

Developing an agent-based adaptive system for scaffolding self-regulated inquiry learning in history education

Eric G. Poitras · Susanne P. Lajoie

Published online: 21 May 2014

© Association for Educational Communications and Technology 2014

Abstract This article presents a methodology for modelling the development of self-regulated learning skills in the context of computer-based learning environments using a combination of tracing techniques. The user-modelling techniques combine statistical and computational approaches to assess skill acquisition, practice, and refinement with the MetaHistoReasoning tool, a single-agent system that supports inquiry-based learning in the domain of history. Data were collected from twenty-two undergraduate students during a 4-h session where user interactions were logged by the system. A logistic regression model predicted user performance in relation to a skill categorization task with 75 % accuracy. The manner in which users apply the skills that are acquired is then assessed through a rule-based reasoning system that allows the pedagogical agent to adapt instruction. The results show that the model allows the agent to detect instances when skills are inappropriately applied as well as what type of goal that is pursued by students. We discuss the implications of these user-modelling techniques in terms of sequencing instructional content and using the tutoring agent to deliver several types of discourse moves in order to enhance learning.

Keywords User modelling · Inquiry-based learning · Pedagogical agent · Tracing system

Introduction

The self-regulation of learning and performance in academic settings is a defining characteristic of autonomous and competent students (Zimmerman and Schunk 2011). This paper outlines an approach for designing open-ended computer-based learning

E. G. Poitras (✉) · S. P. Lajoie

Department of Educational and Counselling Psychology, McGill University, Education Building,
3700 McTavish St., Montreal, QC H3A 1Y2, Canada
e-mail: eric.poitras@mail.mcgill.ca

S. P. Lajoie

e-mail: susanne.lajoie@mcgill.ca

environments (CBLEs) as metacognitive tools to promote the development of self-regulation by responding to the needs of different students (Azevedo 2005; Kinnebrew et al. 2013; Winters et al. 2008). This class of CBLE is characterized by the requirement of supporting students in performing ill-structured and complex tasks, where a large amount of learning material is available. As such, the ability to monitor and control different aspects of learning is a major concern in designing CBLEs.

Several models of self-regulated learning (SRL) skills have been put forth throughout the years, such as the social-cognitive model (Schunk 2005; Zimmerman 2000, 2006, 2008), the information processing model (Butler and Winne 1995; Winne 2001; Winne and Hadwin 1998, 2008; Winne and Perry 2000), and the four phase model of SRL (Pintrich 2000, 2004). Each model shares several common assumptions. Essentially, these models conceptualize SRL as a recurrent process that involves, among other things, the use of cognitive and metacognitive activities (Pintrich 2000; Zimmerman 2001). Self-regulated learners engage in metacognitive monitoring activities by judging their understanding of the learning content, self-questioning, and evaluating progress towards learning goals. Goal-setting activities concern the activation of task- or problem-relevant knowledge and devising the necessary steps to achieve goals. Controlling activities refers to the deployment of strategies to achieve learning goals, such as re-reading, summarizing, paraphrasing, and elaborating information.

Although these models explain SRL in terms of cognitive and metacognitive activities that are applicable across different subject matters, domain-specific accounts of SRL capture the skills that are particular to specific disciplines (Alexander et al. 2011; van der Stel and Veenman 2008). The three-phase model of cognitive and metacognitive activities in historical inquiry states that students engage in monitoring, setting goals, and controlling their own learning in accordance with disciplinary-based practices (Poitras and Lajoie 2013). In doing so, the model accounts for SRL in the domain of history with respect to problem-solving tasks as opposed to text-studying (for the later, see Meijer et al. 2006). Self-regulated learners monitor their own investigations into the causes of historical events by evaluating whether the causes that led to the occurrence of an historical event are well understood. Goal-setting involves intentional efforts to address lack of understanding by arguing in favour of a given cause, weighing alternative explanations, or responding to counter-arguments. In doing so, learning is controlled through the use of disciplinary-based strategies such as evaluating the trustworthiness of sources as well as gathering, corroborating, and contextualizing evidence.

Models of SRL skills can inform the design of CBLEs in order to support students in regulating their own learning. In particular, pedagogical agents are often embedded within such environments to serve as a tutor by interacting with students to provide feedback, hints, prompts, and correct misconceptions (Graesser 2011; Graesser et al. 2008). In doing so, agents are capable of personalizing instruction by capturing and analyzing user interactions and then selecting and delivering the most suitable instructional content (Brusilovsky 2012; Shute and Zapata-Rivera 2012). As an example, AutoTutor traces students' cognitive and metacognitive activities on the basis of their interactions with a computer tutor (Graesser and McNamara 2010; Graesser et al. 2005). The agent intervenes using several types of dialogue moves (e.g., prompts, questions, and corrections) that have been shown to predict positive gains on learning outcome measures (Graesser and McNamara 2010).

An alternative design choice is to embed multiple agents that are each assigned a specific instructional role (Baylor and Kim 2005). The iSTART system models and coaches students to engage in metacognitive activities while reading text by presenting two animated characters that talk to each other (McNamara et al. 2006). For instance, while one

of the agents demonstrates the use of reading comprehension strategies, the other provides feedback to students as they study and practice each strategy. In a large scale evaluation of learning with iSTART, the quality of students' self-explanations increased through extended amounts of practice across a variety of texts (Jackson et al. 2010). MetaTutor trains students in planning, monitoring, and strategy use with the help of four pedagogical agents that are each assigned to different self-regulatory processes (Azevedo et al. 2010). The agents deliver prompts on the basis of prior findings regarding the effectiveness of human tutoring strategies as well as the detection of specific conditions that characterize how students interact with features in the learning environment (e.g., prior knowledge, learning goals, time on task and instructional content) (Azevedo et al. 2011). The Betty's Brain system allows students to teach artificial agents and in the process, to indirectly assess their own learning through quizzes that are followed by feedback (Roscoe et al. 2013). The use of tools embedded as part of the interface was found to be predictive of learning outcomes.

In this paper, we claim that agent-based systems designed to foster SRL skills stand to improve learning through problem-solving in the domain of history. Students experience difficulties in regulating their own learning while studying historical events as they fail to monitor their own progress, set appropriate goals, and make necessary adaptations through the use of tactics and strategies (Greene et al. 2010; Poitras et al. 2012). As such, we examine how students regulate their own learning in the context of the MetaHistoReasoning tool (MHRt), a single-agent system that is designed as a metacognitive tool, by modelling how the requisite SRL skills are acquired, practiced, and refined.

The MetaHistoReasoning tool

The instructional objective of the MHRt is to scaffold the development of SRL skills that are critical in learning through problem-solving within the domain of history (Poitras and Lajoie 2012). These skills refer to the ability to appropriately engage in metacognitive activities, including (a) monitoring their understanding of the causes of historical events, (b) setting goals to guide their inquiries into these causes, and (c) controlling their own learning through the use of disciplinary-based strategies. The MHRt allows students to conduct their own investigations into the causes of historical events and provides explicit representations that support cognitive and metacognitive activities. As such, the design guidelines are meant to problematize the subject matter by allowing students to investigate a problem, while also providing the needed structure to perform the task and develop the relevant skills (see Reiser 2004).

To recreate the conditions in which SRL skills are required to enhance learning, the subject matter is made problematic by describing the circumstances that surround an historical event, but neglecting to explain the causes that led to the occurrence of the event. To do so, the MHRt provides a brief historical narrative of the event in question, which was previously revised by the experimenters to ensure that it fails to explain why the event occurred. The students are asked to search for the missing information by consulting historical sources, including both eye-witnesses and second-hand accounts. Solving this authentic problem requires the use of SRL skills due to the large amount of choices and materials that are made available to the students, notwithstanding the fact that sources may sometimes be incomplete or even contradictory. The MHRt foster the development of SRL skills through the design of the user interface, an explicit representation that influences relevant activities that occur during learning.

The design of the learning environment consists of a series of modules, where each module targets a different stage in skill development. The first module is referred to as the Training Module and it implements example-based skill acquisition as an instructional approach (Renkl 2010; Renkl et al. 2009). Students acquire SRL skills by studying examples of a learner who engages in the requisite cognitive and metacognitive activities involved in learning while performing inquiries into the causes of historical events. The artificial pedagogical agent prompts students to categorize each example in accordance with a list of SRL skills and to generate elaborative inferences (i.e., self-explanations regarding their purpose), in response, students receive corrective feedback. Based on student performance in categorizing examples, additional sets of examples are provided to students if they have not yet mastered the underlying skills. When students correctly identifies less than 70 % of all skills shown in the first four examples, for instance, an additional set of four examples is provided to students as a means to support them in acquiring these skills.

Once students have acquired the SRL skills, they move on to the Inquiry Module, which is designed based on structured inquiry-based learning principles (Hmelo-Silver et al. 2007; Krajcik and Blumenfeld 2006; Levsitk 2011; Loyens and Rikers 2011). This module is designed to allow students to practice and refine the SRL skills that were acquired in the previous module. In terms of monitoring their own understanding of the event, the pedagogical agent begins by asking students to explain why the event occurred. To evaluate their own progress, students rely on the explanation palette to rate their level of confidence in relation to several potential explanations, while the evidence palette displays a record of the arguments that were constructed throughout their investigations. Goal-setting is facilitated by decomposing each investigation in basic steps that are accomplished by students in a linear fashion. This is done in order to allow students to continually evaluate their own progress and to make necessary adjustments by arguing in favour of a potential explanation, considering alternative explanations, or refuting a counter-argument. Students control their own learning by using strategies to achieve their goals. At each step of solving the problem, instructional videos are made available to students, while the annotation tool enables students to write notes in accordance with custom templates, both of which are tailored to a particular strategy. For instance, a typical student first views the instructional video on how to evaluate the credibility of a source document, in order to rate the level of trustworthiness of a particular source in accordance with several criteria, and write a short justification for their choice that is subsequently stored in the evidence palette.

Research objectives and questions

This study aims to improve the adaptive capabilities of the pedagogical agent by modelling how students acquire, practice, and refine SRL skills that are critical in problem-solving within the domain of history. An agent-based tracing system should allow the system to assess skill development and deliver the most suitable instructional content. Based on the four-process adaptive cycle model of Shute and Zapata-Rivera (2012), we use the term *tracing system* to describe the computational processes that underlies how the pedagogical agent *captures and analyzes user profiles* in order to *select and deliver instructional content*. *User profiles* refer to the different patterns of interactions with system features that are generated by each user and recorded in the tracking log data. *User models* consist of the constituents of the tracing system, which determine how the pedagogical agent analyzes user profiles at each stage of skill development. The following questions will be addressed

in the results and discussion sections: (a) How does the agent-based tracing system allow the pedagogical agent to analyze user profiles in terms of the different stages of SRL skill development?; and (b) How should the pedagogical agent select and deliver instructional content in order to facilitate skill acquisition, practice, and refinement?

