Workshop on Metacognition and Self-Regulated Learning in Intelligent Tutoring Systems

Supplementary Proceedings of the 13th International Conference of Artificial Intelligence in Education. Marina del Rey, CA. USA. July 2007
Workshop on Metacognition and Self-Regulated Learning in ITSs

Ido ROLL, Vincent ALEVEN, Roger AZEVEDO, Ryan S.J.d. BAKER, Gautam BISWAS, Cristina CONATI, Amanda HARRIS, Rose LUCKIN, Tanja MITROVIC, Tom MURRAY, Barbara WHITE, Phil WINNE

*Carnegie Mellon University, University of Memphis, University of Nottingham, Vanderbilt University, University of British Columbia, University of Sussex, London Knowledge Lab, University of Canterbury, University of Massachusetts, Amherst, University of California, Berkeley, Simon Fraser University

1. Introduction

A number of studies in educational psychology and cognitive science have shown that students who apply better metacognitive skills and self-regulation strategies during their learning process have better learning outcomes. Educational technology such as intelligent tutoring systems has proven to be effective at the domain (or cognitive) level but similar success has not been achieved yet with regard to tutoring better metacognitive and self-regulation skills. A key question is whether instructional technology can be as effective in fostering metacognitive skills as it is in teaching domain-specific skills and knowledge. On the face of it, the answer is positive. Novel means for interaction, better understanding of learning, mechanisms for tracing students' knowledge, and established domain-level tutoring principles could be applied at the metacognitive level. However, it remains largely unknown exactly how educational technology can help students acquire better metacognitive skills and use them more effectively.

The aim of the workshop is to improve our understanding of the design of goals, instruction, and assessment of tutoring metacognition and self-regulated learning using educational technology.

2. Topics of interest

Topics of interest include, but are not limited to:
- Goals for metacognitive tutoring
- Design guidelines for metacognitive tutors
- Assessment metacognitive knowledge and learning
- Integration of metacognitive and cognitive instruction
- Metacognitive models and representation of metacognitive knowledge
- Pedagogies to teach metacognition
- Empirical or descriptive studies evaluating metacognitive tutoring
- Metacognition and motivation
- Metacognitive awareness of Intelligent Tutoring Systems
# Table of Contents

## Invited Talk
Ken Koedinger, *Toward Understanding When Tutoring Meta-cognition Enhances Domain Learning*  

## Full Papers
Kasia Muldner and Cristina Conati, *Refining Tailored Scaffolding for Meta-Cognitive Skills during Analogical Problem Solving*  
Kwangsu Cho and Moon-Heum Cho, *Self-Monitoring Support in Learning to Write*  
H. Chad Lane, *Metacognition and the Development of Intercultural Competence*  
John Wagster, Jason Tan, Gautam Biswas, and Dan Schwartz, *How Metacognitive Feedback Affects Behavior in Learning and Transfer*  
Amy Witherspoon, Roger Azevedo, and Shanna Baker, *Learners’ Use of Various Types of Representations during Self-Regulated Learning and Externally-Regulated Learning Episodes*
Toward Understanding When Tutoring Meta-cognition Enhances Domain Learning

Kenneth R. KOEDINGER, Vincent ALEVEN, Ryan S.J.d. BAKER and Ido ROLL
Human Computer Interaction Institute, Carnegie Mellon University

We have been pursuing various efforts to extend Cognitive Tutors to support different kinds of meta-cognitive and reflective thinking processes (e.g., Schoenfeld, 1983; White & Frederiksen, 1998). We have had various levels of success in demonstrating enhanced student learning from such meta-cognitive support relative to control conditions without such support. We will summarize work across studies targeting four kinds of meta-cognitive processes:

1. Self-explanation (Aleven & Koedinger, 2002)
2. Error self-correction (Mathan & Koedinger, 2005)
3. Avoiding “gaming” the instruction (Baker et al., 2007)
4. Help-seeking skills (Roll et al., 2007)

These studies illustrate features of “in vivo learning experiments”. All had tight control of the instructional manipulation as it is implemented in computer software within Cognitive Tutors. All involve collection and ‘micro-genetic’ analysis of fine-grained longitudinal data logs. All but #2 were performed in real classrooms. In addition to straightforward post-assessments, measures of robust learning were employed, including transfer (#1-4), long-term retention (#2), and preparation for future learning (#4).

The results are summarized as follows:

1. High school students receiving self-explanation support in a Geometry Cognitive Tutor had improved domain knowledge and transfer relative to control students who were able to do more practice problems without explanation in the same amount of instructional time. Log data provides further evidence that self-explanation support results in less shallow knowledge than problem-solving practice alone.

2. Temporary employment workers receiving tutoring on error self-correction in an Excel Spreadsheet Cognitive Tutor had improved domain knowledge, transfer, and long-term retention relative to controls using the same tutor, but with immediate error feedback. Learning curve analysis of log data shows that the effect was early and probably more relevant to students’ initial declarative knowledge construction than later refinement through practice.

3. Middle school students using a Data Analysis Cognitive Tutor enhanced with a gaming detector and two kinds of responses to detected gaming behavior, emotional and supplementary exercises, had a non-significant improvement on domain knowledge relative to controls using the same tutor without the gaming responses. Treatment students did show reduced overall gaming behavior relative to controls and the number of supplementary exercises received was correlated with learning.

4. High school students using a Geometry Cognitive Tutor enhanced with a help-seeking meta-level tutor did not demonstrate improvement in domain knowledge relative to controls, nor better help-seeking behavior on a transfer environment.
Treatment students became better at deciding when to use help on a declarative help-seeking assessment.

We speculate on what factors may differentiate between larger (#1-2) and smaller (#3-4) impacts on learning. One is whether the goal is primarily a) to support meta-cognitive processes during instruction in order to improve domain learning (#1-2) or b) to improve the meta-cognitive behaviors themselves (#3-4). Another is the extent to which the meta-cognitive process has connections with affect and motivation. It may be that the more connected the meta-cognitive process is to motivational issues (#3-4), the less chance that a scaffolding, monitoring, and tutoring approach will work. A source of suggestive evidence comes from the gaming studies (#3) where surveys of students indicate that a high correlation between negative affect toward mathematics and high levels of gaming.

References

Refining Tailored Scaffolding for Meta-Cognitive Skills during Analogical Problem Solving

Kasia Muldner and Cristina Conati
University of British Columbia, Canada
{kmuldner,conati}@cs.ubc.ca

Abstract. In this paper, we discuss lessons learned from an evaluation of the EA-Coach, an Intelligent Tutoring System that aims to foster meta-cognitive skills during analogical problem solving. The evaluation of the EA-Coach demonstrated that overall, it is effective in its pedagogical approach, but that there are instances where its scaffolding for meta-cognition is too subtle. We discuss implications of this finding in terms of refining the system’s support.

1. Introduction

Research indicates that how students learn from various instructional activities strongly depends on their ability to apply meta-cognitive skills (e.g., [7]). Meta-cognition refers to “one’s knowledge concerning one’s own cognitive processes and products or anything related to them” [10]; more informally, meta-cognition can be referred to as “thinking about thinking”. Meta-cognitive skills are therefore domain-independent abilities that are an important aspect of knowing how to learn in general. Unfortunately, not all students can apply meta-cognitive skills effectively, which hinders learning outcomes [7, 20, 21]. For this reason, we have been working on devising computer-based support to foster meta-cognitive skills in a variety of instructional activities. One of the tutors we have developed, referred to as the Example-Analogy (EA) Coach, provides tailored scaffolding for meta-cognition during analogical problem solving (APS) (i.e., using examples to aid problem solving). This instructional context is motivated by research showing that examples are a natural and effective means of learning: students rely heavily on examples during problem solving when learning a new skill [4, 15], and examples are more effective aids to problem solving than hints [16] or general procedures alone [14]. However, research also shows that as is the case during other instructional activities, learning gains from APS depend heavily on the meta-cognitive skills students bring to bear. Two meta-cognitive skills that are relevant to APS and therefore targeted by the EA-Coach include:

1) min-analogy: solving the problem on one’s own as much as possible instead of by copying from examples [20];

2) Explanation-Based Learning of Correctness (EBLC): learning new domain principles via a form of self-explanation (the process of explaining and clarifying instructional material to oneself [7]). EBLC involves relying on one’s existing commonsense or overly general knowledge to explain how an example solution step is derived [21].

Min-analogy and EBLC are complementary meta-cognitive skills that are beneficial for learning from APS: min-analogy allows students to strengthen their
knowledge through practise and discover knowledge gaps, while EBLC can be used to fill these gaps. Unfortunately, some students prefer more shallow processes that hinder learning, such as copying as much as possible from examples without any proactive reasoning on the underlying domain principles (e.g., [20, 21]). To address this, the EA-Coach includes several levels of scaffolding for meta-cognition, including (1) an innovative example-selection mechanism that chooses examples with the best potential to stimulate in a given student min-analogy and EBLC and (2) interface scaffolding to further encourage the student to use the example effectively. All this scaffolding is fairly subtle. In particular, the EA-Coach does not provide any hints or prompts on meta-cognitive strategies or the target physics domain, leaving much of the responsibility on the student to learn during APS. This design is intended to stimulate students to take initiative in the learning process, rather than enforcing a strict tutorial interaction in which students passively follow a tutor’s directive. The empirical findings from the evaluation of the EA-Coach in a controlled laboratory experiment that we conducted demonstrated its pedagogical effectiveness in terms of encouraging the target meta-cognitive skills [12]. However, the evaluation also showed that some students needed more explicit scaffolding for meta-cognition than what is currently provided by the system. Therefore, we have been working on incorporating this scaffolding, and in this paper, we discuss issues related to its design.

Given the key role that meta-cognitive skills play in the learning process, there has been growing interest in devising computer-based support for meta-cognition. To date, however, the EA-Coach is the only ITS to target meta-cognitive skills during APS. Consequently, it is the only tutor to support min-analogy, which is specific to APS. Although some ITS support the meta-cognitive skill of self-explanation, the EA-Coach is the only ITS that needs to account for how the presence of an example during problem solving will impact students’ self-explanations. The SE-Coach supports self-explanation during pure example studying, without a problem-solving element [8], Normit-SE [11] and the Geometry Tutor [2] support self-explanation during problem solving, without providing students access to examples. The approach for supporting self-explanation adopted by all these tutors is to provide tools to help students derive explanations, and/or prompts to encourage them to self-explain, which as we indicated above the EA-Coach does not do. There has also been work on supporting other meta-cognitive skills, including effective help seeking [17], hypothesis generation during exploration [18] and reduction of gaming [5].

We begin by describing the EA-Coach’s scaffolding for meta-cognition. Next, we describe lessons learned from the empirical evaluation of the EA-Coach, and then discuss their implications in terms of refining the system’s meta-cognitive scaffolding.

