states during complex learning. *Learning and Instruction*
- D'MELLO, S.K., LEHMAN, B., SULLINS, J., DAIGLE, R., COMBS, R., VOGT, K., PERKINS, L., AND GRAESSER, A.C. 2010. A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. In *Proceedings of the 10th Annual Conference of Intelligent Tutoring Systems* (pp. 245-254) Pittsburg, USA: Springer.
- FERGUSON, K., ARROYO, I., MAHADEVAN, S., WOOLF, B., AND BARTO, A. 2006. Improving Intelligent Tutoring Systems: Using Expectation Maximization to Learn Student Skill Levels. M. IKEDA, K. D. ASHLEY, AND T.-W. CHAN, Eds. *Intelligent Tutoring Systems* (Vol. 4053, pp. 453-462). Berlin, Heidelberg: Springer Berlin Heidelberg.
- GRAESSER, A.C., D'MELLO, S.K., HU, X., CAI, Z., OLNEY, A., AND MORGAN, B. in press. AutoTutor. P. M. MCCARTHY, AND C. BOONTHUM, Eds. *Applied natural language processing and content analysis: Advances in identification, investigation and resolution*. Hershey, PA: IGI Global.
- GRAESSER, A.C., AND MCNAMARA, D.S. 2010. Self-regulated learning in learning environments with pedagogical agents that interact in natural language. *Educational Psychologist*, 45, 234-244.
- HADWIN, A., JARVELA, S., AND MILLER, M. 2011. Self-regulated , co-regulated, and socially shared regulation of learning. B. ZIMMERMAN AND D. SCHUNK, Eds. *Handbook of self-regulation of learning and performance* (pp. 65-84). New York: Routledge.
- HALL, M., FRANK, E., HOLMES, G., PFAHRINGER, B., REUTEMANN, P., AND WITTEN, I.H. 2009. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11 (1), 10-18.
- HARLEY, J., BOUCHET, F., AND AZEVEDO, R. 2011. Examining Learners' Emotional Responses to Virtual Pedagogical Agents' Tutoring Strategies. H.H. VILHJÁLMSSON, S. KOPP, S. MARSELLA, AND K.R. THÓRISSON, Eds. *Intelligent Virtual Agents, 11th International Conference*, Lecture Notes in Computer Science (Vol. 6895, pp. 449-450). Presented at the IVA 2011, Reykjavik, Iceland: Springer.
- KINNEBREW, J.S., AND BISWAS, G. 2011. Modeling and Measuring Self-Regulated Learning in Teachable Agent Environments. *Journal of e-Learning and Knowledge Society*, 7 (2), 19-35.

- LEELAWONG, K., AND BISWAS, G. 2008. Designing Learning by Teaching Agents: The Betty's Brain System. *Int. J. Artif. Intell. Ed.*, 18 (3), 181–208.
- MANIKANDAN, C., SUNDARAM, M.A.S., AND MAHESH, B.M. 2006. Collaborative E-Learning for Remote Education; An Approach For Realizing Pervasive Learning Environments. *2nd International Conference on Information and Automation (ICIA'2006)* (pp. 274–278). Colombo, Sri Lanka: IEEE.
- MCQUIGGAN, S., AND LESTER, J. (2009). Modeling affect expression and recognition in an interactive learning environment. *International Journal of Learning Technology*, 4 (3–4), 216–233.
- RODRIGO, M.M.T., ANGLO, E.A., SUGAY, J.O., AND BAKER, R.S.J. 2008. Use of Unsupervised Clustering to Characterize Learner Behaviors and Affective States while Using an Intelligent Tutoring System. *Proceedings of International Conference on Computers in Education* (pp. 49–56). Asia Pacific Society for Computers in Education.
- SHUTE, V.J., AND ZAPATA-RIVERA, D. 2012. *Adaptive educational systems*. P. DURLACH, Ed. *Adaptive technologies for training and education* (pp. 7–27). New York, NY: Cambridge University Press.