Methods

Participants

Twenty-two undergraduate students from McGill University (6 men and 15 women) with an average GPA of 3.3 ($SD = 0.52$; on a 4.0 scale) participated in the study. Participants were recruited through a classified ad posted on the university website. The following inclusion criteria were chosen to sample the participants in this study, based on the student population that is targeted by the design guidelines of the MHRt (i.e., version 1.1): (a) participants had to be native English speakers, (b) not currently enrolled in any history related program, and (c) unfamiliar with the topic to be learned (i.e., the Acadian Deportation, 1755–1763). Participants were compensated \$10 per hour for taking part in the study.

Historical topic and source documents

Causal chain of events in relation to the topic

Students studied the circumstances of a council meeting held in Halifax, in July 1755, where Governor Charles Lawrence gave the order to deport Acadians (see Faragher 2005). Researchers revised the narrative text that described the council meeting by excluding any information in regards to the causes and contributing factors that explained why Governor Charles Lawrence made the decision to begin the deportation. In doing so, the events referred to in the text were represented as a causal chain of events through the use of a replicable and principled method of discourse analysis (Fig. 1; see Montanero and Lucero 2011; Poitras et al. 2012). The chain of events in relation to the Acadian Deportation included 11 causal episodes and 12 links. The causal links were categorized by researchers as consisting of an intentional relationship (motive, desire, or goal that led to the event), a temporal relationship (chronologically ordered series of events), or a causal relationship (influential factors that explain why an event occurred). The causal episodes were reviewed and edited by two researchers until consensus was reached. The text was then edited so that no influential factors were apparent. This was done to induce confusion through knowledge deficit (see Lehman et al. 2012) in order to allow students to develop the SRL skills that are essential to manage and resolve these instances of confusion.

Source documents in relation to the topic

Students investigated the causes of the Acadian Deportation by analyzing six written and two pictorial sources. Half of these sources were primary sources, written by eye-witness accounts of the actual event, and the other half consisted of secondary sources that were written much later after the event had occurred and based on other accounts. The primary sources consisted of a letter and a memorial addressed to the Governor written by the

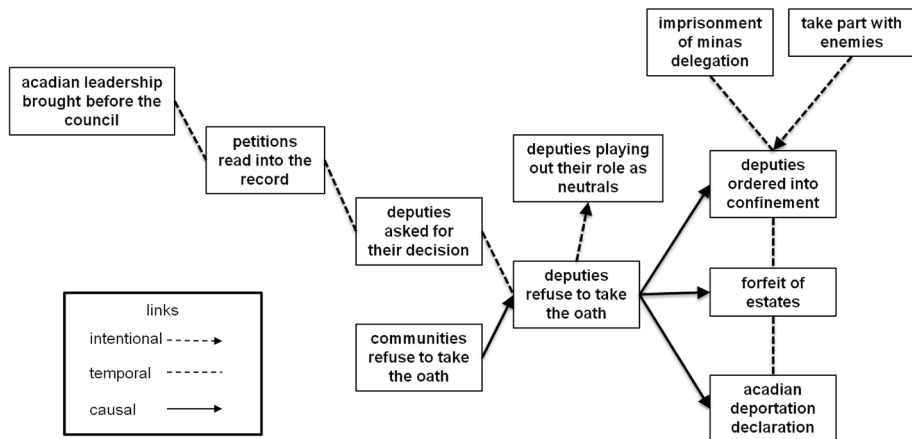


Fig. 1 Causal chain of events in relation to the council meeting held in Halifax, in July 1755, where Governor Charles Lawrence gave the order to deport Acadians

French inhabitants of Acadian communities and an official declaration and a memorial written by the British authorities. The secondary sources consisted of two paintings that illustrated scenes from the expulsion of the Acadians. A historical narrative text and expository text also served as second hand accounts of the event. The documents were selected with the assistance of museum curators in order to identify historical sources that included relevant information regarding possible explanations of the event in question.

The MHRt modules and features

In performing inquiries into the causes of the Acadian Deportation, students first acquired the necessary SRL skills with the benefit of the Training Module. Students began by studying an instructional video to gain basic declarative knowledge in relation to the Acadian Deportation as well as the SRL skills. The instructional video also introduced students to the different features of the interface. After viewing the video, students' studied examples of each SRL skill (see Fig. 2).

The examples were organized in five sets or groups. The first set consists of eight examples that illustrate a single SRL skill each. The second and third sets both include four examples; however, each example illustrates the use of three SRL skills. The fourth and fifth sets also contain four examples, but in this case, five SRL skills are shown in each example. The examples were thus shown in increasing order of complexity. In studying the simpler examples, the agent supported students by providing definitions and explanations (e.g., "This example shows an historian asking a question. In doing so, the historian begins to search for the most important cause of the Acadian Deportation."). Once the examples included three and five skills, the agent prompted students in categorizing each skill that were exemplified (e.g., "Which instance of historical thinking does this example show? Choose the option that best describes what the historian says."). Students were allowed to make as many attempts as was necessary in order to categorize the skill correctly by choosing amongst a list of eight different skills. Furthermore, the sequence in which example sets were delivered to students was determined based on their performance. Students only had to complete the third and fifth sets of examples if they failed to master

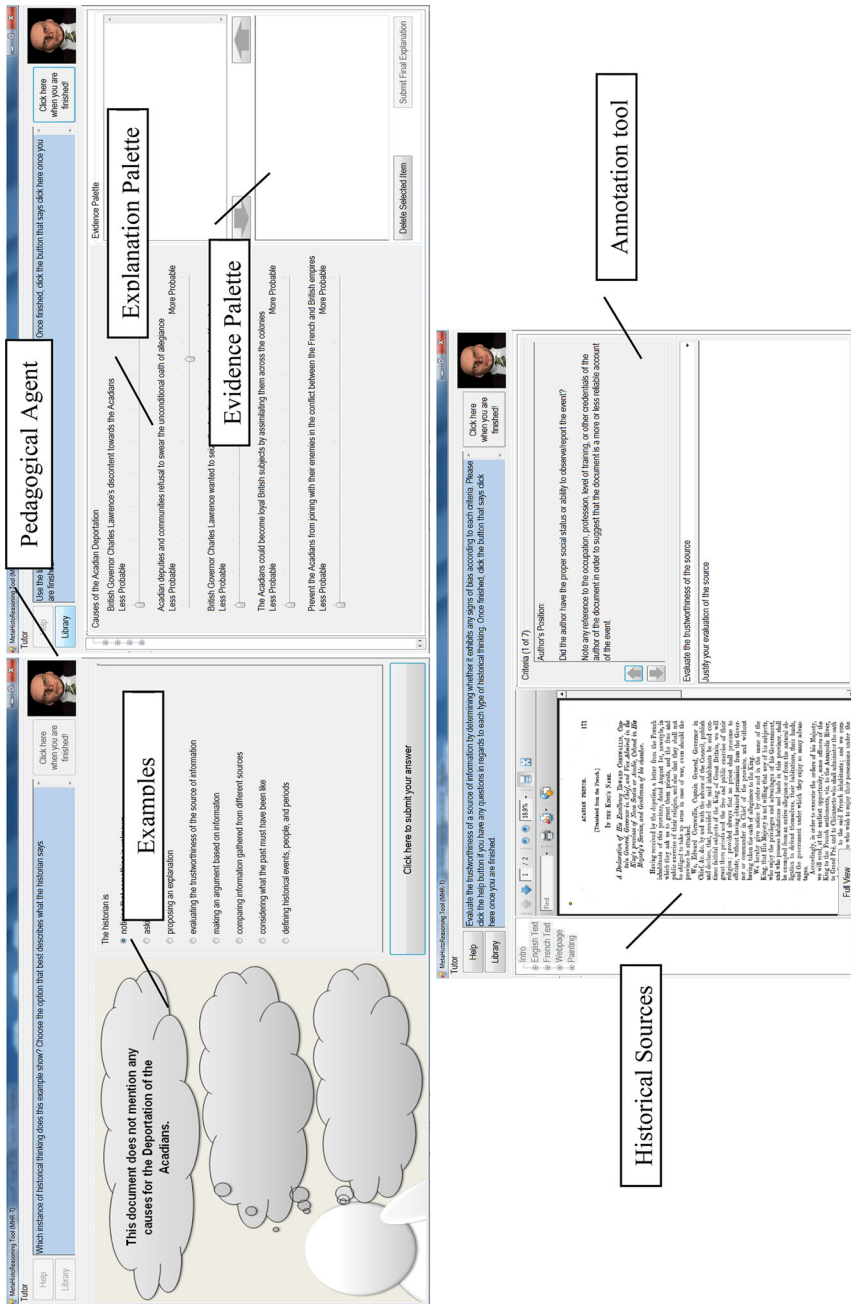


Fig. 2 User interfaces of the MetaHistoReasoning tool Training and Inquiry Modules

Table 1 SRL skills and heuristic examples covered in the Training Module

Strategy	Definition	Example
Noticing unexplained events	Acknowledging being confused or lacking knowledge in regards to an historical event, person, or time period	This document does not mention any causes for the Deportation of the Acadians
Asking appropriate historical questions	Asking questions in relation to historical events, persons, or time periods in order to search for specific information	What is the most important cause for the Acadian Deportation?
Formulating historical explanations	Proposing historical explanations by identifying multiple causes which lead to the occurrence of an historical event	The deportation was caused by the Acadian's refusal to swear the oath of allegiance
Evaluating the trustworthiness of historical sources	Determining whether a source of information (e.g., journal, eye witness account, history textbook, painting) provides a more or less reliable account of the event	Since the letter of the Acadian deputies was written during the time period and by knowledgeable participants in the event, it is considered as being a more trustworthy source of information regarding the Acadian Deportation
Gathering evidence	Making claims about the past (i.e., historical explanations that specify the causes of an historical event) that are supported by information gathered from sources	The letter states that the Acadians refused to swear the oath since they did not want to bear arms in case of War, as such, they were deported
Corroborating evidence	Searching for information that corroborates an account of an event by noticing similarities and differences amongst different sources of information	The fact that the Acadians refused to swear the oath was mentioned in both a letter and the transcription of a council meeting
Contextualizing evidence	Attempting to imagine themselves as being part of an event by describing in details certain aspects in relation to the time it occurred, its physical location, as well as other relevant common attitudes and cultural values	Acadians had close ties with the French; therefore, they must have felt apprehensive about going to war against them alongside the British
Using substantive concepts	Using concepts, in relation to historical events, persons, or time periods, to better understand and organize information in regards to the past	The Seven Year's War is the first global conflict which involved both French and British

the skills, as indicated by obtaining lower than an average performance of 70 % accuracy for the previous set. Table 1 includes a list of the SRL skills, including an example for each skill that was shown to the students.

Additionally, the agent helped students differentiate between each skill by providing corrective feedback in relation to their choices (e.g., “Your answer is correct.”; “Your answer is incorrect, try again.”). The agent also prompted students to elaborate on the underlying rationale for using these skills (e.g., “Explain how each instance of historical thinking relates to the historian’s goal, which is to explain why the Acadian Deportation occurred.”). Students wrote their self-explanations using a textbox that was shown next to the examples. Other features included brief descriptions of each skill that were made available to students as tooltips.