2. The EA-Coach

The EA-Coach aims to foster the meta-cognitive skills of min-analogy and EBLC in the domain of introductory Newtonian physics. The system’s key form of support corresponds to its example-selection mechanism. We describe this mechanism after we present the scaffolding for meta-cognition embedded in the EA-Coach interface.

2.1. Scaffolding for Meta-Cognition via the EA-Coach Interface

The EA-Coach Coach interface includes two windows that students use to solve problems and refer to examples (problem and example windows in Fig. 1). The
Workshop on Metacognition and SRL. AIED 2007

problem window’s design is directly based on that in Andes [8], an ITS supporting pure problem solving without access to examples. This window allows students to enter the problem solution by drawing free-body diagrams and by entering equations via the keyboard. The system does not constrain input of the problem solution, in that students are free to enter the solution steps in any order and/or skip steps if they wish. The EA-Coach provides immediate feedback for correctness on students’ problem-solving entries, by coloring correct vs. incorrect entries red or green, respectively. This feedback is the first form of scaffolding for effective APS provided by the EA-Coach.

As evidence in cognitive science demonstrates (e.g., [7]) and as we confirmed through our pilot studies, some students lack self-monitoring skills and so are unable to diagnose their own misconceptions or errors. We argue that immediate problem-solving feedback can help trigger min-analogy and EBLC in these students. For instance, suppose a student with a poor min-analogy tendency is generating the problem solution by indiscriminately copying from an example that includes some differences blocking ‘correct’ copying of its solution. Immediate feedback for correctness can make the student aware of the incorrectly copied steps and so discourage excessive copying by highlighting its limitations. As a second example, consider a student who inferred an incorrect rule via EBLC from an example and applied it to generate the problem solution (students may need several attempts before a correct rule is inferred [6]). Feedback for correctness can make the student aware of the misconception and encourage her to repair it.

While working on a problem, students can ask for an example, which the EA-Coach adaptively selects and presents in the example window (see Fig. 1b). The format of the example shown in the example window evolved from our pilot evaluations, and is intended to mirror the problem window’s format. The example window includes mechanisms to provide further scaffolding for the targeted meta-cognitive skills. One form of this scaffolding corresponds to the masking interface that covers the example specification and solution steps (see Fig. 2; note that the masking interface is not shown in Fig. 1). Moving the mouse over a region in the masking interface uncovers the region and covers whatever region was previously uncovered. The masking interface is

Figure 1: The EA-Coach interface
intended to (1) help focus students’ attention on individual example lines, and (2) discourage copying, because of the effort needed to explicitly uncover the example solution. To further discourage copying, another form of scaffolding corresponds to the lack of Copy and Paste functionality between the example and problem windows. This design is based on findings from an earlier study where we conducted that showed students abused Copy/Paste functionality to indiscriminately copy example solutions.

2.2. Scaffolding for Meta-Cognition via the EA-Coach Example-Selection Mechanism

In addition to the interface scaffolding described in the previous section, another form of scaffolding provided by the EA-Coach for min-analogy and EBLC corresponds to the system’s example-selection mechanism. Specifically, when a student asks for an example during APS, the EA-Coach selects the one from the pool of available examples stored in the system that has the best potential to meet the following two objectives: (1) learning: the example triggers learning by supporting min-analogy and EBLC and (2) problem-solving success: the example helps to solve the problem.

In order to find appropriate examples that enable learning and successful problem solving, a key factor that needs to be taken into account is the similarity between the problem the student is working on (target problem from now on) and the selected example, because it impacts the APS process. For instance, differences between the target problem and the example may hinder both learning and problem-solving success during APS, if students lack the skills to bridge the differences [13]. However, there is also evidence that although highly-similar examples support problem-solving success, they may interfere with learning, possibly because they allow the problem solution to be generated by copying the example solution [15]. Unfortunately, to date there is a lack of full understanding on how different levels of similarity impact students’ APS behaviors and whether some kinds of similarity support both learning and problem-solving success. In our work, we propose: (1) a novel classification of problem/example differences and (2) hypotheses regarding their impact on APS. Since this is described in [9], in this paper we only highlight some of the key features of our classification and associated hypotheses.

We classify two solution steps as structurally identical if they are generated by the same rule and structurally different otherwise. To illustrate this, Fig. 3 shows fragments of the specifications/solutions for the problem/example pair in Fig. 1, including two structurally-identical pairs of steps (see $P_{step_n}/E_{step_n}$ and $P_{step_{n+1}}/E_{step_{n+1}}$, Fig. 3). Now, two structurally-identical steps may be superficially different (e.g., see $P_{step_n}/E_{step_n}$ and $P_{step_{n+1}}/E_{step_{n+1}}$, Fig. 3). We classify differences between structurally identical steps as trivial or non-trivial, based on the type of transfer from problem to example that a given difference allows: (1) trivial differences allow example steps to be copied because they can be resolved by transformational analogy (i.e., using the problem/example specifications as a guide to substitute example constants via ones needed for the problem solution; as is the case for $P_{step_n}/E_{step_n}$, Fig. 3); (2) non-trivial differences cannot be resolved by transformational analogy and so require more in-depth reasoning such as EBLC to be resolved (as is the case for $P_{step_{n+1}}/E_{step_{n+1}}$, Fig. 3). Given that trivial differences can easily be resolved, we argue that (1) non-trivial
differences have better potential than trivial differences to encourage min-analogy and EBLC for students with poor knowledge and meta-cognitive skills, and (2) neither type of difference (trivial, non-trivial) interferes with problem-solving success (since the problem/example solution steps corresponding to the difference are generated by the same rule, meaning that the example affords the student the opportunity to learn the rule needed to derive the problem step).

Our hypotheses regarding the impact of problem/example differences on learning and problem-solving success are embedded into the EA-Coach’s example-selection process. A challenge with our approach, however, is how to find examples that are different enough to discourage copying and trigger min-analogy and EBLC but still provide enough scaffolding to help students learn and solve the problem, given the great variance that exists between students in terms of knowledge and meta-cognitive abilities. To address this challenge, we incorporated relevant factors (student characteristics, problem/example similarity) into a probabilistic student model that corresponds to a dynamic Bayesian network. The student model plays a crucial role in the EA-Coach example-selection process. Specifically, the suitability of a candidate example is quantified via a two-phase decision-theoretic process:

1. **[Simulation phase]** First, the EA-Coach student model is used to predict how a given student will solve the problem and learn from doing so in the presence of the candidate example.

2. **[EU calculation phase]** Second, the model’s prediction is quantified via a utility function that calculates the candidate example’s expected utility for enabling learning and problem-solving success.

This two-phase process is repeated for each candidate example, and the example with the highest expected utility is presented to the student (for details, see [12]). Our approach for example selection has two key advantages. First, it allows us to take into account a student’s knowledge and meta-cognitive skills during the example-selection process and thereby tailor the choice of example to a given students’ needs. Second, it handles the uncertainty that is inherent in our modeling process in a principled manner.

### 3. Evaluation of the EA-Coach

We evaluated the EA-Coach’s pedagogical effectiveness using a controlled laboratory experiment with 16 UBC students. Details on the study methodology and results are provided in [12]. In this paper, we summarize the key findings, and focus on
describing lessons learned from the evaluation pointing to how to refine the EA-Coach scaffolding. Since the example-selection mechanism is the primary form of support for min-analogy and EBLC, the study focused on evaluating how the mechanism realized its learning and problem-solving success objectives. To do so, we had students solve two problems; for each problem, we gave students access to an example. For one of the problems, the example was selected by the EA-Coach (adaptive selection), while for the other (static selection condition), an example most similar to the target problem was provided. In both conditions, the version of the EA-Coach interface presented in section 2.1 was used. All actions in this interface were logged; all sessions were videotaped and the talk-aloud method was used to capture subjects’ reasoning.

The evaluation of the EA-Coach showed that its example-selection mechanism meets the two selection goals (learning, problem-solving success). The results are summarized in Table 1. As far as learning is concerned, we focused on analyzing APS behaviors that foster learning: EBLC and min-analogy (i.e., degree of copying). On average, students generated significantly more EBLC self-explanations and copied significantly less when presented with the EA-Coach’s examples, as compared to statically selected examples. Since cognitive science research shows that copying is detrimental to learning and EBLC fosters learning, these findings provide support regarding the EA-Coach’s pedagogical effectiveness. The evaluation also showed that overall, students were successful in generating the problem solution in the presence of statically and adaptively selected examples. Specifically, all of the students generated the problem solution in the static condition, while in the adaptive condition, all but two did so (two students generated a partial solution). This difference between conditions, however, is not statistically significant (sign test, p=0.5). We also analyzed students’ performance during the problem-solving process: in the adaptive condition, students had a significantly higher task time and significantly more errors while generating the problem solution, as compared to the static condition. However, from a pedagogical standpoint, these are not negative findings because increased task time/errors are by-products of learning, i.e., learning takes time and make require several attempts before correct principles are inferred [6]. Therefore, in general the evaluation validated the EA-Coach’s approach for supporting meta-cognition during APS. However, our evaluation also showed that there were some instances where the EA-Coach’s scaffolding was too subtle to encourage min-analogy or EBLC, suggesting that there is room for improving the system’s support. Here, we provide details on some of these instances, since they guided the refinements to the system we discuss in section 4.

**Min-Analogy.** The EA-Coach’s adaptively-selected examples are intended to encourage min-analogy by discouraging copying. However, two students copied more in the adaptive condition than the static condition. One of these students had an above average number of copy events in both conditions, suggesting that she had a poor min-

---

**Table 1: Summary of results**

<table>
<thead>
<tr>
<th></th>
<th>Mean Adaptive</th>
<th>Mean Static</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copy Episodes</td>
<td>5.9</td>
<td>8.1</td>
<td>7.2</td>
<td>0.023</td>
</tr>
<tr>
<td>EBLC Episodes</td>
<td>2.92</td>
<td>1.14</td>
<td>12.8</td>
<td>0.005</td>
</tr>
<tr>
<td>Task Time</td>
<td>42min, 23sec</td>
<td>25min, 35sec</td>
<td>31.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Errors</td>
<td>22.4</td>
<td>7.6</td>
<td>11.5</td>
<td>0.007</td>
</tr>
</tbody>
</table>
analogy tendency. Thus, it appears that more explicit scaffolding than what is currently provided by the EA-Coach may be needed to encourage a shift over to min-analogy for students with a poor min-analogy tendency. The second student had an average number of copy episodes in the adaptive condition, and about half of the average number of copy episodes in the static condition. One explanation as to why this student copied more in the adaptive condition may be related to the order in which this student experienced the conditions (the conditions were counterbalanced to account for order effects). Specifically, the student solved the problem in the adaptive condition first. Although we did not find overall that condition order had an effect, for this student the adaptively selected example may have discouraged copying, but this did not become apparent until the subsequent (static) condition. We should also point out that although students copied less in the adaptive condition than the static condition, students still did copy in the adaptive condition (see Table 1). Thus, there is room for improvement with respect to the EA-Coach’s scaffolding to encourage students to engage in min-analogy.