- TABACHNICK, B.G., AND FIDELL, L.S. 2007. *Using multivariate statistics* (5th ed.). New York: Harper Collins.
- TALavera, L., AND GAUDIOSO, E. 2004. Mining Student Data To Characterize Similar Behavior Groups In Unstructured Collaboration Spaces. *Proc. of the ECAI 2004 Workshop on Artificial Intelligence in Computer Supported Collaborative Learning* (pp. 17–22). Presented at the 16th European Conference on Artificial Intelligence, Valencia, Spain.
- TENG, C.-W., LIN, C.-S., CHENG, S.-Y., AND HEH, J.-S. 2004. Analyzing User Behavior Distribution on E-Learning Platform with Techniques of Clustering. *Society for Information Technology & Teacher Education International Conference, 2004* (1), 3052–3058.
- TIAN, F., WANG, S., ZHENG, C., AND ZHENG, Q. 2008. Research on e-learner personality grouping based on fuzzy clustering analysis. *12th International Conference on Computer Supported Cooperative Work in Design (CSCWD'2008)* (pp. 1035–1040). Xi'an, China: IEEE
- TOBII STUDIO™ (Version 1.5) [Computer Software]. 2009. Stockholm, Sweden: Tobii Technology.
- TOBII T60 [Apparatus]. 2009. Stockholm, Sweden: Tobii Technology.
- VANLEHN, K. 2011. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring system. *Educational Psychologist*, 46 (4), 197–221.
- VELLIDO, A., CASTRO, F., AND NEBOT, A. 2010. Clustering Educational Data. C. ROMERO, S. VENTURA, M. PECHENIZKIY, AND R.S.J.D. BAKER, Eds. *Handbook of Educational Data Mining, Data Mining and Knowledge Discovery* (pp. 75–92). CRC Press.
- VELLIDO, A., CASTRO, F., NEBOT, A., AND MUGICA, F. 2006. Characterization of atypical virtual campus usage behavior through robust generative relevance analysis. *Proceedings of the 5th IASTED international conference on Web-based education* (pp. 183–188). Anaheim, CA, USA: ACTA Press.
- VICARVISION. 2010. FaceReader (3.1) [Computer software]. Wageningen, The Netherlands: Noldus Information Technology.
- WHITE, B.Y., FREDERIKSEN, J.R., AND COLLINS, J. 2009. The interplay of scientific inquiry and metacognition. D.J. HACKER, J. DUNLOSKY, AND A.C. GRAESSER, Eds. *Handbook of Metacognition* (pp. 175–205). New York, NY: Routledge.
- WINNE, P., AND HADWIN, A. 1998. Studying as self-regulated learning. D. HACKER, J. DUNLOSKY, AND A. GRAESSER, Eds., *Metacognition in educational theory and practice* (pp. 227–304). Mahwah, NJ: Erlbaum.
- WINNE, P., AND HADWIN, A. 2008. The weave of motivation and self-regulated learning. D. SCHUNK & B. ZIMMERMAN, Eds. *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). Mahwah, NJ: Erlbaum.
- WINNE, P.H., AND NESBIT, J.C. 2009. Supporting self-regulated learning with cognitive tools. D.J. HACKER, J. DUNLOSKY, AND A.C. GRAESSER, Eds. *Handbook of metacognition in education*. Mahwah, NJ: Erlbaum.
- WOOLF, B. 2009. *Building intelligent interactive tutors: Student centered strategies for revolutionizing e-learning*. Burlington, MA: Morgan Kaufmann.
- ZAKRZEWSKA, D. 2008. Using Clustering Technique for Students' Grouping in Intelligent E-Learning Systems. A. HOLZINGER, Ed. *USAB'08, LNCS* (Vol. 5298, pp. 403–410). Berlin / Heidelberg: Springer-Verlag.