Whereas students were acquiring SRL skills in the Training Module, they got to practice and refine these skills in the Inquiry Module. As an example, students first acquired the ability to evaluate the trustworthiness of sources, and in the Inquiry Module they were

Table 2 Causal antecedents included in the Explanation Palette

Causal factor	Definition	Example
Personalistic	The influence of the intentions and motives of personal agents	British Governor Charles Lawrence's discontent towards the Acadians
Political	The political situation immediately prior to the event	Acadian deputies and communities refusal to swear the unconditional oath of allegiance
Economic	The economic situation existing at those times	British Governor Charles Lawrence wanted to seize the Acadians land, property, and livestock
Ideological	The influence of the ideological context	The Acadians could become loyal British subjects by assimilating them across the colonies
International policy	The contemporary international context	Prevent the Acadians from joining with their enemies in the conflict between the French and British empires

Note The causes were organized according to the framework proposed by Jacott et al. (1998), where they distinguish between intentionalist and structuralist explanations of historical events

asked to judge the credibility of sources against eight different criteria pertaining to the author or the document itself. Students were supported in using SRL skills while performing inquiries into the causes of the event, through the assistance of the pedagogical agent, a series of instructional videos, the annotation tool, the explanation and evidence palette, as well as a digital library.

Instructional videos were first shown to students in order to provide declarative knowledge in relation to skill. These videos explained the purpose and rationale of each skill, compared each of them, and illustrated how these are used with the system features.

The artificial pedagogical agent guided students in performing inquiries into the causes of the event by assigning sub-goals at each stage of the inquiry process. In doing so, the agent structured the students' investigation into the past by instructing them in using skills to formulate explanations. Moreover, the agent modelled the use of SRL skills by drawing their attention to the fact that the causes of the Acadian Deportation are unknown (i.e., "Read this text and you will notice that it does not explain why Charles Lawrence made the decision to deport the Acadians.") and asking an appropriate question (i.e., "What was the most important cause of the Acadian Deportation?").

At the beginning and end of each line of inquiry into the causes of the Acadian Deportation, students formulated and revised their explanations with the benefit of the explanation and evidence palette. Table 2 shows the different causes that were proposed to explain the occurrence of the Acadian Deportation. Students moved a slider along a trackbar to indicate whether a cause was believed to be less or more probable. On the right hand side, the students' notes (written as a result of the inquiry process) were recorded and displayed in a listbox referred to as the evidence palette. Students clicked on the "up" and "down" arrow to review and edit their notes.

During their inquiry into the causes of the Acadian Deportation, students evaluated the trustworthiness of sources and gathered, corroborated, and contextualized evidence obtained from historical sources with the benefit of the annotation tool. The annotation tool appeared on the right of the screen, while the historical sources appeared on the left. Students used the annotation tool to write notes while analyzing the historical sources. The written notes were not analyzed by the system for the purposes of adapting instruction. The annotation tool served two purposes: (1) It served as an external memory device since students' notes were recorded and could be consulted at a later time; (2) It restricted students' notes to specific configurations and subgoals set by the pedagogical agent.

The digital library included a wide range of factual and declarative information regarding the historical concepts and themes that were addressed in the source documents. The library provided students with information in relation to relevant historical figures (i.e., Charles Lawrence, Edward Cornwallis, Acadians, French Deputies, and Board of Trade and Plantations), the broader context surrounding the Deportation (i.e., the Seven Year's War), as well as policies that were enacted during that time period (i.e., Treaty of Utrecht, Oath of Allegiance, Deportation).

Measures

Pre-test measures

A list of demographic questions were used to characterize the sample and determine eligibility, including GPA, gender, age, years of study, and nature of the degree. Familiarity with the topic under investigation and with the domain of history was assessed with a questionnaire adapted from Hicks and Doolittle (2009). Students rated their knowledge of inquiry and the Acadian Deportation on a 5 point Likert scale ranging from very low (1) to very high (5). Also, students checked whether an item applied to them from a seven-option list (e.g., "I have an undergraduate degree in history, or a history related field"). The scores for knowledge of historical inquiry ranged from 1 to 12, while the scores for knowledge of the historical topic ranged from 1 to 5.

Process measures

Data were gathered on user interactions in the context of learning with the MHRt. Log-file entries and video-screen captures were collected unobtrusively while students learned about history (see Azevedo et al. 2010; Aleven et al. 2010). The log-file recorded user interactions with the interface features at a scale of milliseconds (10^{-4} s). An analysis of student performance on specific MHRt features was conducted including: students' categorizations of the heuristic examples, the sequence in which example sets were delivered, as well as the use of the explanation palette and annotation tool. The time-stamped video screen captures were used to corroborate log-file trace data by capturing the sequence of events and behaviours as they unfolded on the computer screen during task performance.

Post-test measures

Sentence recognition was assessed with a sentence verification task (Royer et al. 1996). Researchers have used this technique to assess students' understanding of the meaning of historical sources (see Wiley and Voss 1999). The sentence verification task consisted of 24 binary-response items, 12 true and 12 false (for example items, see Appendix A). In the sentence verification task, students were instructed to indicate whether the item was semantically similar to a sentence mentioned in one of the source documents. Reliability using the Kuder-Richardson 20 formula was calculated at .54. Content validity was established by randomly selecting sentences from each source document to ensure that the items were representative of the content that was covered in the learning environment.

Inference recognition was assessed with an inference verification task (Royer et al. 1996). Researchers have used this technique to assess students' understanding of inferences drawn from historical sources (see Wiley and Voss 1999). The inference verification task

consisted of 24 binary-response items, 12 true and 12 false (for example items, see [Appendix A](#)). In the inference verification task, students were instructed to indicate whether the item could be inferred from source documents. Reliability using the Kuder-Richardson 20 formula was calculated at .34. Content validity was established by randomly selecting sentences from each source document to ensure that the items were representative of the content that was covered in the learning environment.

Following Nokes et al. (2007), the students were asked to write an argumentative essay and answer four open-ended questions. To do so, students were provided with (a) a final report that summarized their notes taken while learning with the MHRt Inquiry Module and (b) an additional set of historical sources that were different from the ones provided while learning with the MHRt, but that still referred to the same topic. Students were instructed to write a 200-word essay with the aim of explaining the most important cause that led to the occurrence of the Acadian Deportation, the event under investigation. Students were then asked to rank the documents in terms of their reliability and usefulness while writing the essay. These answers were used by the researchers to revise the set of historical sources embedded in the MHRt.

Procedure

The following procedure was used to collect the data. The study was conducted over 1 day and lasted approximately 4 h. Prior to undergoing the study, a consent form was administered to students. Students were informed that all instructions would be provided through videos and that the experimenter would only answer questions in regards to the use of the features of the software. Students first completed the demographic questionnaire. Students then reported their knowledge of the topic and domain. Within the first 3 h, students received SRL training with the MHRt. Instructions was administered through the instructional videos and the pedagogical agent in the training and inquiry modules respectively. Students first completed the Training Module, and then moved on to the Inquiry Module.

During the last hour of the study, students completed the learning assessment tasks. The learning outcome measures assessed gains in knowledge about the topic and the arguments about the causes of the event under investigation, and were administered in the following order: (a) the sentence verification task, (b) the inference verification task, (c) the historical essay-writing task, and (d) the open-ended questions. At the end of the experiment, students were debriefed and compensated for their participation.

Results

Evaluating learning outcomes

Knowledge gained about the topic under investigation varied greatly across students. Before learning with the MHRt, students were not knowledgeable about the topic under investigation, since the average level of familiarity with the topic was quite low ($M = 1.14$, $SD = 0.48$, *Range*: 1–5). The mean accuracy score for the sentence verification and inference verification tasks suggest that students were moderately capable of recognizing sentences and inferences drawn from the source documents after learning with the MHRt. The students were able to recognize an average of 62 % of sentences ($SD = 11.7$ %, range: 41.7–87.5 %) and 69 % of inferences ($SD = 11.3$ %, range: 45.8–91.7 %). It is noteworthy to mention that both the range and deviation scores

obtained on these measures are quite large, suggesting that there are significant individual differences in the learning gains of students.

The range of argumentative techniques used to write the essays was examined by sampling the least and most elaborate essay, made up of the smallest and largest amount of characters (as shown in [Appendix B](#)). Before learning with the MHRt, students were unfamiliar with the domain of history and its practices, as the average level of familiarity with the domain was relatively low ($M = 3.33$, $SD = 2.03$, *Range*: 1–12).

On the one hand, in terms of their explanation of the event under investigation, both the least and most elaborate essay mention a similar cause, the refusal to swear the oath of allegiance, as an important factor in explaining why the event under investigation occurred. The most elaborate essay, however, challenges the neutrality of the Acadians in the conflict with the British government, mentioning that the Acadians had joined and supplied French soldiers at Fort Beauséjour. In doing so, the most elaborate essay mentions more than one influential factor in its explanation of the event.

On the other hand, the essays supports the explanation in a similar manner, as both the least and most elaborate essay refers to evidence drawn from the source documents. The most elaborate essay is different from the other in terms of the manner in which the evidence is described and evaluated. The most elaborate essay is more specific, relying on quotations to the actual sources as opposed to a general reference to re-occurring themes across documents. In addition, the most elaborate essay takes a critical stance by reflecting on the intentions of the author, explaining that Charles Lawrence may have been biased by writing the letter as a means to get the approval of the Lords of Trade. In summary, the exam of the argumentative essays and the learning outcome measures suggest that there is a great deal of variability in how students gained knowledge about the topic and reasoned about the causes of the event under investigation.