**Gaming Behavior.** Although overall, adaptively selected examples effectively discouraged copying, they did not always encourage behaviors beneficial to learning. In particular, when copying resulted in errors, some subjects would switch from copying to rapidly entering alternative guesses to generate the problem step (a behavior that may be referred to as ‘gaming the system’ [5], which can interfere with learning). Since the system provided feedback for correctness, students could resort to gaming instead of deriving the step by learning the corresponding rule (although this was not a practical strategy given the large number of alternatives students had to try before finding the right one). For instance, one student produced forty attempts to generate a problem step, after unsuccessfully trying to copy it from the example (the example and problem shared a non-trivial difference at this point that blocked copying). This student, who generated very few self-explanations, continued to produce subsequent solution attempts quite quickly, leaving little or no time for reasoning of any kind. Gaming behavior was sometimes apparent in subjects’ utterances - for instance, another student said “I’m just trying things, I don’t feel like thinking about it”. This suggests that explicit scaffolding to discourage gaming should be incorporated into the system.

**EBLC.** The EA-Coach’s adaptively-selected examples are intended to encourage EBLC better than statically-selected examples. Although none of the students expressed fewer EBLC explanations in the adaptive condition than in the static condition, five students generated an equal number of explanations in the two conditions. Three of these students expressed a below-average number of explanations, and so appeared to have a low EBLC tendency, suggesting that they required more explicit scaffolding to encourage them to self-explain. The remaining two students generated an average number of EBLC explanations, and so it is not clear why the adaptively selected examples did not encourage them to self-explain better than statically-selected examples. In addition to the cases discussed above, two other students in the adaptive condition were not able to learn all the rules that were required to generate the problem solution, which hindered their problem-solving success, which we discuss below. Thus, as is the case with min-analogy, there is room for improvement with respect to the EA-Coach’s scaffolding for EBLC.

**Problem-Solving Success.** The EA-Coach aims to find examples that help to generate the problem solution. As we mentioned above, two students generated a correct but incomplete solution in the adaptive condition. Both of these students received an example with non-trivial superficial differences that blocked copying of some of its solution steps, because the EA-Coach student model predicted that this would trigger
learning of the corresponding rules via EBLC. However, one of these students showed no desire to engage in any in-depth reasoning during the study and so likely had a very low EBLC tendency, indicating that she needed more support for EBLC than that offered by the EA-Coach. On the other hand, the other student under consideration did generate a number of EBLC self-explanations and invest effort during problem solving to learn some of the rules needed to solve the problem (as we found by comparing her pre and post-test answers on related questions). However, although the student model simulation predicted she would learn all the necessary rules and thus generate the full problem solution, she was unable to do so within the allotted time. We can’t predict whether this student would have eventually generated a full solution or whether she would have become overwhelmed and frustrated by the process. Thus, even if students have good APS tendencies, the EA-Coach’s scaffolding may be too subtle to ensure that they will learn all the rules needed to generate a full problem solution.

4. Refining the EA-Coach’s Meta-Cognitive Scaffolding

As described above, our evaluation showed that sometimes students required more explicit scaffolding for meta-cognition during APS than what is currently provided by the EA-Coach. Therefore, as our next step, we have been working on designing this scaffolding. Given that some students had difficulty generating EBLC explanations during our study, one form of scaffolding we are working on corresponds to tools to help students infer the appropriate domain principles via EBLC. To illustrate how EBLC operates, Fig. 4 shows how a student could use it to explain how Estep3 is derived in the example in Fig 3, which corresponds to inferring the rule that a normal force exists [21]. Specifically, the student relies on her existing (1) common sense knowledge to infer that since the ramp supports the crate, there is a force on the ramp applied by the crate (2) overly general knowledge to infer that this force is an ‘official physics force’; (3) domain knowledge to infer that there is a reaction force exerted by the ramp on the crate. As a consequence of this line of reasoning, the student learns a new domain rule. The challenge with incorporating tools for EBLC is how to support the overly general and commonsense reasoning that characterizes this type of self-explanation. Traditionally, tools for self-explanation support domain-based reasoning, and involve selecting the explanation (or portion of) from a list (e.g., see Fig. 5, which shows a tool scaffolding self-explanation provided by the SE-Coach [8]). Recently, some work has also been done on incorporating free-form explanations into the Geometry Explanation Tutor, where students can enter the explanation by typing [1]. Since with this latter approach students are not constrained in how they generate explanations, they could express EBLC-reasoning to the system. However, as far as we are aware, the Geometry Tutor can not provide feedback on EBLC, since it does not have the capabilities to recognize overly general/commonsense reasoning. Thus, in general, it is an open question how tools for EBLC should be designed to provide the necessary scaffolding, or how to incorporate the complex domain and student models needed to allow the system to capture and provide feedback on EBLC.

Instead of tools for scaffolding EBLC-style explanations, another option could involve providing tools for self-explanations that only support domain-based reasoning, of the type shown in Fig. 5. Note that although this tool does not provide any guidance for overly general/commonsense reasoning involved in deriving the explanation shown in Fig. 4, it could still help the student learn the necessary rule. A possible advantage of tools scaffolding EBLC is that they enable students to gain experience with this type of reasoning, which may help them learn to reason via EBLC in the absence of tools.
see if and how this occurs, it would be interesting to compare how tools scaffolding EBLC vs. domain-based explanations impact learning outcomes. Incorporating tools to scaffold the self-explanation process into the EA-Coach means that its student model will need to be refined. The model currently corresponds to a dynamic Bayesian network that takes into account information on problem/example similarity, student characteristics (knowledge, meta-cognitive skills), and problem-solving and example-viewing actions in the interface to infer whether a student is reasoning via EBLC. Tools provide additional information on how students are reasoning, and so the model will need to be extended to take it into account.

In addition to tools, another form of scaffolding we are working on incorporating into the EA-Coach corresponds to interventions encouraging min-analogy and EBLC and discouraging gaming, which our evaluation suggested could be beneficial. A common approach for realizing meta-cognitive interventions involves generating hints, e.g., to encourage explanation [8], or effective help-seeking [3]. Another approach involves using animated pedagogical agents to express approval or lack of it, based on whether students are gaming the system in ways that interfere with learning [5]. Yet a third possibility involves making the use of tools, such as the ones to scaffold EBLC, mandatory if the system detects that students are not reasoning effectively. Our goal in the design of any intervention is to maintain freedom of interaction during the learning process, and have the system intervene only if it detects shortcomings in a given student’s meta-cognitive abilities that are hindering her learning outcomes.

5. Conclusions and Additional Future Work

The EA-Coach supports APS via scaffolding for the meta-cognitive skills of min-analogy and EBLC. The scaffolding is realized through adaptively selected examples and interface design to encourage students to use the selected examples effectively. Although our evaluation of the EA-Coach demonstrated the system’s pedagogical effectiveness, it also showed that in some cases, students needed more explicit scaffolding than what is currently provided to engage in min-analogy and EBLC. To address this, we have been working on devising this scaffolding. To date, we have concentrated on the design of tools to support EBLC and prompts to encourage it as well as min-analogy during APS. Another avenue we are exploring involves incorporating eye-tracking technology into the framework. Currently, the EA-Coach’s ability to track students’ example usage is due to its masking interface. An eye tracker would provide more accurate information on example usage than the masking interface, without the need to cover the example solution. This would facilitate evaluating if and how the masking interface scaffolds APS (e.g., does it discourage copying as it is...
intended to?). In the longer term, we planning on investigating the EA-Coach’s role in cognitive skill acquisition. The EA-Coach is integrated into an ITS framework that includes two other tutors, which support pure example studying and pure problem solving. Since research suggests that skill acquisition starts with example studying, moves to APS and then to pure problem solving [19], this opens up the possibility for investigating whether there are optimal trajectories that students can follow through these activities. If so, how can they be defined and supported in an intelligent learning environment that includes all three coaches?

References


Self-Monitoring in Learning to Write

Kwangsu CHO and Moon-Heum CHO
University of Missouri, Columbia, USA

Abstract. The purpose of this study was to examine the role of self-monitoring support for writing skill improvement in a reciprocal peer review of writing system called SWoRD [5]. As monitoring is found critical in self-regulated learning research [3], [7], [13], [21], students were provided with opportunities of self-monitoring their writing through self-evaluations and peer evaluations on both their own writing and peer writing. With 601 undergraduate and graduate students from 16 courses in three U.S. universities, it was found that although not all the students did develop successful monitoring skills, the students who developed good monitoring drastically improved their writing compared to those who did not. Finally, we discussed the results and suggested future research.

Keywords. Self-monitoring, Peer review, Writing, CSCL, SWoRD

Introduction

While writing is considered as one of the most important skills that learners are expected to master for professional as well as academic success, writing well is a fundamental skill that most students lack across any ages in the U.S. [15] and also very likely in other countries. A recent study reports that 69% of 8th graders and 77% of 12th graders have only basic or lower levels of writing skills [26]. Not surprisingly, these students are entering colleges ill-prepared to engage in written communication needed for academic success in college.

As part of the Writing-Across-the-Curriculum or Writing-In-the-Discipline movement, most U.S. universities commit considerable financial and pedagogical resources to having first-year composition courses and writing-intensive courses in disciplines. As a vicious cycle, however, lower levels of writing skills are funneling more and more resources away to provide basic level reading and writing skills. More than 50 percent of first-year undergraduates are still incapable of writing without fairly free of basic language errors [24]. Moreover, undergraduates are often found to have serious difficulty writing in their disciplines even after completing composition courses [25], which cast a blight on getting good salaried jobs upon graduation [27].

Accordingly, the U.S. National Commission on Writing consisting of more than 4,300 schools and colleges in the U.S. declares of great urgency the increased emphasis on writing at all levels of education [15]. One of the major impediments in writing instruction is that instructors show the near total neglect of writing due to the instructor’s demanding workload of reading, commenting on, and grading student papers [15].

To address the impediment while improving writing instruction, the Scaffolded Writing and Rewriting in the Discipline (SWoRD) system (http://www.missouri.edu/~chokw/SWoRD.html) [5] has been developed. Based on the social-cognitive framework of writing skill acquisition [17], [20], [28], the SWoRD system implements a familiar but understudied tool, reciprocal peer reviewing of writing: Students in SWoRD play two roles, one of writer and one of reviewer. 

1 Corresponding Author: School of Information Science & Learning Technologies, University of Missouri, Columbia, MO 65211, USA; E-mail: chokw@missouri.edu.
While it is a practical solution, reciprocal peer reviewing is challenging for writing instruction mainly because students are very likely to be novices in their disciplines and also inexperienced in writing and generating constructive feedback. Therefore, SWoRD implements peer reviewing in a new way to overcome the typical challenge with peer reviewing. Specifically, SWoRD provides more pedagogical structure to guide creation of good writing assignments through its case-based reasoning module, more accountability so that students must take writing and reviewing tasks seriously through its writing and reviewing support, and infrastructure using easy-to-use technology to make the tasks manageable. Since the fall of 2002, SWoRD has been quite successfully used in about 60 courses from 10 universities. SWoRD has supported a wide spectrum of writing instruction settings from first year composition to doctoral seminar courses, from argumentation writing to lab reports to grant proposals, and from a class of five students to that of 304 students.