Modelling skill acquisition

Model variables

A logistic regression analysis was conducted to predict the odds of incorrectly categorizing examples of skills that were taught in the context of the Training Module. The model was fit to a total of 1,032 correct and 393 incorrect user categorizations. We excluded five outliers from the data. Users categorized each example by identifying the skill that was exemplified amongst a multiple-choice list with eight options. We first added the fixed parameters to the model, which consisted of several behavioural and contextual variables. In terms of user-behaviours, the system tracked the number of previous categorization attempts (i.e., *attempt* parameter) and the time taken to categorize the example from the moment when it is first shown to the user (i.e., *time* parameter). In terms of the contextual features that characterized the learning environment, the system captured the type of skill that was exemplified (i.e., *example category* parameter) and the frequency of previous exposure to each category of example (i.e., *exposure* parameter). We then added a subject variable to the model as a random effect using a scale identity covariance structure in order to take into account the independence of observations between users. Although the dependencies between each response could have been captured through a repeated measures effect, it was not possible to add it to the model due to the unequal amounts of categorizations performed by each user. However, fixed parameters such as the amount of prior exposure and previous attempts captured the dependencies between series of examples that demonstrated similar skills. Table 3 shows the model parameters, including

Table 3 Binomial logistic regression model parameters

Model parameter	Coefficient	Std. error	<i>t</i>	Sig.	Exp (coefficient)	95 % confidence interval for exp (coefficient)	
						Lower	Upper
Intercept	2.74	0.34	8.05	.00*	15.50	7.95	30.24
Skill category							
(A)	-1.87	0.33	-5.69	.00*	0.15	0.08	0.29
(B)	-1.01	0.36	-2.29	.01*	0.37	0.18	0.73
(C)	-2.32	0.33	-7.00	.00*	0.10	0.05	0.19
(D)	-1.53	0.34	-4.52	.00*	0.22	0.11	0.42
(E)	-0.48	0.37	-1.27	.20	0.62	0.30	1.30
(F)	0.02	0.41	0.04	.97	1.02	0.45	2.29
(G)	0.79	0.55	1.44	.15	2.19	0.75	6.42
(H)	0 ^a						
Exposure	0.04	0.02	2.17	.03*	1.04	1.01	1.07
Attempt	-0.20	0.05	-4.13	.00*	0.82	0.75	0.90
Time	-0.02	0.01	-2.99	.01*	0.99	0.98	1.00

Note (A) Gathering evidence; (B) Using concepts; (C) Contextualizing evidence; (D) Corroborating evidence; (E) Explaining events; (F) Noticing that causes are unknown; (G) Asking questions; (H) Evaluating the trustworthiness of sources

the regression coefficients, standard errors, significance tests, odds ratios, and confidence intervals.

Model evaluation

The category of skill that was exemplified was found to have a statistically significant effect on the class predictions generated by the model, $F(7, 1,414) = 18.61, p < .05$. The odds of correctly categorizing examples of skills were related to the following skill categories: (a) 0.098 times greater for contextualizing evidence, (b) 0.154 times greater for gathering evidence, (c) 0.217 times greater for corroborating evidence, and (d) 0.365 times greater for using substantive concepts. The number of attempts that were made to categorize each example was also shown to be predictive of the correctness of user categorizations, $F(1, 1,414) = 17.02, p < .05$. The odds of correct user categorizations were 0.823 times greater when users made an increasing number of attempts to categorize an example. The amount of exposure to each category of skill as well as the time that was taken by the users to categorize a particular example had a statistically significant effect on the predictions of the logistic regression model, $F(1, 1,414) = 4.713, p < .05$, and $F(1, 1,414) = 8.929, p < .05$, respectively. The odds of making a correct categorization are 1.035 times greater when the amount of prior exposure was increased, but only 0.985 times greater when the time taken to categorize a particular example was increased. These findings suggest that students who have acquired a particular skill made less prior attempts and more rapidly categorized an example. Furthermore, skill acquisition was less likely to occur when studying examples of using concepts as well as gathering, corroborating, and contextualizing evidence. Alternatively, skill acquisition was more likely to occur after repeated exposure to particular types of examples.

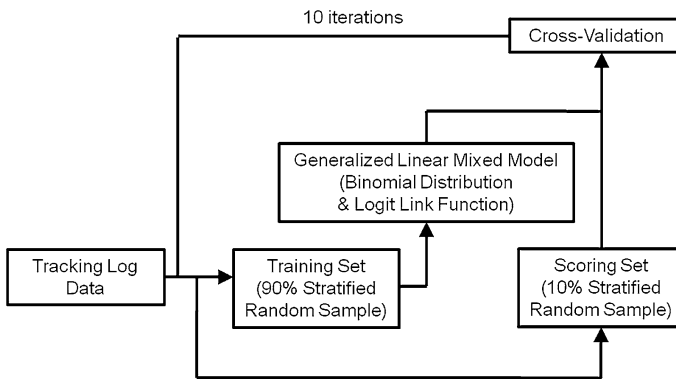


Fig. 3 The stratified 10-fold cross-validation evaluation process

Table 4 The original confusion matrix of the binary logistic regression model

Observed	Predicted		% Correct
	Incorrect	Correct	
Incorrect	122	271	31.0
Correct	73	959	92.9
% Correct	62.6 %	78.0 %	75.9

The performance of the prediction model was evaluated over ten folds of stratified cross-validation (see Fig. 3). The tracking log data is a system component that records user interactions with the features embedded in the MHRt. At each fold, the tracking log data was randomly partitioned into two equal size subsamples. Nine folds (approximately 929 correct and 354 incorrect user categorizations) served as the training data set in developing the logistic regression model, while one fold was used as the scoring data set to test the predictions of the model (approximately 103 correct and 39 incorrect user categorizations). The fold that served as the scoring data set, was changed over the ten iterations of the cross-validation evaluation. Table 4 provides an overview of the prediction outcomes of the logistic regression model that were obtained using the original tracking log data. In the following paragraphs, we outline the average prediction outcomes over the ten iterations of the stratified cross-validation procedure.

First, we evaluated the outcomes of the cross-validation procedure by calculating the average number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The results showed that an average of 99.5 correct user categorizations were accurately classified as correct ($SD = 2.1$, Min = 96, Max = 102) and 3.7 were inaccurately classified as incorrect ($SD = 1.9$, Min = 1, Max = 7). In regards to the incorrect user categorizations, an average of 8.7 were accurately recognized as incorrect ($SD = 2.1$, Min = 7, Max = 14) and 30.1 were inaccurately recognized as correct ($SD = 2.1$, Min = 25, Max = 32).

Second, we calculated the accuracy $((TP + TN)/(TP + TN + FP + FN))$, sensitivity $(TP/(TP + FN))$, specificity $(TN/(FP + TN))$, positive predictive value $(TP/(TP + FP))$, and negative predictive value $(TN/(FN + TN))$ of the prediction model (see Fig. 4). The prediction model achieved 76.2 % accuracy for recognition of user categorization correctness while studying examples of skills ($SD = 1.81$, Min = 73.2, Max = 78.9; 95 %

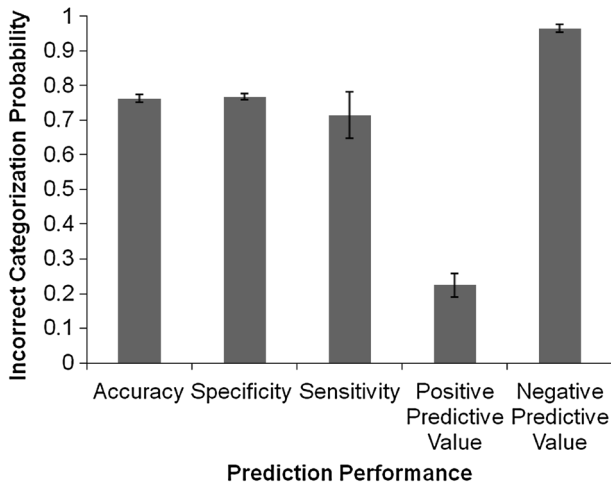


Fig. 4 The average performance outcomes of the stratified 10-fold cross-validation

confidence interval, 75.1–77.3). The prediction model achieved moderately high levels of sensitivity and specificity in that 71.4 % of incorrect ($SD = 10.86$, $Min = 53.3$, $Max = 87.5$; 95 % confidence interval, 64.69–78.16), and 76.8 % of correct classifications ($SD = 1.30$, $Min = 75.2$, $Max = 79.7$; 95 % confidence interval, 76.0–77.6) were accurately distinguished by the model. On the one hand, the negative predictive value of the model was substantially higher at 96.4 % of correct classifications that were accurately recognized by the model ($SD = 1.84$, $Min = 93.2$, $Max = 99.0$; 95 % confidence interval, 95.3–97.6). On the other hand, the positive predictive value was much lower at 22.4 % of incorrect classifications ($SD = 5.41$, $Min = 17.9$, $Max = 35.9$; 95 % confidence interval, 19.1–25.8). The outcomes of the stratified ten-fold cross-validation suggest that the prediction model performs reasonably well in estimating skill acquisition, albeit with a tendency to inaccurately recognize incorrect categorizations.

Modelling skill practice and refinement

Model variables

A rule-based reasoning system was developed to detect the type of goal set by students and whether strategies are applied in an appropriate manner. In a similar manner to Gulz et al. (2011), we designed the reasoning capacities of the agent using a design-based research methodology (see Collins et al. 2004). In doing so, each component of the architecture was designed, and will be iteratively revised, on the basis of the empirical evidence in regards to the strengths and weaknesses that are exhibited by users of the Inquiry Module. Figure 5 shows the components of the rule-based reasoning system.

The system classifies goal-setting and strategy use on the basis of 32 types of user interactions. For example, users wrote notes in a textbox and chose options from listboxes embedded in the annotation tool. These different features of individual user profiles, whether the users' behaviours (e.g., option selected) or discourse (e.g., length of characters), are extracted by the system and analyzed through the rule-based inference engine.

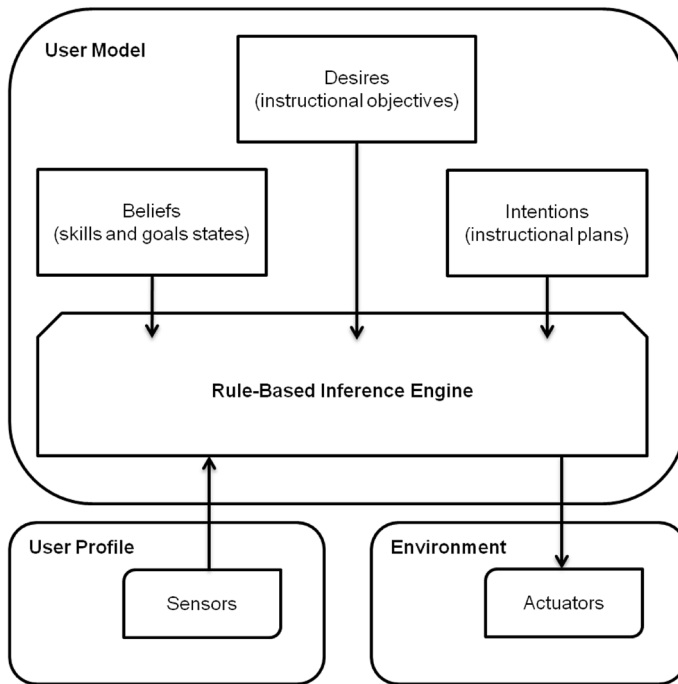


Fig. 5 Components of the rule-based reasoning system

The rule-based inference engine applied a series of IF-THEN decision rules for the purposes of classification. The recognition process determined the agents' *beliefs* in regards to the students' inquiries. Beliefs that were in direct contradiction to the agents' *desires* signaled the need to intervene through remedial activities. As such, the agents' desires are driven by the instructional objectives that were set by the instructor. In order to support users in achieving each instructional objective, the inference engine selected amongst different *intentions*. Agent intentions refer to instructional plans that determine what content is available and how it should be delivered to the student. Table 5 shows a sample of beliefs, desires, and intentions that were represented in the rule-based inference engine. In this paper, we limit our discussion of the agents' intentions in terms of the discourse moves that are selected and delivered by the pedagogical agent.

In detecting the types of goals that were set by users while performing lines of inquiries into the causes of historical events, the inference engine relied on features extracted at multiple time points. In doing so, a line of inquiry was defined as the process of formulating and revising an explanation for the event under investigation with the benefit of the explanation palette. During this process, students argued about their explanation with the help of the annotation tool by gathering and corroborating evidence. The system inferred when users confirmed their own explanation based on the analysis of the user profile when the following conditions were met. First, the claim for the argument in the annotation tool referred to the most likely cause for the event, as determined by the rankings of each cause in the explanation palette, both before and after performing a line of inquiry. Second, the grounds for the argument warranted the claim. Third, the grounds for the argument were corroborated by similar information obtained from additional sources. When these

conditions were met, the agent inferred that students had confirmed an explanation for the event, only later to suggest that students changed their goal to anticipate a counter-argument.

As an example, the system monitored the inquiry process by detecting whether users improperly engaged in certain skills. For instance, the user formulated an argument by stating a claim using the annotation tool. When a user failed to specify a claim, the beliefs of the agent were updated in that the system inferred that a skill was used inappropriately. Since the agent was programmed to desire that skills should be used appropriately, this conflict is resolved in that the intentions of the agent are updated. In this case, the agent may simply select the following discourse move to be delivered to the student—“*State your claim in relation to the most important cause for the event under investigation*”. As shown in Table 5, the degree of specificity of each rule varies as a function of what skill was targeted and how it was misused.

Model evaluation

A simulation of the rule-based decision system was performed with the tracking log data for 36 lines of inquiry into the causes of a particular historical event (from a total of 45 lines of inquiry since we excluded 9 due to missing data or technical errors). The tracking log data included the user rankings of each cause at the beginning and end of each line of

Table 5 Sample of rules taken from the inference engine

Features	States	Beliefs	Desires	Intentions (e.g., discourse moves)
Data	Is identical to previous line of inquiry Has less than 100 characters	Evidence is gathered inappropriately	Argumentation structure should be appropriate	Search for new evidence Elaborate further your argument for the explanation of the event under investigation
Warrant	Is not specified	Warrant is implicit		Indicate whether the factual information confirms or refutes your claim
Claim	Is not specified	Claim is implicit		State your claim in relation to the most important cause for the event under investigation
Claim and cause rankings	Claim is ranked as the most probable	Line of inquiry confirms an explanation	Multiples should be used while performing lines of inquiry	Okay, now that you have confirmed that “X” is the most probable explanation, now anticipate and respond to counter-arguments against your own position
Warrant	Claim is warranted by the data			
Corroborating sources	More amounts of corroborating information rather than dis corroborating information			

inquiry as well as the relevant annotations, including the drop down list values of the claims and warrants, the written factual information, and the checkbox values for each document that confirmed or refuted the grounds of the argument. The users each made an average of 2 lines of inquiry into the causes of the historical event ($SD = 1.9$). The average time duration of these lines of inquiries was 29 min and 38 s ($SD = 33$ min and 48 s).

First, the system reasoned about features that characterized the user profiles in order to identify instances when skills were used inappropriately. The system analyzed users' arguments in terms of whether the factual information was identical in two consecutive lines of inquiry and was insufficiently elaborated. The same arguments were also analyzed on the basis that users could fail to warrant as well as indicate the claim. In corroborating sources, the system detected whether users searched across the sources and found equal amounts of information that confirmed and refuted the claim, suggesting that the grounds of the claim were still uncertain. In a similar manner, the users' rankings of each cause were analyzed by the system, and if these were found to be similar, the explanation was judged to be uncertain. Finally, the system recorded the time when the explanation was first formulated and then revised in order to determine whether users' investigation was too superficial.

The result of the simulation shows that the rule-based reasoning system detected that students often misused skills while investigating the causes of an event. Two lines of inquiries were recognized as mentioning the same grounds as the ones that were previously used. Fifteen lines of inquiries included factual information that was not sufficiently elaborated. Although all the lines of inquiry indicated the claim of the argument, three lines failed to indicate whether the claim was more or less likely given the factual information. On six occasions the argument was found to be uncertain. Two of these lines referred to an equal amount of corroborating and dis corroborating information, while four of them rated multiple causes as the most important factors in explaining the event.

Second, the system reasoned about features that characterized user profiles with the aim of inferring the goal that was set by users. Goals were recognized and classified into one of three categories: confirming the most probable cause for the event, refuting counter-arguments, and weighing alternative explanations. In *confirming the most probable cause for the event*, users argued in favour of a claim by gathering factual information across several sources of information. Users *refute counter-arguments* by anticipating factual information that could be used against their own position and responding to the argument by searching for factual information that contradicted the evidence. Users *weigh alternative explanations* by formulating an argument that confirmed or refuted another claim or plausible explanation.

The results of the simulation of the rule-based decision system suggests that students engaged in confirmation bias, arguing solely in favour of their own claims while failing to consider alternative view points and respond to counter-arguments. Of the students who used skills appropriately, the system detected that those students performed lines of inquiry with the aim of confirming an explanation on every occasion. Table 6 provides a summary of these findings. In conducting lines of inquiries into the causes of the historical event, users had a tendency to confirm that the political situation was the most important factor that led to the occurrence of the event. Although this tendency could also be due to the fact that users all gathered evidence from the same source document while performing the first line of inquiry, users failed to pursue additional lines of inquiries. Therefore, this effect is not simply attributable to the availability of the factual information.

Table 6 Goal setting-activities detected by the rule-based reasoning system

Plausible explanation for the event under investigation		Lines of inquiry that confirmed an explanation			Lines of inquiry that anticipated counter-arguments		Lines of inquiry that weighed alternative causes	
Definition	Description	First line	Second line	Third line	First line		First line	
<i>Personalistic</i> : The influence of the intentions and motives of personal agents	British Governor Charles Lawrence's discontent towards the Acadians	0 (0 %)	0 (0 %)	0 (0 %)	0 (0 %)		0 (0 %)	
<i>Political</i> : The political situation immediately prior to the event	Acadian deputies and communities refusal to swear the unconditional oath of allegiance	8 (80 %)	1 (10 %)	1 (10 %)	0 (0 %)		0 (0 %)	
<i>Economic</i> : The economic situation existing at those times	British Governor Charles Lawrence wanted to seize the Acadians land, property, and livestock	0 (0 %)	0 (0 %)	0 (0 %)	0 (0 %)		0 (0 %)	
<i>Ideological</i> : The influence of the ideological context	The Acadians could become loyal British subjects by assimilating them across the colonies	0 (0 %)	0 (0 %)	0 (0 %)	0 (0 %)		0 (0 %)	
<i>International Policy</i> : The contemporary international context	Prevent the Acadians from joining with their enemies in the conflict between the French and British empires	0 (0 %)	0 (0 %)	0 (0 %)	0 (0 %)		0 (0 %)	

Discussion

The objective of this study was to model skill acquisition, practice, and refinement in order to develop a tracing system that allows the pedagogical agent to adapt instruction to the different needs of each user. We argued that there was a need to implement assessment mechanisms in the MHRt in order to allow the system to support users in using SRL skills that are particular to learning through historical inquiry. This objective is warranted by the great deal of variability in learning outcome results among students after learning with the benefit of the MHRt. Taken together, our findings not only outline the functioning of such a system, but also provide the grounds for its development and implementation. The agent-based tracing system relies on sensors that capture user interactions with the features of both modules in order to generate user profiles (see Fig. 6). We used two different approaches to analyze the user profiles and model the different stages of skill development.

First, a mixed-effects binary logistic regression model was fitted to the tracking log data recorded by the Training Module. The statistical technique used to model skill acquisition was adapted from Performance Factors Analysis (PFA), an adaptive data mining model that estimates knowledge acquisition and sequences instructional content in the context of solving well-structured problems (Pavlik et al. 2009). Empirical studies that have compared the predictive accuracy of PFA against other widely used techniques, including Bayesian Knowledge Tracing (Baker et al. 2008a, b), have demonstrated its effectiveness in predicting how students assimilate problem-solving skills (Gong et al. 2011). In a similar manner to PFA, we modeled skill acquisition based on user performance, using logistic regression to estimate accumulated learning across different types of skills as a function of practice. In doing so, the parameters of the prediction model are modified to better correspond with the tracking of log data recorded by the MHRt. While the model fit the data quite well, incorrect categorizations were often misclassified as correct ones.

Second, a rule-based reasoning system was used to analyze how users practiced and refined SRL skills while investigating the causes of historical events in the context of the Inquiry Module. The formalism was adapted from a rational agent-oriented system known as Procedural Reasoning System (PRS) and based on the Belief-Desire-Intention (BDI)

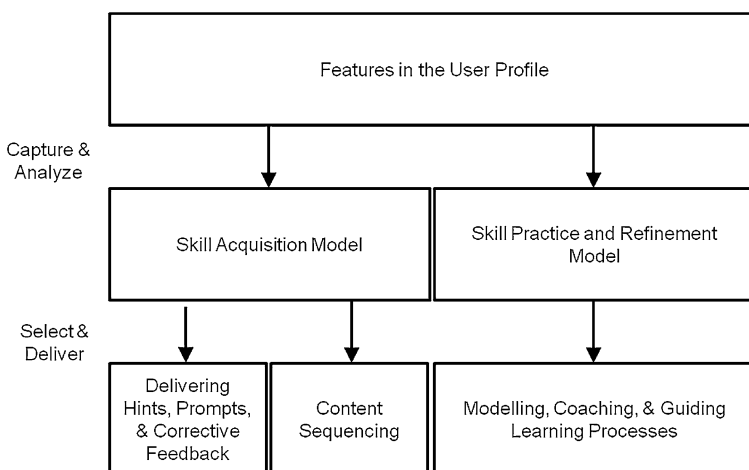


Fig. 6 Pedagogical agent tracing system of the MHRt

architecture (Ingrand et al. 1992; Rao and Georgeff 1995). PRS-based systems that use a BDI model have been applied to several domains: to monitor and control processes involved in student appraisals and emotions (Jacques and Vicari 2007); learning styles (Sun and Joy 2005); and collaboration patterns (Liu et al. 2006). The need for a system that recognizes goal-setting and strategy use was clearly established. Users often failed to use skills appropriately as well as to conduct investigations by setting goals not only to confirm their own points of view, but also to explore alternative explanations and refute counter-arguments.

Each profile is analyzed by the system in order to make inferences in relation to skill development. In doing so, the agent is capable of selecting and delivering the most suitable type of instruction. In the following sections, we raise several challenges in adapting instruction with the MHRt on the basis of individual differences in skill acquisition, practice, and refinement. Each section addresses a specific application of the models that were developed in the context of this study, namely: (a) delivering hints, prompts, and corrective feedback, (b) sequencing instructional content, and (c) modelling, coaching, and guiding learning processes. In support of our claims, we compare the non-adaptive system features with the adaptive ones.