As part of an ongoing SWoRD research project, this study focuses on the role of self-monitoring in writing skills. Research often argues that students should develop good self-regulation skills to become proficient writers [20], [21], [29], [30] beyond basic writing skills and knowledge encompassing word, syntax, grammar, genre, style, rhetoric, and audience. For example, experienced writers successfully set goals, monitor, and reflect their learning process through iterative writing and rewriting. Self-monitoring among various self-regulation skills is known to be critical in learning to write [7], [19]. Self-Monitoring as a skill that makes writers accurately perceive their own writing from readers’ perspectives is known to develop by observing oneself compared to what others perceive [3], [14]. Thus, peers in a reciprocal peer reviewing context can be a great source of developing accurate self-monitoring which in turn may lead to self-regulation skills of writing.

Despite the theoretical importance of self-monitoring in writing [7], there is little research that investigated the role of self-monitoring in writing performance. Therefore, this study examines the role of self-monitoring support in improving writing skills in the SWoRD system. To serve the goal, specifically, we investigate two research questions in SWoRD: 1) How students self-monitoring changes over time and 2) how the self-monitoring changes are related to writing quality improvement.

1. Self-Monitoring

Existing research agrees upon the critical role of monitoring in writing [4], [11], [18]. While expert writers have good monitoring skills, learning writers do not have. For example, before writing, experienced writers monitor various components in texts for better communication with readers. Experienced writers are found to be aware of their writing limitations, necessary processes, and how audience would respond to their writing [20]. By contrast, learning writers have severe difficulties of monitoring their writing process. For example, learning writers are often unsuccessful in detecting problems in texts and also unsuccessful in fixing problems even if learning writers identify the same problems as experienced writers [2]. Learning writers tend to monitor local problems such as words, grammatical errors, unlike experienced writers who monitor global or structural levels [8], [16]. In addition, not surprisingly, learning writers are rarely aware of readers’ perspectives [9]. Therefore, learning writers need monitoring support to complement the cognitive deficit of learning writers. Because self-monitoring skills develop by observing what one perceive compared to what others perceive [3], [14], it is expected that self- and peer-evaluation experiences may help writers strategically function while working on writing.
Although participating in both self- and peer-evaluation would improve self-monitoring skills, a reasonable concern is that students may not develop accurate monitoring skills. When student writers overestimate or underestimate their writing quality, their inaccurate monitoring may hinder the students from setting realistic goals and from using appropriate learning strategies. Indeed, it was found that inaccurate monitoring may undermine its positive role in writing improvement [12]. Hence, in their coregulation model, McCaslin and Hickey emphasize the importance of consistency of self-monitoring results between self- and others [14]. If students over- or under-estimate their learning processes, inaccurate monitoring may cause less skillful self-regulation. Thus, inaccurate self-monitoring might result in inappropriate use of learning strategies, which may have students deviate from an established route to writing improvement.

1.1. Self-Monitoring Support in SWoRD

While self-monitoring must be a valuable source for learning to write [7], [9], [10], [19], the intervention of self-monitoring support seems not straightforward in the reciprocal peer reviewing of writing situation. Based on the theoretical review, first, students in SWoRD were asked to participate in evaluating their own writing (self-evaluation) as well as peer writing (peer evaluation). Therefore, students might compare self- and peer-evaluation in their memories. Then, a newer SWoRD version provides students with monitoring interfaces that allowed students to make explicit comparisons between self- and peer evaluations on both self-writing and peer writing. As shown in Figure 1, SWoRD [5] is equipped with two types of self-monitoring supports through evaluations: One is for their own writing and the other is for others’ writing.

1.1.1. Comparing self-evaluation and peer-evaluations on their own writing

With reciprocal peer reviewing, student reviewers are often asked to review multiple peer drafts (typically 4-5 in SWoRD) unlike traditional classrooms where instructors are the most typical source of evaluation [5], [6]. Therefore, reciprocal peer reviewing allows students to receive diverse evaluations. In addition, students as authors self-evaluate their own writing when they submit the writing. Self-evaluation may help authors develop more explicit awareness of their writing quality or problems on the same scale that their reviewers use.

Therefore, students are allowed to compare their self-evaluations and peer evaluations on their own writing. The discrepancy between the two evaluations may play a role of triggering authors to act upon the gap to improve their writing. However, it should be noted that SWoRD did not provide students to make explicit comparisons between self-evaluation and peer evaluation on their own writing when we collected the data for this study.

1.1.2. Comparing self- and peer-evaluation on peer writing
The other type of the monitoring support was designed to allow each reviewer to compare their own evaluations with other evaluations on peer papers. It was expected that reviewing peer writing may help students better understand writing assignment, evaluation rubrics, what to do, and what not to do in writing [3], [5]. This active review experiences may enable student as reviewers to view writing from various reader perspectives. Figure 2 and Figure 3 show partial interfaces supporting self-monitoring for a reviewer. The interface visualizes the extent to which a reviewer’s evaluation is consistent with that of others who reviewed the same papers. The pattern in Figure 2 shows that the reviewer’s grades are consistent with those of others, while there is a visually significant difference with 400lbs Gorilla. Pseudo names such as 400lbs Gorilla or River are used to keep students from identifying reviewers. Also, if a reviewer clicks on the author name, then the reviewer can read both her own review and others’ reviews on the same writing. Figure 3 is another SM interface that allows reviewers to compare written comments and ratings.

In sum, the goal of this study was to examine the role of the self-monitoring support in improving writing skills. More specifically, we investigated how self-monitoring would change
over time, and also how self-monitoring is related to writing improvement. We examined the questions with a large number of participants in three U.S. universities, unlike the past self-regulation research conducted generally with a small number of elementary or adolescent students [10], [14], [17].

2. Method

2.1. Participants

601 students participated in this study over a 2-year span from three research universities in the U.S. across 16 courses including various writing genres and disciplines (i.e., physics, cognitive psychology, cognitive science, and health psychology) (see Table 1). Three of the courses were graduate courses, 13 were undergraduate courses, and four among the 13 undergraduate courses were for non-majors. The courses used the SWoRD system to support their writing assignments and used the default evaluation rubric consisting of prose flow, argumentation, and insight that will be described later. The students participated in this research as their regular course activities. Typically writing and reviewing assignments together accounted for approximately 40% of the final course grade in each course. However the actual grade proportion could vary depending upon courses.

<table>
<thead>
<tr>
<th>Course ID</th>
<th>Discipline</th>
<th>School Year</th>
<th>University</th>
<th>No. of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>G01</td>
<td>Health Psychology</td>
<td>2005</td>
<td>A</td>
<td>30</td>
</tr>
<tr>
<td>G02</td>
<td>Instructional Technology</td>
<td>2004</td>
<td>C</td>
<td>12</td>
</tr>
<tr>
<td>G03</td>
<td>Biomedical Informatics</td>
<td>2005</td>
<td>B</td>
<td>14</td>
</tr>
<tr>
<td>U01</td>
<td>Cognitive Science</td>
<td>2004</td>
<td>B</td>
<td>20</td>
</tr>
<tr>
<td>U02</td>
<td>Cognitive Science</td>
<td>2005</td>
<td>B</td>
<td>17</td>
</tr>
<tr>
<td>U03</td>
<td>Cognitive Psychology*</td>
<td>2005</td>
<td>B</td>
<td>67</td>
</tr>
<tr>
<td>U04</td>
<td>Cognitive Psychology*</td>
<td>2005</td>
<td>B</td>
<td>78</td>
</tr>
<tr>
<td>U05</td>
<td>Cognitive Psychology*</td>
<td>2004</td>
<td>B</td>
<td>70</td>
</tr>
<tr>
<td>U06</td>
<td>Cognitive Psychology*</td>
<td>2005</td>
<td>B</td>
<td>25</td>
</tr>
<tr>
<td>U07</td>
<td>Physics</td>
<td>2005</td>
<td>B</td>
<td>81</td>
</tr>
<tr>
<td>U08</td>
<td>Physics</td>
<td>2005</td>
<td>B</td>
<td>26</td>
</tr>
<tr>
<td>U09</td>
<td>Writing</td>
<td>2005</td>
<td>B</td>
<td>10</td>
</tr>
<tr>
<td>U10</td>
<td>Writing</td>
<td>2005</td>
<td>B</td>
<td>19</td>
</tr>
<tr>
<td>U11</td>
<td>History</td>
<td>2005</td>
<td>B</td>
<td>97</td>
</tr>
<tr>
<td>U12</td>
<td>Leisure</td>
<td>2005</td>
<td>C</td>
<td>19</td>
</tr>
<tr>
<td>U13</td>
<td>Leisure</td>
<td>2005</td>
<td>C</td>
<td>16</td>
</tr>
</tbody>
</table>

Note. G = graduate; U = undergraduate, * Courses for non-majors.

2.2. Writing Task

The exact writing task assigned to students varied across the courses, as one would expect across courses from many different disciplines. The required length of the assigned papers varied from shorter (5-to-8 pages) to longer papers (10-to-15 pages). Writing genres included 1) a research paper, 2) an application to real life of research findings, 3) technical writing, and 4) a research proposal.

2.3. Self- and Peer Evaluation
Students in this study used same evaluation criteria for their self-evaluation and peer evaluation. Also the evaluation rubrics were always available for students to access. The default evaluation dimensions were prose flow, argumentation, and insight on a 7-point rating scale.

- Prose flow concerned how easily the main points of the argument could be followed, including aspects of sequencing and transitions.
- Argument concerned the quality of the claims and support, including relevance and consideration of counter-arguments.
- Insight concerned the new ideas, information, and inferences that the paper contributed to the class, beyond assigned course texts and materials.

2.4. Self-Monitoring (SM)

SM is defined as an absolute difference between an average of self-evaluation (SE) and that of other evaluation (OE) on their own writing. If the difference is closer to zero, a learner is assumed to have better self-monitoring. Students had SE and OE on the first writing and also on their revised writing.

\[ SM = | SE - OE | \] (1)

2.5. Self-Monitoring Development (SMD)

SMD is defined as a unidirectional change from the first self-monitoring (SMt1) to the second self-monitoring (SMt2). If SMD is positive, SMD was defined as developed because positive values means SMt2 is smaller than SMt1. This means the gap between self- and peer-evaluation is reduced over time. By contrast, if SMD is equal to or less than zero, SMD was defined as not-yet-developed. This means the gap between self- and peer-evaluation is either the same or is not reduced over time.

\[ SA = SMt1 - SMt2 \] (2)

2.6. Procedure

The experiment followed the built-in processes in SWoRD. Specifically, in step one, students create an account in the system and specify a pseudonym. Papers are later distributed to authors under this pseudonym to reduce any status biases that may occur in peer review. Reviewers are only identified to authors by number (e.g., reviewer #1, reviewer #2, etc) to ensure there is no retribution between particular authors and reviewers. At this time, instructors set writing and reviewing assignments by using its case-based reasoning module (see Figure 4), due dates, and assignment policies.