Delivering hints, prompts, and corrective feedback

Hints, prompts and feedback serve as discourse moves for pedagogical agents that are delivered to students in order to facilitate skill acquisition. In training SRL skills, for instance, AutoTutor classifies students' answers to questions while solving physics problems in terms of either anticipated good answers or misconceptions. For example, hints (e.g., When the collision provides a large force to accelerate the head, what could happen to the neck?) and prompts (e.g., When the head and body are moving at the same acceleration, they are moving at the same _____?) are provided to students to help them articulate their own knowledge and cover all the expectations of what constitutes a good answer. Pedagogical agents use discourse moves that are determined based on what level of help a student needs given their performance at meeting the self-regulated model expectations (Graesser and McNamara 2010). Alternatively, MetaTutor and Betty's Brain employ dialogue prompts in order to assist students in using particular SRL skills, such as making an inference, with the help of features embedded in the system interface (Azevedo et al. 2010; Kinnebrew et al. 2013). These dialogue prompts (e.g., Can we double-check that these links are correct by reading the resources?) are delivered in response to patterns of user interactions detected in the system log files (e.g., Student has just added a number of links to the concept map). Feedback can be provided in relation to both knowledge construction and SRL skills. On the one hand, AutoTutor and MetaTutor provide feedback on the correctness of student answers and summaries. On the other hand, iSTART provides students with feedback in relation to the length and relevance of self-explanations, while MetaTutor provides feedback when goals are poorly stated (Graesser and McNamara 2010).

In studying examples of skills in the context of the MHRt Training Module, the pedagogical agent tracing system tracks user interactions with system features in order to build user profiles. Each profile captures individual differences in skill acquisition, and is continually updated through the tracking log data that is recorded by the system. These user profiles are analyzed through the mixed-effects logistic regression model that was fitted to the tracking log data obtained from previous users. The model thus enables the pedagogical agent to make inferences regarding the different levels of skill acquisition. Table 7

Table 7 Discourse moves of the pedagogical agent in the Training Module

System features	Discourse move category	Agent dialogue text-to-speech conversion
Adaptive version	Hints	
	Less detailed	This skill is used in the early stages of the inquiry process
	More detailed	In the early stages of the inquiry process, noticing instances of confusion and identifying missing information is critical to searching across sources of information
	Prompts	
	Specific elaborative	Why is it important to notice that the causes of the event are unknown?
Non-adaptive version	General elaborative	Explain the rationale for engaging in these skills; what is the underlying purpose, and how do these skills allow an historian to achieve this objective?
	Corrective feedback	
	Correct response	Your response is correct
	Negative response	Your response is incorrect, try again

provides an overview of the discourse moves that are used by the agent to support skill acquisition.

The tracing system complements the existing features embedded in the Training Module by allowing the pedagogical agent to adapt discourse moves based on the specific needs of each user. In the adaptive version of the Training Module, the pedagogical agent selects from three major types of discourse moves when users have difficulty acquiring skills, including hints, prompts, and corrective feedback. *Hints are designed to support users in the process of studying examples of skills that are difficult to acquire.* The agent delivers hints when there is a high probability that the user will incorrectly categorize an example. The agent selects a hint that contains the most amount of detail when the user profile suggests a larger than expected amount of previous attempts. *Prompts are meant to support users in assimilating each skill by elaborating on their underlying rationale.* General elaborative prompts support users in elaborating on the rationale for using a series of skills, as opposed to specific elaborative prompts that target isolated skills for elaboration. Although the probability of incorrectly categorizing an example is the main factor in delivering elaborative prompts, the amount of specificity decreases as the amount of prior exposure to each skill increases with practice. *Corrective feedback aims to support the users in differentiating the defining characteristics of each skill.* The agent provides users with feedback in relation to their responses after categorizing an example by indicating whether it is correct or incorrect. We explain how the tracing system allows the agent to adapt discourse moves in order to support skill acquisition by simulating a user profile.

In conducting the simulation, we calculated the probability of correctly categorizing an example of contextualizing evidence, the most difficult skill to acquire by students (see Fig. 7). The values of the time parameter were varied from 0 to 40 s with 2 s increment. We also set the other parameters of the model, averaging the number of attempts made by previous users (1.64 attempts) and amount of prior exposure (5.70 previous exposures). In evaluating the pedagogical agent tracing system, the threshold values that determine when

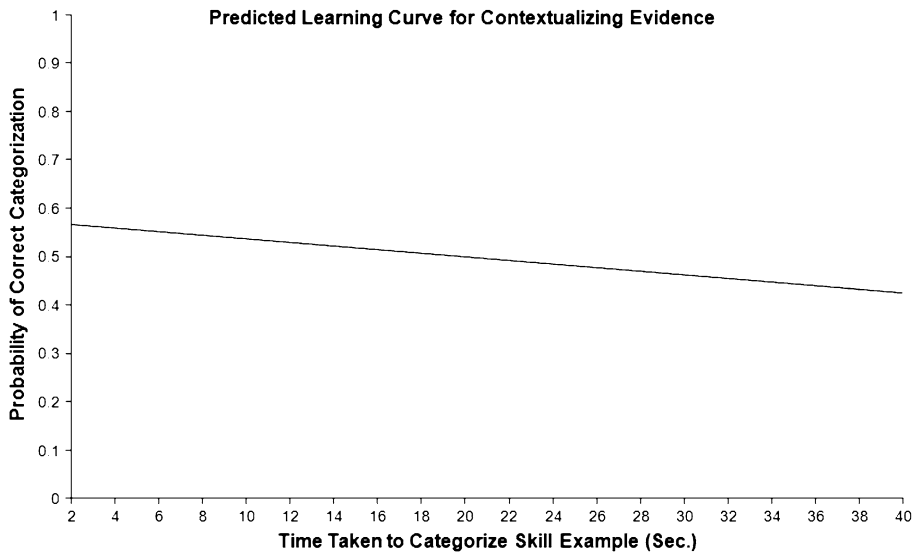


Fig. 7 Predicted learning curve for contextualizing evidence as a function of time taken to categorize skill examples

and how the agent intervenes can be manipulated in order to compare different instructional decision making processes. For instance, the threshold value regarding the amount of previous attempts can be decreased in order to allow the pedagogical agent to deliver more detailed hints. Specific elaborative prompts can also be provided more often by decreasing the value of the prior exposure threshold value.

In summary, the tracing system allows the pedagogical agent to determine what instructional content is most suitable as well as when it should be delivered to students in the Training Module. Hints, prompts, and corrective feedback are provided to students based on the specific characteristics of their own user profiles as well as the parameters of the skill acquisition model. In the next section, we discuss how the tracing system determines the best order in which each example should be provided to students.

Sequencing instructional content

Examples of SRL skills may be provided to students for training purposes prior to learning. For instance, MetaTutor involves a training regimen where pedagogical agents provide declarative knowledge in relation to self-regulatory processes that are critical to learning about science topics (Azevedo, Johnson, Chauncey, and Graesser 2011). Instructional videos of humans that model the use of these skills, are then studied by the students, who then have to complete both a discrimination task (i.e., distinguishing between instances when strategies are used correctly or incorrectly) and recognition task (i.e., identifying the strategies that are used by agents during learning).

In learning with the MHRt Training Module, the pedagogical agent tracing system facilitates skill acquisition by adapting the delivery of examples on the basis of the specific needs of each user profile. The current version of the system provides users with four sets of examples. The second and fourth sets are delivered on the basis of user performance on the first and third sets. When a user categorizes correctly more than 70 % of examples in a

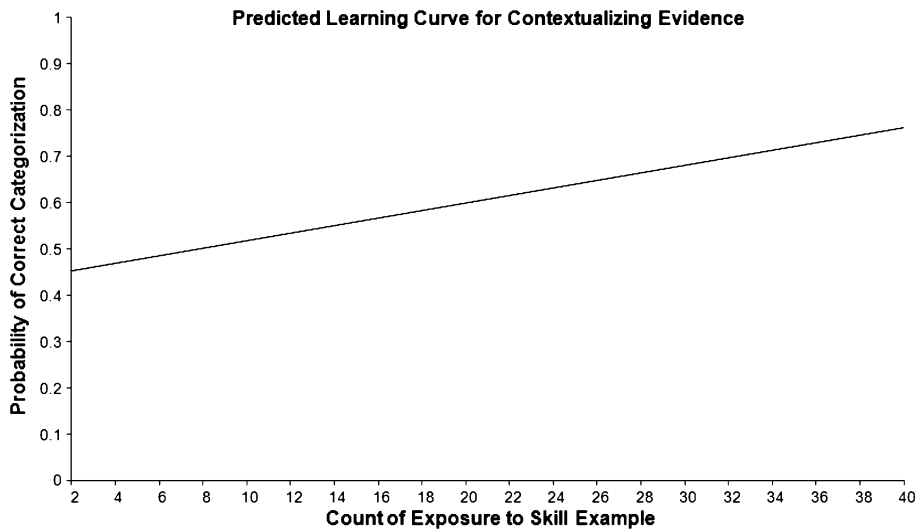


Fig. 8 Predicted learning curve for contextualizing evidence as a function of count of exposure to skill examples

set, the subsequent set of example is faded, and the user is thus allowed to progress more rapidly in the Training Module. We argue that this fading mechanism could be further improved by tailoring the delivery of examples based on the specific needs of each user, rather than individual differences in performance. More specifically, the system should deliver examples in order to ensure that users are more often exposed to skills that are particularly difficult to assimilate.

In support of our claim, we conducted a simulation where the probability of correctly calculating an example from the second set of was calculated (see Fig. 8). The values of the time and prior attempt parameters were set using the average values of previous users who completed the first set of examples (16.80 s and 1.43 attempts, respectively). The values of the amount of previous exposure were varied from 0 to 40 with an increment of 2. The results of the simulation support our claim that the instructional content should be sequenced on the basis of individual differences in user profiles. When users have been exposed to lesser amounts of examples, the probabilities of incorrectly categorizing an example varies across each skill category. However, the skill category had no effect on skill acquisition when the amount of exposure is increased. This finding suggests that the amount of exposure to each skill category should be tailored to each user, by sequencing each example to maximize the amount of exposure to the skills that are most difficult to acquire.

In summary, the adaptive version employs the tracing system to sequence each example in order to ensure that users are exposed to examples of skills that are the most difficult to acquire based on their own user profile. In contrast, the non-adaptive version of the Training Module fades entire sets of examples based on performance in categorizing examples. In doing so, the sequence of the instructional content fails to differentiate between each type of skill, and how specific skill categories may be more difficult for users to assimilate. In future studies, we will compare different algorithms to optimize the delivery of examples, thereby improving the efficiency of the Training Module.