In step two which was optional, students practice the review criteria with three sample papers and receive feedback from SWoRD based on past expert review and peer reviews on the same papers.

In step three, student authors upload their draft paper before the first draft deadline. When submitting their draft, students also self-evaluate their writing quality based on the three dimensions. The self-evaluation opportunity helps students self-monitor their writing. Once the submission deadline has passed, each author’s draft is assigned to n peers, where n is prespecified by the instructor (usually 5 or 6). We use a moving window algorithm so that no two drafts are assigned to the same set of n peers. Also, student reviewers are selected based on the probability of each student completing reviews and that of being a fair reviewer to maximize the chance of writers’ receiving constructive peer reviews.
In step four, peer reviewers submit their evaluations on the papers assigned to them. They generate written comments on and rate each draft on three 7-point evaluation dimensions. SWoRD requires written comments to be entered for each evaluation dimension before the evaluation rating is made. This order encourages reviewers to base ratings on substance rather than intuition. These evaluations and comments are made available to authors after the peer evaluation deadline. The system evaluates each reviewer’s evaluation in terms of three measures; problems in relative ordering of paper quality, systemically high or low evaluation, and systematic problems in how broadly or narrowly evaluations are made. The goal of the grades is to force accountability on the peer-grading task and to encourage reviewers to consider a broader audience than just themselves.

In step five, when students begin revising the draft, they can see the full set of comments on their paper, the system’s assessment of each reviewer’s consistency (marked with stars), their overall writing grade so far in relation to the class mean, the system’s assessment of their own reviewing consistency, and their overall reviewing grade so far relative to the class mean. Students upload their final draft and it is distributed to the same peer reviewers as in the first round of reviewing.

In step six, once the draft has been submitted, each author is asked to rate the helpfulness of each review they received. They use a 7-point helpfulness scale, from Not helpful at all (1) to Very helpful (7). These ratings constitute the other half of the reviewer’s reviewing grade and serve to encourage reviewers to take the written review task seriously.

In step seven, reviewers download the final drafts assigned to them and begin the final draft review process. The same rating rubric is used as for the first draft.

In step eight, students see the full set of comments on their draft paper, the system’s assessment of each reviewer’s consistency, their overall writing grade so far in relation to the class mean, the system’s assessment of their own reviewing consistency, and their overall reviewing grade so far relative to the class mean. Students are asked to grade the helpfulness of the final draft comments. SWoRD automatically places equal weight on first and final draft activities and equally weights reviewing rating consistency and comment helpfulness.

3. Results

The results of the study were presented with each research question.
3.1 How did SM change over time?

We first examined if SM would change over time while students self-evaluated and received feedback from peers. 601 students’ SMt1 was 3.04 (SD = 2.6) and SMt2 was 3.32 (SD = 3.1). Among 601 students, 287 students developed their SM skills over time and 314 students did not develop their SM skills. In general, SM significantly decreased from SMt1 to SMt2, t(600) = -2.23, p < .05. Paired t-tests were conducted to investigate SM change over time in each course of the 16 courses. Among the 16 courses, significant changes appeared only in the three courses (G03: t(29) = 2.67, p < .05; U02: t(16) = -3.47, p < .05; U05: t(69) = 2.06, p < .05).

3.2 How was SM change associated with writing improvement?

To address the second question, the overall Pearson correlation was first computed to analyze a linear association between the amount of SMD and the writing quality improvement. The correlation was statistically significant, r (600) = .66, p < .001, showing that SMD and writing quality improvement are highly associated.

Then the Pearson correlations were carried out in each of the 16 courses. Figure 5 shows individual correlations between SMD and the writing improvement with 95% confidence intervals bars. Two graduate courses’ Pearson correlation out of three was significant and 11 undergraduate courses’ Pearson correlations out of 13 was significant, p < .05. Also, the ranges of correlations were diverse from -.18 to .82. The average Pearson correlation over 16 courses was .56, the average Pearson correlation over the three graduate courses was .58, and the average Pearson correlation over the 13 undergraduate courses was .55.

![Figure 5. Pearson Correlation between SA and Non-SA and Writing Improvement](image)

Note. G = graduate course; U = undergraduate course. * indicates significance of Pearson Correlation p < .05, n indicates the number of students in each course.

4. Discussion

In this study, we examined the role of self-monitoring skills in learning to write in the SWoRD system that provides students with the self-monitoring support. With the large-scale field data, the results empirically found that SM is important for undergraduate students to improve writing. Consistent with self-regulated learning theories, this study empirically showed that skillful self-regulated writers tend to be correctly aware of their own learning [12], [27], [18] which also leads...
to performance improvement. For example, the results are consistent with Graham and Harris’s self-regulated strategy development (SRSD) model in that self-regulated writing positively influences writing performance [10]. In addition, Graham, Harris, and Mason found that students who received SRSD training significantly improved their writing in their posttest as well as outperformed those who did not receive SRSD training in terms of number of writing length, story element, persuasive elements, and quality [11].

The results also revealed that SM skills do not develop easily although about the half of the participants improved their SM skills over time. Indeed, past research showed that students in the computer environment are less likely to activate self-regulated learning strategies than those in the human-agent environment [1]. Although further research is necessary to investigate individual differences of SM skills to see why some students changed their SM skills while others did not, there are at least two possible conjectures. First, the SM interfaces might not be enough to trigger accurate self-monitoring. Although students self-evaluated their own writing, they were provided with any scheme or interface to make an explicit comparison between self- and peer-evaluations on their own writing. Thus, some students might regularly avoid comparing self- with peer-evaluations, or once started, abandon it because of the high memory demands required to effectively administrate the comparison. Therefore, it would be expected that students would develop SM skills with the explicit comparison between self- and peer-evaluations on their own writing. Second, students may better benefit from the SM support when they are aware of the positive role of SM in learning to write. Zimmerman, Bonner, and Kovach suggest that students be explicitly instructed on the benefits of using self-regulated learning strategies before and also while performing tasks [21]. However, we did not provide students with instructions on why they need develop SM skills in the SWoRD system. Therefore, it would be expected that training on the benefit of SM skills would help students develop the skills.

Further research may reveal a more detailed picture of how SM works. For example, the use analysis of the SM interfaces with student characteristics (such as age, gender, ethnicity, gender, past writing experiences) would allow researchers to look into self-monitoring behaviors at a more detailed level. This information would be also important for machine learning to categorize who would and would not develop SM skills.

Finally, feedback quality may influence self-monitoring in learning to write. An interesting hypothesis would be that students receiving quality feedback from peers are more likely to use the feedback, accept suggestions, and self-regulate to bridge the gap between self-evaluation and peer evaluation. By contrast, students receiving low quality feedback from peers are less likely to use the feedback, ignore suggestions, and keep their own self-evaluation. To address this situation, machine learning can be used to diagnose feedback quality. For example, student writing improves as a function of comments that have specific, detailed explanations on a problem of writing and a solution to the problem [23]. Therefore, a discourse tree can be used to detect the absence of explanation in feedback to diagnose feedback quality. The open source SPADE [22] could be used to produce a discourse tree for each sentence in feedback. SPADE is based on the formal model of text coherence developed for a free text and generates a three consisting of three sets: Elementary text units or etus, discourse relations, some of which are indicative of explanation (such as Explanation, Elaboration, and Cause), and derived text units (internal nodes) via discourse relations applied to etus.

Acknowledgement

This research was supported by the grants from the Andrew Mellon Foundation to the first author. We thank Carla Bates for her comments on an earlier draft.
References


[22] ICAS (Interssegmental Committee of the Academic Senates of the California Community Colleges, the California State University, and the University of California Academic Literacy) (2002) A statement of competencies expected of students entering California’s public colleges and universities.


Metacognition and the Development of Intercultural Competence

H. Chad LANE  
Institute for Creative Technologies  
University of Southern California, USA

Abstract. We argue that metacognition is a critical component in the development of intercultural competence by highlighting the importance of supporting a learner’s self-assessment, self-monitoring, predictive, planning, and reflection skills. We also survey several modern immersive cultural learning environments and discuss the role intelligent tutoring and experience management techniques can play to support these metacognitive demands. Techniques for adapting the behaviors of virtual humans to promote cultural learning are discussed, as well as explicit approaches to feedback. We conclude with several suggestions for future research, including the use of existing intercultural development metrics for evaluating learning in immersive environments and to conduct more studies of the use of implicit and explicit feedback to guide learning and establish optimal conditions for acquiring intercultural competence.

Keywords. intercultural competence, metacognition, intelligent tutoring systems, serious games, immersive learning environments, experience manipulation

Introduction

Learning and adapting to a new culture is a significant challenge. In different cultural contexts, interpersonal and communicative behaviors that seem natural may produce unexpected results. For example, simple habits such as nodding and other forms of backchannel feedback can lead to unintended agreements that may, in turn, negatively affect trust, reputation, and so on. It is certainly important for someone who will be spending time in a new cultural context to prepare and be prepared for what awaits them. This is the problem cross-cultural education programs attempt to solve.

A common approach is to provide a learner with a long list of “do’s and don’ts” specific to the country or culture they will be experiencing. While straightforward and easy, this approach relies heavily on rote learning and produces little or no deep understanding of culture. It also ignores empirical evidence that to develop intercultural competence in a general way, people need to move through identifiable stages of development [3,11,13,20]. Rushing to the point of behavior adjustment with limited or no understanding of the underlying cultural reasons can be problematic. True intercultural competence requires (at least) a heightened sense of self-awareness, an ability to self-assess, enhanced perceptive abilities, and a proclivity to reflect on experience. In other words, intercultural development requires metacognitive maturity. This paper is about this process and how immersive learning environments and intelligent tutoring can be used to promote intercultural learning.
1. Metacognition in learning

Metacognition involves active control over cognitive processes during problem solving. For example, when one is solving an algebra equation, cognition refers to the activities necessary to solve it, such as identifying rules to apply, applying them, finding a solution, and so on. Metacognition refers to a higher order of thinking that operates on these cognitive activities, such as planning, analyzing, assessing, monitoring, and reflecting on problem solving decisions and performance. Metacognition also enables more effective learning. A learner who is able to accurately gauge his or her own understanding is better equipped monitor his or her own progress. This typically involves self-questioning and is part of the larger notion of metacognitive regulation [6].

Metacognitive skills can be taught. Numerous classroom studies have shown that explicitly teaching metacognitive strategies in the context of a specific domain (e.g., physics) can improve learning outcomes [5, p.19]. Strategies taught in these studies integrate metacognitive activities with cognitive and seek to make the steps of analyzing, planning, assessing, and reflection habitual in the learner. Studies have also shown that learning is more effective when learners explain worked out solutions to themselves [7]. This phenomenon, which better learners do spontaneously, is known as the self-explanation effect. It can also be taught [8]. More recently, computer tutors focusing on teaching metacognitive skills have shown positive effects on learning behaviors (e.g., [1]).

2. Developing intercultural competence

2.1. The Peace Corps model

Peace Corps volunteers are told to expect four levels of cultural awareness when they begin a new assignment in a foreign country [10, p.199]:

1. **Unconscious incompetence**: minimal awareness of cultural difference or mistakes made; a state of “blissful ignorance”.
2. **Conscious incompetence**: basic realization of cultural difference; minimal understanding of underlying reasons or their significance.
3. **Conscious competence**: increased understanding of differences; deliberate behavioral adjustments are made to reduce cultural errors and misunderstandings.
4. **Unconscious competence**: culturally appropriate behavior is more or less automatic; one’s “instincts have been reconditioned.”

Peace Corps educators have therefore recognized that a certain level of metacognitive maturity is needed to become interculturally competent. For example, to move from unconscious to conscious incompetence, one must realize that what seems “normal” may be considered strange by people from another culture. It must be recognized that any peculiar observed behavior is likely in reaction to the learner. For people with underdeveloped metacognitive abilities, there is a real risk these connections will fail to be made, leading them to conclude the strangeness they perceive is inevitable and beyond their understanding.

2.2. The Developmental Model of Intercultural Sensitivity

The Peace Corps model is useful as a teaching tool, but was not designed as an explanatory scientific model. The Development Model of Intercultural Sensitivity (DMIS), created by Bennett [3], is such a model. It is intended to explain how people construe cultural difference
and how this ability becomes more flexible with time. By construe, Bennett is referring to Kelly’s [16] constructivist view that experience is a function of how one assigns meaning to events that occur in their lives. It is not simply a matter of being present during some event or set of events, but rather how those events are interpreted, encoded into memory, and later remembered. An underlying assumption of the DMIS is that as one’s ability to construe cultural differences evolves, intercultural competence also increases. According to Bennett, “it is the construction of reality as increasingly capable of accommodating cultural difference that constitutes development” [3, p. 24].

What constitutes a cultural difference for someone? It depends on that person’s cultural worldview, which is defined as the set of distinctions the person draws from to construe events in the world. A monocultural person – one who has primarily experienced a single culture in his or her life – will be unable to construe perceived differences from outside that cultural worldview. On the other hand, a person with a broader understanding is generally able to understand, even assume, other cultural worldviews. Hammer and Bennett summarize: “The crux of the development of intercultural sensitivity is attaining the ability to construe (and thus to experience) cultural difference in more complex ways” [11, p.423]. The DMIS is a posits two broad worldview orientations: ethnocentrism and ethnorelativism. Each consists of three stages and are described below. The model is summarized in figure 1.

Bennett defines ethnocentrism as an assumption that “the worldview of one’s own culture is central to all reality” [3, p.30]. Further, an ethnocentric person will implicitly assume that all others share this same worldview. The first ethnocentric stage is denial of difference in which the learner ignores or neglects differences. The next stage is defense against difference which includes recognition of cultural difference, but with negative evaluation. This stage is characterized by an “us vs. them” mindset and overt, negative stereotyping. The last ethnocentric stage is minimization of difference and includes the first signs of considering another cultural worldview. In this stage, the learner emphasizes similarities between cultures and recognizes only superficial cultural differences. Comments such as “we are all the same” are common at this stage. Guidance is especially important because some learners believe minimization is the ultimate stage of growth. When reality sets in that cultural differences are truly significant, there is a risk of withdrawal [3, p. 44].

The remaining three stages represent a shift to the ethnoretative orientation and are characterized by a basic understanding that one’s culture is but one out of many valid worldviews. The first ethnorelative stage is acceptance of difference in which the learner recognizes and appreciates cultural differences. Cultural difference evokes positive feelings in the learner for the first time. The next stage, adaptation to difference, is akin to the Peace Corps “conscious competence” stage in that the learner makes an asserted effort to take the perspective of others. Because of this “frame shifting” ability, the learner can more easily interact with people from other cultures. The final stage is integration of difference: the learner has internalized multiple cultural worldviews and can easily assume different perspectives. Integration is an advanced stage often requiring years of experience to achieve.

**Figure 1.** The Developmental Mode of Intercultural Sensitivity [3].
2.3. Metacognition and the DMIS

Metacognitive skills are critical for advancement through the DMIS stages. Given that the model is based on how one construes cultural differences, it follows that a learner must become aware of the construal process [16], how it works and how to assess their own construal abilities. The following metacognitive skills (at least) are related to the DMIS:

- *Enhanced perceptive abilities* are needed to consciously recognize cultural differences without reacting to them immediately.
- *Self-assessment* failures on one’s own interactions can hinder progress through the DMIS stages. This feeds into *self-monitoring* and tracking through the stages.
- *Cultural self-awareness*, defined as an understanding of one’s own culture, is an important aspect of movement from ethnocentrism to ethnorelativism [2].
- *Self-regulated learning* is part of intercultural developmental: “Generally, people in the later phase of adaptation know how to orchestrate their own learning” [3, p. 59].
- *Planning and goal-setting* can support progression through DMIS stages, such as seeking to reach a specific stage or to understand specific cultural differences.

Bennett addresses metacognition in his discussion of the final stage, *integration of difference*. He explains how one’s identity is not lost in an ethnorelative state:

“Rather, the integrated person understands that his or her identity emerges from the act of defining identity itself. This self-reflective loop shows identity to be one act of constructing reality, similar to others that yield concepts and cultures. By being conscious of this dynamic process, people can function in relationship to cultures while staying outside the constraints of any particular one. (p.60)"

The risk is that some learners may feel like their culture and individuality is lost once they reach this advanced stage. Bennett’s point is that the learner must also accept this redefined and more advanced understanding of self – one that relies on metacognitive maturity. Of course, reaching this is well out of the scope of any educational approach; but, it does make a strong case for nurturing an intercultural learner’s metacognitive skills. The Peace Corps’ approach is consistent with it [10], as are other training programs [14,18]. The rest of this paper explores how immersive learning environments may provide additional support.

3. Intelligent techniques for guided cultural learning

3.1. Immersive cultural learning environments and virtual humans

Technologies such as virtual reality and photoreal 3D graphics are particularly important when considering cross-cultural training. High fidelity simulations make it possible to create realistic portrayals of the products of different cultures, such as architecture, dress, sounds, art, and even smells. This can promote the learner’s sense of immersion and provide a foundation for identifying *objective* cultural differences. Two such environments are shown in figure 2. The first is the Tactical Iraqi Language and Culture Training System (TLCTS) developed by Tactical Language Training, Inc. [15]. TLCTS teaches Arabic language and cultural skills. In the mission environment (shown in the screenshot), a learner is free to explore an Iraqi village, hear the sounds, speak to locals, and make gestures. The clothing, buildings, and surroundings are realistic and thus can give a learner a sense of what it might be like to walk around an Iraqi village. In this way, the system is already in a position to aid in the learner’s identification of cultural differences.
Figure 2. Examples of immersive cultural learning environments: (left) Tactical Iraqi Language and Culture Training System [15] and (right) the Adaptive Thinking and Leadership simulation [21]. Used with permission.

The screenshot on the right in figure 2 is from the Adaptive Thinking and Leadership (ATL) simulation game [21]. ATL is a team-training system that uses human role players for both sides in intercultural scenarios. In-game assessment is performed by peers and instructors who observe play and after-action review (AAR) facilities are available to convey the outcomes to trainees. Learners are often assigned to role play as people from different cultures, with appropriate backstories and goals. Role-playing is a well-developed technique in the crosscultural training literature [18] and consistent with the DMIS with respect to the goal of understanding different cultural worldviews.

In multi-player environments, like the ATL system, inhabitants are human roleplayers. This can be costly and sometimes challenging to control from an educator’s point of view. Research in virtual humans provides an alternative or supplement to cultural team-training in immersive learning environments. Virtual humans combine artificial intelligence (AI) research in cultural and emotional modeling, speech processing, dialogue management, natural language understanding, and gesturing, among others, to enable natural feeling communication and interaction with computer-controlled characters that listen and respond to the user. Virtual humans are driven by rich models of tasks, emotion, body language, and communication [22]. The underlying representations readily support explanation, which can be useful for learning [9]. In the case of intercultural education, it is therefore important to endow virtual humans with models of culture and the ability to explain their actions and reactions in terms of their cultural worldview. It is also important that their behavior be controllable in order to establish conditions that best promote learning.

3.2. Experience manipulation and implicit feedback

Generally speaking, computer simulations simulate real world phenomena as accurately as possible. There are circumstances when it is appropriate to consider goals other than fidelity when deciding how a simulation should behave and what events should occur. For instance, to enhance entertainment value, a popular basketball video game includes special modes that allow players to jump well over ten feet high. In this case, the goal of entertaining the human player trumps the goal of simulating basketball completely realistically. In the case of learning, the same idea applies: if a certain event or situation will promote learning, then the simulation should seek to make that event happen. We refer to this general technique as experience manipulation and now discuss several ways it might be used to promote metacognitive growth and cultural learning.
When a cultural error is committed, or when appropriate actions are taken, learners need support in (at least):

- recognizing that an error was committed (or that a good action was taken)
- finding a causal link between the action taken and the observed reaction
- understanding the reason(s) and culpable underlying cultural differences
- learning how to avoid the same mistake in the future (or sustain good actions)

It is important to go beyond simply concluding to avoid the same behavior in the future since this will contribute little in the learner’s progression through the DMIS stages. Also, the stage a person is in impacts how cultural differences are interpreted. Someone in the denial phase may not even be willing to accept the fact that a cultural error even occurred, for example. Someone in the other two ethnocentric stages (defense and minimization) may be aware an error occurred, but unwilling to take blame or perhaps place the onus on the virtual human to be the one who should adapt. Based this understanding of cultural growth, the reaction of a virtual human to a cultural error should be appropriate for that learner.

Feedback from the simulation itself, such as the oral and gestural reactions of virtual humans, is called *implicit feedback*. To support recognition of cultural errors, there are several strategies that can be used adjust implicit feedback to promote learning. One of the simplest is to accentuate verbal responses of characters to draw more attention to anger or negative feelings, in the case of an error. Similarly, implicit positive feedback can be achieved by accentuating positive and laudatory responses to correct user actions. In some cases, it may even be appropriate for the virtual human to deliver an impassioned mini-lecture regarding the cultural issues in question. The choice of words by the virtual human can be designed to refer directly to actions taken by learner to support the pedagogical goal of linking cause and effect in the learner’s mind. In addition, the virtual human might also drop hints regarding the underlying cultural differences. Body language and gestures can have a dramatic effect on the communicative power of utterances. Figure 3 shows several virtual humans in different emotional states and displaying a variety of gestures. The timing and emphasis of these gestures can be adjusted to meet pedagogical goals in a way similar to the utterance content. Aside from body language, other features that might be adjusted in virtual humans are facial expressions, speech rate, intonation, and tone, emotional state, and personality traits.
3.3. Experience management and interactive narrative

The techniques described in the previous section all focused on emphasizing specific details of interacting with virtual humans to support the learner in recognizing cultural difference and improving their ability to self-assess in an interpersonal context. Of course, explanations for why certain behaviors are culturally offensive can be very complex. They may involve fundamental differences between worldviews, varying ethical standards, social structure, historical and geographical factors, and so on. A deep understanding of culture includes these advanced notions and may enable a learner to go beyond rote learning by providing the knowledge needed to reconstruct appropriate surface behaviors later. Cultural simulations should provide diverse cultural experiences that go beyond one-on-one interactions and carefully manage how events are presented and experienced by the learner.