Modelling, coaching, and guiding learning processes

Pedagogical agents can provide SRL training by modelling and coaching how to use skills as well as guiding students in setting goals and sub-goals. As an example, the iSTART Demonstration Module includes two agents called Merlin and Genie that model the use of skills by interacting with each other. Genie produces self-explanations while reading text sentences aloud, while Merlin coaches Genie in using strategies by providing feedback on its quality. In the Practice Module, the roles are reversed in that the student self-explains by using the different reading strategies while receiving feedback from the agent (Graesser et al. 2005). Furthermore, MetaTutor guides students in setting goals while studying text. Goal-generation is supported by the agent in that the instructor sets a major goal, and students are allowed to generate several sub-goals. These sub-goals are then recognized by the system, which asks students to clarify or elaborate them further, and this information is used to guide instructional decisions (Azevedo et al. 2010).

In learning with the MHRt Inquiry Module, the pedagogical agent tracing system analyzes how users perform inquiries into the causes of historical events in order to infer their goals. Each line of inquiry is classified by the system as confirming an explanation, responding to counter-arguments, and weighing alternative causes. When the system recognizes a particular goal, the agent can interact with students by selecting the most suitable discourse moves. Table 8 outlines the discourse moves that are delivered by the pedagogical agent to support users in practicing and refining skills.

The tracing system allows the pedagogical agent to respond to lines of inquiries performed by users by coaching skill usage as well as guiding goal-setting activities. In doing so, the tracing system complements the current features embedded in the Inquiry Module, which were limited to guiding sub-goals and modelling skill usage. *In coaching skills, the pedagogical agent trains users in employing skills appropriately, by drawing their attention to common mistakes and issues that need to be addressed.* The pedagogical agent intervenes by drawing their attention to the issue and explaining how to address it. The pedagogical agent selects from several discourse moves by identifying the particular issues that are faced by users through the rule-based reasoning system. *In guiding goal-setting, the pedagogical agent supports users in strengthening their arguments by varying the type of goal that they pursue throughout the inquiry process.* The agent prompts users in

Table 8 Discourse moves of the pedagogical agent in the Inquiry Module

System features	Discourse move category	Sample agent dialogue text-to-speech conversion
Adaptive version	Coaching skills	The explanation for the event under investigation is uncertain since the reliability of the factual information is low
	Guiding goals	
	Goals	Okay, now that you have confirmed that “X” is the most probable explanation, now anticipate and respond to counter-arguments against your own position
Non-adaptive version	Sub-goals	For each explanation, indicate whether it is more likely to be true or false given the evidence you have gathered so far
	Modelling skills	This text describes the council meeting where Charles Lawrence ordered to begin the Deportation of the Acadians. Read this text and you will notice that it does not explain why Charles Lawrence made the decision to deport the Acadians

shifting from one type of goal to another, for instance, the agent might detect that a user is confirming an explanation (e.g., “*Okay, now that you have confirmed that “X” is the most probable explanation*”) and then suggest an alternative (e.g., “*now anticipate and respond to counter-arguments against your own position*”). Each discourse move is selected by the agent on the basis of the rule-based reasoning system.

As such, the rule-based reasoning system allows the pedagogical agent to support students in using skills and setting goals while learning with the Inquiry Module. The system adapts instruction based on the particular features of different user profiles. The agent coaches students when skills are misused, while guidance is provided when students exhibit confirmation biases. The agent also models the use of skills, in particular, how to ask questions appropriately. Question asking is essential to investigating the past when reading a text that does not mention the causes that explain why certain events occur.

Conclusion

In summary, the findings obtained in this study demonstrate the use of different analytical techniques to model processes that are critical to skill development. These models guide efforts to develop and implement dynamic assessment mechanisms that are capable of capturing and analyzing user interactions with system features. The inferences that are drawn from the models allow the system to select and deliver instructional content that is the most suitable to the specific needs of different users.

Future research will compare the impacts of the adaptive and non-adaptive versions of the MHRt towards learning outcomes. The results obtained from the sentence verification task, inference verification task, and written summaries will establish a standard, against which the revised version of the system will be compared. We predict that the revised version of the system will assist students to engage in SRL skills more appropriately, in doing so, this should result in more accurate recognition of sentences and inferences as well as more sophisticated arguments in the written essays.

There are limitations to modelling skill development using binary logistic regression and rule-based decision approaches to classification; for example, the logistic regression model is excessively conservative in predicting classes of categorization accuracy. The model wrongfully classifies incorrect categorizations as correct ones. This is an important issue since students will often experience difficulties while studying examples of skills that may not be recognized by the system. Educational data mining techniques offer several alternative approaches to classifying binary data, including Naïve Bayes, decision tree, and rule induction classifiers (for review, see Hämmäläinen and Vinni 2010). We will conduct series of classification experiments to evaluate the performance of these models in comparison to the generalized linear mixed model. Furthermore, the rule-based reasoning system neglects the characteristic of the source documents in modelling skill practice and refinement. Recommender systems predict users’ preference for certain documents on the basis of their characteristics (Manouselis et al. 2011). Since searching across source documents is an important aspect of gathering and corroborating evidence, the rule-based decision approach to modelling skill practice and refinement could be made more comprehensive by integrating a recommendation function. We are currently in the process of developing such a system for directing users to certain sections in the digital library, and are making efforts to expand the model to include the source documents as well.

One remaining question is to improve these prediction models by integrating different types of models. Ensemble-based classifiers or systems use different methods to combine

several prediction models with the aims of improving prediction performance (for review, see Dietterich 2000; Kuncheva 2004; Polikar 2006; Rokach 2010). For instance, log file data includes notes that are written by students in the context of both the Training and Inquiry Module. These notes could provide valuable insights on a range of phenomena that impact skill development, such as how users elaborate on the examples of skills and interpret information obtained from source documents. Text mining techniques are commonly used in educational systems for the purposes of assessment, whether to classify self-explanations (McNamara et al. 2007) and essays (O'Rourke and Calvo 2009) in terms of their quality, or to recognize different types of affective states (D'Mello et al. 2007). Separate classifiers can be trained independently on the data obtained from these different information sources, and their predictions can be combined to form a meta-classifier in order to improve the inferences that are made by the tracing system.

Appendix A

Sentence verification task

Original item

"8. Having received by the deputies, a letter from the French inhabitants of this province, dated August 1st, newstyle, in which they ask us to grant them priests, and the free and public exercise of their religion, and also that they shall not be obliged to take up arms in case of war, even should the province be attacked."

Paraphrase item

"14. We hereby warn that by command and in the name of the King, that His Majesty does not want any of his subjects, who benefit from the rights and rewards of his Government, and who own habitations and lands in this region, shall be excused from a total commitment or from the expected duty to preserve themselves, their habitations, their lands, and the government under which they enjoy so many rewards."

Meaning change item

"6. Accordingly, in order to execute the orders of his Majesty, we will send, at the earliest opportunity, some officers of the King to the French settlements, viz. to the Annapolis River, to Grand Pré, and to Chicanecto who shall swear to the oath of allegiance to the said King; and we command all those who wish to enjoy their possessions under the happy government of his Majesty, to present themselves in order to administer the oath of allegiance before the 15/26 October, which will be the last day granted by us."

Distractor item

"4. It is His Excellency's order that this decree [oath of allegiance] be published in a small subset of departments as soon as possible, that no person may pretend ignorance of the same."

Inference verification task

Near inference, true item

“23. The dispute between the Acadians and Edward Cornwallis concerns the obligation, mentioned in the oath of allegiance, to take up arms in case of war, should the province be attacked.”

Near inference, false item

“5. Edward Cornwallis would allow the Acadians to refuse the oath should they consent to remain neutral and not fight alongside the French army.”

Far inference, true item

“3. The Acadians refused to swear to the oath of allegiance that was administered by the officers of the king at Annapolis River, Grand Pré, and Chicanecto.”

Far inference, false item

“15. The order to swear the oath of allegiance, which was published across all of the French settlements, was ignored by the French deputies.”

Appendix B

Elaborate argumentative essay

The Acadian Deportation of 1755 was a complicated event that involved many factors and motivations, thus it can be difficult to identify the most important cause. Nevertheless, I believe that the primary explanation for the Deportation was that the British suspected that the Acadians may not have been as neutral as initially thought and in fact were supporting the French in the war. This suspicion was based on the fact that the Acadian deputies and communities refused to swear the unconditional oath of allegiance, and furthermore, some of the Acadians were found to be secretly supporting the French. When the British first took control of the province of Nova Scotia, occupied by French descendents, they wanted the Acadians to take an unconditional oath, however the Acadians proposed a compromise, whereas they would not need to take up arms in the time of war; this was known as the unconditional oath. Later, when the 7 year war was beginning and there was a new Governor, Charles Lawrence, the British again asked the Acadians to take the unconditional oath, but the Acadians refused and wished to remain neutral in the war. In Doc 3 (Extract from a Letter of Governor Lawrence to Lords of Trade, 1754), the Governor argues that a large number of Acadians were working for the French and giving them supplies. I initially believed that Lawrence was only saying this so that he could get the approval of the Lords of Trade, however Doc 5 (The Deportation of the Acadians, 2011), confirms this fact by acknowledging that there were 270 Acadians found at the French Fort Beausejour when it fell. In conclusion, I believe that the best explanation for the Acadian Deportation was that the British suspected that the Acadians were loyal to the French and felt threatened by this. This suspicion was based off of the fact that the Acadians would not

swear the unconditional oath of allegiance and that there were several instances where they were found to be in fact supporting the French.

Shallow argumentative essay

Given that it is a recurring theme in many of the documents concerning the Acadians, especially those written by persons in positions of authority at the time that the events were unfolding, I believe that one of the main reasons the Acadians were deported or imprisoned was due to their refusal to swear the oath of allegiance to the British monarchy.

References

- Aleven, V., Roll, I., McLaren, B. M., & 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.
- Alexander, P. A., Dinsmore, D. L., Parkinson, M. M., & Winters, F. I. (2011). Self-regulated learning in academic domains. In B. Zimmerman & D. Schunk (Eds.), *Handbook of self-regulation of learning and performance*. New York: Routledge.
- Azevedo, R. (2005). Computer environments as metacognitive tools for enhancing learning. *Educational Psychologist*, 40(4), 193–197.
- Azevedo, R., Johnson, A., Chauncey, A., & Graesser, A. (2011). Use of hypermedia to convey and assess self-regulated learning. In B. Zimmerman & D. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 102–121). New York: Routledge.
- Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. D. (2010). Measuring cognitive and meta-cognitive regulatory processes during hypermedia learning: Issues and challenges. *Educational Psychologist*, 45(4), 210–223.
- Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education*, 15(1), 95–115.
- Brusilovsky, P. (2012). Adaptive hypermedia for education and training. In P. Durlach & A. Lesgold (Eds.), *Adaptive Technologies for Training and Education* (pp. 46–68). Cambridge: Cambridge University Press.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65, 245–281.
- Baker, R. S. J. d., Corbett, A. T., & Aleven, V. (2008a). More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In *Proceedings of the 9th international conference on intelligent tutoring systems* (pp. 406–415). Montreal, Canada.
- Baker, R. S. J. d., Corbett, A. T., & Aleven, V. (2008b). Improving contextual models of guessing and slipping with a truncated training set. In *Proceedings of the 1st international conference on educational data mining* (pp. 67–76). Montreal, Canada.
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences*, 13(1), 15–42.
- D'Mello, S. K., Picard, R. W., & Graesser, A. C. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*, 22(4), 53–61.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. In J. Kittler & F. Roli (Eds.), *First International Workshop on Multiple Classifier Systems* (pp. 1–15). New York, NY: Springer Verlag.
- Faragher, J. M. (2005). *A great and noble scheme: The tragic story of the expulsion of the french Acadians from their American homeland*. New York, NY: W. W. Norton.
- Gong, Y., Beck, J. E., & Heffernan, N. T. (2011). How to construct more accurate student models: Comparing and optimizing knowledge tracing and performance factor analysis. *International Journal of Artificial Intelligence in Education*, 21(1–2), 27–46.
- Graesser, A. C. (2011). Learning, thinking, and emoting with discourse technologies. *American Psychologist*, 66(8), 746–757.
- Graesser, A. C., Jeon, M., & Dufty, D. (2008). Agent technologies designed to facilitate interactive knowledge construction. *Discourse Processes*, 45, 298–322.
- Graesser, A. C., & McNamara, D. (2010). Self-regulated learning in learning environments with pedagogical agents that interact in natural language. *Educational Psychologist*, 45(4), 234–244.