One such technology focusing on experience management is automated story directing (ASD) [23]. The goal is similar to that of a traditional tutoring system: allow users to feel as much freedom as possible, but keep them on certain paths that consist of certain experiences. The “path” in an interactive storytelling system is a storyline consisting of plots, arcs, events, and other narrative elements. Users may “break” a storyline by taking actions in the virtual world, and so ASD systems use a variety of techniques, like reactive planning, to repair storylines and re-plan when new events are deemed desirable. Often, the aim is to maximize engagement. For cultural training, the additional aim is to create situations and conditions that promote practice and learning.

Metacognition comes into play when we consider the learner’s role in the narrative. She or he must be aware that the actions being taken are being observed by the AI agents in the simulation and that choices being made have observable outcomes. Just as minute details of interaction can be manipulated to highlight differences, so can story elements. For example, if a learner makes a gender-related error early in a game, the ASD may decide to propagate this knowledge through the social network to force the learner to enter future encounters with this baggage. Here the goal is not only to teach gender specific cultural differences, but also to encourage consideration of unintended cultural consequences of earlier actions. This requires the metacognitive abilities to self-assess over an extended period, reaching back further than just the most recent action, and the ability to predict (another metacognitive skill). After dealing with negative consequences of actions, it is hoped that a learner will become more likely to consider possible unintended outcomes of actions before taking them.

3.4. Guidance and feedback

Inherent risks exist in unguided environments, such as inefficient learning, the formation of misconceptions, and development of incomplete or fragmented knowledge [17]. Experience manipulation and implicit feedback can certainly mitigate these risks to a certain degree, but to adequately address the needs of novice and intermediate learners, explicit feedback from a human tutor, pedagogical agent, disembodied coach, or other form of intelligent tutor has the potential to greatly enhance learning (e.g., [4]). Explicit guidance can come in different forms in an immersive learning environment. A pedagogical agent who plays a role in the underlying simulation is a popular approach. The TLCTS [15] and the mission rehearsal exercise described in [22] both provide pedagogical agents in the form of a knowledgeable Sergeant who maintains an understanding of the cognitive goals and can give hints on how to succeed. Another form is that of a disembodied tutor that posts messages in a special area of a GUI or via speech. No matter what the modality, explicit feedback provides more direct and understandable guidance than implicit – this is especially important for novices [17].
Immersive learning environments can be overwhelming at times. With respect to cultural learning, explicit hinting and feedback can help learners in several ways:

1. confirm a learner’s interpretation of observed virtual humans behaviors
2. explain the cultural differences in play during specific interactions
3. explain the “under the hood” reasoning of a virtual human
4. hint about ideal actions to take or warn against certain risks
5. suggest the learner identify possible outcomes and desirable end states

Explicit tutorial feedback removes a level of interpretation for the learner. Rather than guessing or inferring the cognitive and emotional states of virtual humans, a clear statement by a tutor can act as a strong scaffold for learning in immersive cultural environments. There are certainly cognitive aspects to the tactics listed above, but they also address the metacognitive demands of intercultural development. Tactics 1-3 encourage self-assessment by describing the impact of a learner’s actions on a virtual human. Because this is feedback being delivered in a real-time environment, the issues of distraction and cognitive overload need to be considered. Thus, it is ideal to keep “in-game” feedback short and precise, saving the longer explanations for post-practice reflection. This is the approach taken in the ELECT Bi-Lateral negotiation game [12]). Hinting (tactic 4) can be direct (and at the cognitive level), but also can be used to encourage the learner to think about pros and cons of taking different actions – this is especially important in ill-defined domains where assessment is inherently challenging [19]. The content of tactics 1-4 are precisely the things we want the learner speculating on before acting in the environment. In other words, the ultimate aim is for the learner to self-guide in these precise ways. Such cognitive activities by the learner would constitute attempts at self-explanation. Tactic 5, identifying potential end states, is a purely metacognitive tactic that is geared towards supporting goal formation and identifying “what right looks like.” Encouraging the learner to “think before acting”, to engage in planning and simulate hypothetical actions, and “reflect after acting” are at the core of metacognitive maturity and a fundamental requirement for growth through the DMIS stages.

Tutorial sub-dialogues are rarely possible in real-time environments, and so there is time only for very brief periods of reflection. However, once a practice session or exercise is complete, there is time to carefully target metacognitive skill development. Immersive learning environments should therefore provide supporting tools such as video playback. Reflective tutoring is an appropriate supplement to guide the use of these tools and to fill in the gaps from feedback that was delivered during practice. The reflective tutoring system built for the virtual humans [9,12] walks the student through three questions: (1) What happened? (2) Why did these events occur the way they did? and (3) How can good performance be sustained and poor performance be improved? A promising approach here is to leverage explanation facilities of virtual humans to uncover their thinking via explanation [9] and discover what other actions may have produced better outcomes. An advanced tactic is to restart the simulation to give the learner a second chance (a “mulligan”). Reflection at this point may enhance self-assessment skills and intercultural growth.

4. Conclusions and suggestions for future research

This paper has argued that achieving intercultural competence requires strong metacognitive abilities. Although cross-cultural training programs frequently adopt metacognitive approaches to teaching intercultural competence, the connection is rarely made explicit. The Peace Corps approach is an example of this that describes growth as a continuum from unconscious incompetence (not knowing anything and being blissfully unaware of
differences) to unconscious competence (full awareness of differences and appropriate behavior is second nature). By describing these stages, the learner put in a position to self-monitor their advancement. This then requires application of other metacognitive skills, such as self-assessment, self-explanation and self-regulation, to progress through the stages. The Developmental Model of Intercultural Sensitivity (DMIS) is an empirically derived and validated model of intercultural development based on notion of how cultural differences are “construed” by a learner [3,11]. To develop intercultural competence, it is necessary to construe cultural difference in progressively more complex ways, such as from different cultural worldview perspectives. Growth here also requires mature metacognitive abilities and it may be possible to promote these skills in modern immersive learning environments through a combination of experience manipulation and explicit guidance. The DMIS suggests cultural difference as the pivot point for intercultural growth, and so careful direction of virtual humans and delivery of feedback that targets self-assessment, predictive, and reflection skills has the potential to speed the growth to intercultural competence.

Early evaluations of TLCTS [15] and the ELECT BiLAT [12] serious games have shown some promising results with respect to learning aspects of specific cultures. Most of the computer simulations built for cultural education have not undergone rigorous experimental evaluations for learning or for intercultural development. Hammer and Bennett’s Intercultural Development Inventory (IDI) [11] has been used to validate the DMIS and may provide a suitable metric for determining the value-added, if any, that comes from augmenting cultural training programs [18] with immersive learning environments — especially if the IDI can be administered repeatedly and rates of change can be tracked. Other more general questions about feedback are suggested for further study. For example,

- Does implicit pedagogical feedback break immersion?
- If so, what is the cost (if any) to breaking immersion with respect to learning?
- What is the interplay between implicit and explicit feedback?
- What situations merit the use of explicit feedback?

The use of implicit feedback and experience manipulation is perhaps one of the most important open questions to address. The instructor interface to the ATL serious game [21] provides the ability to throw “curve balls” to teams during their missions, such as helicopter fly-overs. These are intended to support the development of adaptive thinking skills under stress. This is related to research in the ITS community on intelligent problem selection and finding the appropriate level of challenge for a learner. These connections need further exploration, as does the reasoning behind expert instructors’ decisions to throw curve balls: What are the triggers? How do instructors decide which curve ball to throw? The answers may not always involve metacognitive skills, but as this paper has attempted to argue, manipulations of this sort in an intercultural context may be ideal to highlight cultural differences to give learners practice in dealing with them. This may enhance the development of intercultural competence, but empirical research is certainly needed to confirm this hypothesis.

Acknowledgements

I thank Dr. Jonathan Kaplan and Dr. Jim Belanich of the U.S. Army Research Institute for many interesting conversations on experience manipulation, implicit feedback, and difficulty that seeded the ideas presented in this paper. The project described here has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM).
Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

References


How Metacognitive Feedback Affects Behavior in Learning and Transfer

John Wagster¹, Jason Tan¹, Gautam Biswas¹, and Dan Schwartz²

¹Department of EECS/ISIS
Vanderbilt University
²School of Education
Stanford University

Abstract. We have developed learning environments that use learning by teaching with metacognitive support to help middle school students learn about complex science topics. We study the role of metacognitive feedback in learning by teaching environments by examining student behaviors and performance across a primary task and a transfer task, noting how the behavior patterns for the high and low performers change as they progress from the primary to the transfer task. In this paper, we examine the apparent behavior shifts that occur as students observe, practice, and then internalize self-regulation skills. Our results show the benefit of metacognitive feedback, and we discuss approaches for incorporating adaptive metacognitive feedback in future systems.

Keywords. learning-by-teaching; metacognitive support, learning behaviors.

Introduction

We have been using learning by teaching models to create learning environments for middle school students that promote the development of higher-order cognitive skills for problem solving in science and math domains [1][2][3]. The resulting teachable agents (TA’s) are software programs where students teach a computer agent using well structured visual representations, and these interactions help shape their thinking [2][4].

Previous studies showed evidence that learning-by-teaching with metacognitive support helped students develop better learning and self-monitoring strategies, and this prepared them for future learning on related topics, even when this learning happened outside of the support provided by the TA environment [2]. In this paper, we perform systematic analyses to draw links between student learning and their observed behaviors during the learning task. The learning measure is defined by the students’ performance on a preparation for future learning (PFL) task. We notice a shift in the behaviors exhibited by both high and low performers from the main to the transfer study. Others, such as Witherspoon et al. [5] have found that learners initially try a variety of strategies when learning a new domain, and as they gain a better perspective they begin to adopt more sophisticated planning strategies for learning in that domain. More generally, we believe a similar behavior shift from observation and cognitive acquisition to self-control and self-regulation occurs as students gain a deeper understanding of the reasoning mechanisms and the metacognitive feedback they receive in our TA environment [6]. The rest of the paper provides an overview of Betty’s Brain, our learning by teaching system, the metacognitive support provided by the system, a description of our experimental study, and a summary of our findings and future work.
1. Learning by Teaching: Betty’s Brain

The teaching process in Betty’s Brain, illustrated in Figure 1, is implemented as three primary activities: (i) *teach*: students explicitly teach Betty using a concept map representation, (ii) *query*: students use a template to generate questions to see if Betty can answer questions based on what she has been taught, and (iii) *quiz*: students observe Betty’s performance on a set of predefined questions. Betty uses qualitative reasoning methods to reason through chains of links [2], [7] to answer questions, and, if asked, explains her reasoning using text and animation schemes. Betty also provides feedback to get the students to adopt more metacognitive strategies in their learning tasks [8]. Students reflect on Betty’s answers and her explanations, and revise their own knowledge as they make changes to the concept maps to teach Betty better. Details of the Betty’s Brain system and experiments that we have conducted with this system are summarized in [2], [8]. Next we discuss the metacognitive support provided to students as they learn about river ecosystems.