- Graesser, A. C., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point & Query, AutoTutor, and iSTART. *Educational Psychologist*, 40(4), 225–234.
- Greene, J. A., Bolick, C. M., & Robertson, J. (2010). Fostering historical knowledge and thinking skills using hypermedia learning environments: The role of self-regulated learning. *Computers & Education*, 54(1), 230–243.
- Gulz, A., Haake, M., Silvervarg, A., Sjöden, B., & Veletsianos, G. (2011). Building a social conversational pedagogical agent—Design challenges and methodological approaches. In D. Perez-Marin & I. Pascual-Nieto (Eds.), *Conversational agents and natural language interaction: Techniques and effective practices*. IGI Global.
- Hämäläinen, W., & Vinni, M. (2010). Classifiers for educational technology. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. Baker (Eds.), *Handbook on educational data mining* (pp. 54–74). Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, CRC Press.
- Hicks, D., & Doolittle, P. E. (2009). Multimedia-based historical inquiry strategy instruction. Do size and form really matter? In J. Lee & A. Friedman (Eds.), *Research on technology in social studies education*. Greenwich, CT: Information Age Publishing.
- Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42(2), 99–107.
- Ingrand, F. F., Georgeff, M. P., & Rao, A. S. (1992). An architecture for real time reasoning and system control. *IEEE Expert*, 7(6), 34–44.
- Jackson, G. T., Boonthum, C., & McNamara, D. S. (2010). The efficacy of iSTART extended practice: Low ability students catch up. In V. Alevan, J. Kay, & J. Mostow (Eds.), *Proceedings of the Tenth International Conference on Intelligent Tutoring Systems (ITS 2010)* (pp. 349–351). Berlin: Springer-Verlag.
- Jacott, L., López-Manjón, A., & Carretero, M. (1998). Generating explanations in history. In J. F. Voss & M. Carretero (Eds.), *Learning and reasoning in history. International review of history education* (Vol. 2, pp. 294–306). London: Woburn.
- Jacques, P. A., & Vicari, R. (2007). A BDI approach to infer students' emotions in an intelligent learning environment. *Computers & Education*, 49(2), 360–384.
- Kinnebrew, J., Biswas, G., Sulcer, B., & Taylor, R. (2013). Investigating self-regulated learning in teachable agent environments. In R. Azevedo & V. Alevan (Eds.), *International Handbook of Metacognition and Learning Technologies*. Berlin: Springer.
- Krajcik, J. S., & Blumenfeld, P. (2006). Project-based learning. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 317–334). Cambridge: New York, NY.
- Kuncheva, L. (2004). Classifier ensembles for changing environments. In *Proceedings of the 5th international workshop on multiple classifier systems* (pp. 1–15). Springer.
- Lehman, B., D'Mello, S. K., & Graesser, A. C. (2012). Confusion and complex learning during interactions with computer learning environments. *The Internet and Higher Education*, 15(3), 184–194.
- Levstik, L. S. (2011). Learning history. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 108–126). New York, NY: Routledge.
- Liu, Z., Jin, H., & Fang, Z. (2006). Collaborative learning in e-learning based on multi-agent systems. In *Proceedings of the 10th international conference on computer supported cooperative work in design* (pp. 1–5). Nanjing.
- Loyens, S. M. M., & Rikers, R. M. J. P. (2011). Instruction based on inquiry. In R. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction*. Routledge Press.
- Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., & Koper, R. (2011). Recommender systems in technology enhanced learning. In L. Rokach, B. Shapira, P. Kantor, & F. Ricci, (Eds.), *Recommender systems handbook* (pp. 387–415). Springer.
- McNamara, D. S., Boonthum, C., Levinstein, I. B., & Millis, K. (2007). Evaluating self-explanations in iSTART: Comparing word-based and LSA systems. In T. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *LSA: A Road to Meaning* (pp. 227–241). Mahwah, NJ: Erlbaum.
- McNamara, D. S., O'Reilly, T., Best, R., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research*, 34, 147–171.
- Meijer, J., Veenman, M. V. J., & van Hout-Wolters, B. H. A. M. (2006). Metacognitive activities in text-studying and problem-solving: Development of a taxonomy. *Educational Research and Evaluation*, 12, 209–237.
- Montanero, M., & Lucero, M. (2011). Causal discourse and the teaching of history. How do teachers explain historical causality? *Instructional Science*, 39(2), 109–136.
- Nokes, J. D., Dole, J. A., & Hacker, D. J. (2007). Teaching high school students to use heuristics while reading historical texts. *Journal of Educational Psychology*, 99, 492–504.

- O'Rourke, S. T., & Calvo, R. A. (2009). Analysing semantic flow in academic writing. In *Proceedings of the international conference on artificial intelligence in education (AIED 2009)*. Brighton, UK.
- Pavlik, P. I., Cen, H., & Koedinger, K. R. (2009). Performance factors analysis—A new alternative to knowledge tracing. In *Proceedings of the 14th international conference on artificial intelligence in education* (pp. 531–538). Brighton, UK.
- Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientations in learning and achievement. *Journal of Educational Psychology*, 92, 544–555.
- Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16, 385–407.
- Poitras, E. G., & Lajoie, S. P. (2012). The MetaHistoReasoning tool: A metacognitive tool for fostering domain-specific metacognitive activities in learning through historical inquiry. In A. Weerasinghe, R. Azevedo, I. Roll, & B. du Boulay (Eds.), *Proceedings of the Workshop on Self-Regulated Learning in Educational Technologies: Supporting, Modelling, Evaluating, and Fostering Metacognition with Computer-Based Learning Environments* (pp. 41–44). Greece: Crete.
- Poitras, E. G., & Lajoie, S. P. (2013). A domain-specific account of self-regulated learning: The cognitive and metacognitive activities involved in learning through historical inquiry. *Metacognition & Learning*, 8(3), 213–234.
- Poitras, E. G., Lajoie, S. P., & Hong, Y.-J. (2012). The design of technology-rich learning environments as metacognitive tools in history education. *Instructional Science*, 40(6), 1033–1061.
- Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6, 21–45.
- Rao, A. S., & Georgeff, M. P. (1995). BDI agents: From theory to practice. In *Proceedings of the first international conference on multi-agent systems (ICMAS-95)* (pp. 312–319). San Francisco, CA.
- Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, 13(3), 273–304.
- Renkl, A. (2010). Instruction based on examples. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction*. New York, NY: Routledge.
- Renkl, A., Hilbert, T., & Schworm, S. (2009). Example-based learning in heuristic domains: A cognitive load theory account. *Educational Psychology Review*, 21(1), 67–78.
- Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1–2), 1–39.
- Roscoe, R. D., Segedy, J. R., Sulcer, B., Jeong, H., & Biswas, G. (2013). Shallow strategy development in a teachable agent environment designed to support self-regulated learning. *Computers & Education*, 62, 286–297.
- Royer, J. M., Carlo, M. S., Dufresne, R. J., & Mestre, J. P. (1996). The assessment of levels of domain expertise while reading. *Cognition and Instruction*, 14(3), 373–408.
- Schunk, D. H. (2005). Self-regulated learning: The educational legacy of Paul R. Pintrich. *Educational Psychologist*, 40, 85–94.
- Shute, V. J., & Zapata-Rivera, D. (2012). Adaptive educational systems. In P. Durlach & A. Lesgold (Eds.), *Adaptive Technologies for Training and Education* (pp. 7–27). Cambridge: Cambridge University Press.
- Sun, S., & Joy, M. (2005). An autonomous multi-agent solution for adaptive education. In *Proceedings of the 23rd IASTED international multi-conference artificial intelligence and applications*. Innsbruck, Austria.
- Van der Stel, M., & Veenman, M. V. J. (2008). Relation between intellectual ability and meta-cognitive skillfulness as predictors of learning performance of young students performing tasks in different domains. *Learning and Individual Differences*, 18, 128–134.
- Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, 91, 301–311.
- Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 153–189). Mahwah, NJ: Erlbaum.
- Winne, P., & Hadwin, A. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in Educational Theory and Practice* (pp. 279–306). Hillsdale, NJ: Erlbaum.
- Winne, P., & Hadwin, A. (2008). The weave of motivation and self-regulated learning. In D. Schunk & B. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). Mahwah: Erlbaum.
- Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). San Diego: Academic.

- Winters, F. I., Greene, J. A., & Costich, C. M. (2008). Self-regulation of learning within computer-based learning environments: A critical analysis. *Educational Psychology Review*, 20, 429–444.
- Zimmerman, B. J. (2000). Attainment of self-regulation: A social cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation research, and applications* (pp. 13–39). Orlando, FL: Academic Press.
- Zimmerman, B. J. (2001). Achieving academic excellence: A self-regulatory perspective. In M. Ferrari (Ed.), *The pursuit of excellence through education* (pp. 85–110). Mahwah, NJ: Erlbaum.
- Zimmerman, B. (2006). Development and adaptation of expertise: The role of self-regulatory processes and beliefs. In K. Ericsson, N. Charness, P. Feltovich, & R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 705–722). New York: Cambridge University Press.
- Zimmerman, B. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45, 166–183.
- Zimmerman, B., & Schunk, D. (Eds.). (2011). *Handbook of self-regulation of learning and performance*. New York, NY: Routledge.

Eric G. Poitras is a Postdoctoral researcher at the Learning Environments Across Disciplines (LEADS) Research Partnership, a Social Sciences and Humanities Research Council of Canada Partnership Grant.

Susanne P. Lajoie is a First Tier Canadian Research Chair of Advanced Technologies for Learning in Authentic Settings and Director of the Learning Environments Across Disciplines (LEADS).