![Figure 1: Betty's Brain System with Query Window](image)

2. Metacognitive Support in Betty’s Brain

Cognitive science researchers have established that metacognition and self-regulation are important components in developing effective learners in the classroom and beyond [9], [10], [11]. Pintrich differentiates between (i) *metacognitive knowledge* that includes knowledge of general strategies and when they apply, as well as knowledge of one’s own abilities, and (ii) *metacognitive control* and *self regulatory processes* that learners use to monitor and regulate their cognition and learning [12]. We believe the TA environments when combined with adequate scaffolding and feedback can provide appropriate opportunities for students to develop both metacognitive knowledge and control, and thereby, improve their subsequent learning.

We adopt a self-regulated learning (SRL) framework that describes a set of comprehensive skills that start with setting goals for learning new materials and applying them to problem solving tasks, deliberating about strategies to enable this learning, monitoring one’s learning progress, and then revising one’s knowledge, beliefs, and strategies as new materials and strategies are learnt. In conjunction with these higher level cognitive activities, social interac-
tions and motivation also play an important role in the self-regulation process [11]. We believe that two interacting factors of our TA implementations are particularly supportive of self regulation. The first is the visual shared representation that the students use to teach their agents. The second factor, shared responsibility, targets the positive effects of social interactions to learning. This manifests in the form of a joint effort where the student has the responsibility for teaching the TA (the TA knows no more and no less than what the student teaches it), whereas the TA has the responsibility for answering questions and taking tests.

Betty’s persona in the SRL version incorporates metacognitive knowledge that she conveys to the students at appropriate times to help the student develop and apply monitoring and self-regulation strategies [8]. For example, when the student is building the concept map, Betty occasionally responds by demonstrating reasoning through chains of events. She may query the user, and sometimes remark (right or wrong) that the answer she is deriving does not seem to make sense. The idea of these spontaneous prompts is to get the student to reflect on what they are teaching, and perhaps, like a good teacher check on their tutee’s learning progress. These interactions are directed to help Betty’s student-teacher understand the importance of monitoring and being aware of one’s own abilities. On other cues, the Mentor (and sometimes Betty herself) provides suggestions on cognitive strategies the students may employ to improve their own learning and understanding of the subject matter under consideration.

3. Experimental Design

To study the effect of metacognitive and self-regulation strategies on learning behaviors, we designed three versions of the TA system. We refer to the system used in the control condition as the intelligent tutoring system (ITS) because this directed learning environment contains some aspects of traditional ITS’s [13]. In this condition, the students were taught instead of teaching someone else. Mr. Davis, the Mentor agent, asked the students to construct a concept map to answer three sets of quiz questions. When students submitted their maps for a quiz, Mr. Davis provided corrective feedback that was based on errors in the quiz answers [2]. System 2 was a Learning by Teaching (LBT) environment, where students were asked to teach Betty by creating a concept map. The students were told that Betty needed help to pass a test so she could join the high school science club. Students using the LBT system could query Betty to see how well she was learning, and they could ask Betty to take quizzes at any time during the teaching process. After Betty took a quiz, Mr. Davis graded the quiz, and provided Betty and her student-teacher with corrective feedback. The text of the feedback was identical to what was provided in the ITS system. System 3 was a learning-by-teaching system with Self Regulated Learning (SRL). Students in this condition also taught Betty but the primary differences between the LBT and SRL systems were in Betty’s behavior and interactions with the student, as well as the feedback that the Mentor provided after Betty took a quiz. Betty’s persona in the SRL version incorporated metacognitive knowledge, which she communicated to the students to help them develop and apply monitoring and self regulation strategies to aid their own learning [8]. The PFL study used a version of the system similar to the LBT version, in that it did not incorporate metacognitive feedback.

4. Experimental Study and Results

The study was conducted in two 5th grade science classrooms in a Metro Nashville school. 53 students from the two classrooms were divided into three equal groups using a stratified sam-
pling method based on standard achievement scores in mathematics and language. The three groups, ITS, LBT, and SRL, worked for seven 45 minute sessions over a period of two weeks to create their concept maps on aquatic ecosystems. A PFL study was conducted approximately 8 weeks after the main study where students focused on creating concept maps for the land-based nitrogen cycle. Students were administered pre and post-tests before and after the main study.

4.1. Analysis of Students’ Behaviors

Student activity sequences in each session of the main study were extracted from the system log files. The sequences contained six primary activities: (i) Edit Map (EM), (ii) Ask Query (AQ), (iii) Request Quiz (RQ), (iv) Resource Access (RA), (v) Request Explanation (RE), and (vi) Continue Explanation (CE). Actions where the students were adding, modifying, or deleting concepts and links in their concept map were classified as EM activities. The RQ and RA activity labels are self explanatory. Students in the LBT and SRL groups could ask Betty queries (AQ), and then check Betty’s reasoning by asking for explanations (RE). Betty’s explanations often involved multiple steps that mirrored the multiple steps used by the reasoning process to generate an answer. Betty provided an initial response to a request for an explanation (RE), and then followed it up with more details if the student clicked on the “Continue Explanation” (CE) button. The ITS group also had access to the query and explanation features for debugging their concept maps. Explanations were provided by the Mentor agent. An example activity sequence for a student working on the LBT system in one of the seven sessions appears below:


In previous work [8][14] we used intuition and empirical observations to link behavior sequences to manifestations of metacognitive control and self regulation [11][12]. A primary finding in the earlier studies was that students who frequently exhibited the Quiz-Edit-Quiz behavior (defined as RQ_EM_RQ or EM_RQ_EM) were more likely to have concept maps with low scores. The pattern appeared to reflect trial and error (edit map, see if it worked using the quiz, then repeat to fix problems). On the other hand, students who asked queries to check on the changes they had made to their concept map (EM_AQ) and requested explanations after asking queries (EM_AQ_RE) were more likely to produce high scoring concept maps. Preliminary analysis showed that students in the SRL condition used the EM_AQ and EM_AQ_RE patterns more frequently than the other groups, and the ITS group used the EM_RQ_EM pattern more often than the LBT and SRL groups. We concluded that the metacognitive support helped the SRL students learn good monitoring behavior. Furthermore, the SRL group also produced better concept maps than the ITS and LBT groups.

In this paper we compare behavior sequences that correspond to high and low learning performance in the main and PFL studies. Our findings, like [5], show a definite shift in self regulatory behavior as students gain a better understanding of the domain and learn self regulation strategies by observation and practice. We speculate whether this implies a progression from cognitive acquisition to full self-regulation as outlined in [6].

4.2. Identifying Behavior Patterns Indicative of High and Low Performance

We use correlational analysis to identify activity patterns in the main and PFL study that are indicative of high and low performance. Correlations are computed between the frequency of pattern use by the students and their learning performance. A normalization factor (i.e., the
number of occurrences of the first activity in the sequence) is applied to the frequency computation. We define students’ learning performance by the quality of their concept map at the end of the transfer (PFL) study. Concept map quality is computed as the sum of the correct concepts and correct links in the student’s concept map. Concepts and links are defined to be correct if they appear in the expert map1 or if they are graded to be relevant by two coders because they demonstrated a correct understanding of the domain (even if they were not necessary to answer the quiz questions).

Correlational analysis provides a preliminary method for linking behavior patterns to levels of learning performance. In future work we will develop methods that more definitely establish causal relation between observed behaviors and student learning.

We first identify behaviors in the main study that are indicative of high and low PFL performance. For the correlation computations, we restricted the number of considered activity patterns in the main study to lengths of two and three.2 The mean correlation value for these patterns with the transfer map was 0.087 (SD = 0.146). The activities with large positive correlations were associated with high performance, and the activities with large negative correlations were associated with low performance. A cutoff criterion of $M \pm 2.5D$ was used to select the highest and lowest performance patterns. Table 1 lists the activity patterns with correlation values above the high cutoff of 0.379 and below the low cut off of -0.205.

Table 1. Activity patterns with high and low correlation values with transfer study concept map score

<table>
<thead>
<tr>
<th>High correlation</th>
<th>Low correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity Pattern</td>
<td>Correlation Value</td>
</tr>
<tr>
<td>AQ_RA_EM</td>
<td>0.460</td>
</tr>
<tr>
<td>EM_AQ_RA</td>
<td>0.419</td>
</tr>
<tr>
<td>AQ_RA</td>
<td>0.414</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The three activity patterns that correlated well with high performance included two activities: (i) RA, resource access, for seeking more information about the domain, and (ii) AQ, asking queries to check on answers generated by the concept map. These students used the AQ_RA_EM and EM_AQ_RA activity patterns to check the correctness of their concept maps by asking queries and then looking up the resources to see if the answers were correct. AQ_RA_EM would imply that the students then went on to make changes in their concept maps, and EM_AQ_RA would imply that students used resources to check on the changes they had just made to their concept maps.). We should clarify that the online resources were organized like a textbook with hyperlink structures and keyword search features. Students had to read relevant portions of the text and infer the relations between entities in their construction of the concept map.

Three of the four patterns that showed strong correlations with low performance, i.e., EM_RQ_EM, RQ_EM_RQ, and RQ_EM were linked to the suboptimal Quiz-Edit-Quiz strategy that we have discussed before [2][8]. AQ in the fourth pattern AQ_EM may be a good activity, however, the fact that students went on to edit their concept maps instead of performing

---

1 The expert map was used by the mentor agent, Mr. Davis, to grade the students’ concept maps and provide feedback. However, the students did not have access to the expert map.

2 A maximum length of 3 was chosen to reduce computation time. In future work, we will look at longer behavior patterns.
monitoring activities, such as RA (resource access) or RE (request explanation), led us to believe that these students were not using the AQ feature in a very useful way.

In previous studies, we had conjectured and demonstrated qualitatively that significant use of activity patterns that included the query and explanation mechanisms (AQ, RE, CE) was indicative of high performance. The pattern AQ_RE is the 4th highest ranked activity pattern (correlation value = 0.35) was a little below the high cutoff level. The high rank for the AQ_RE activity pattern is encouraging, but from this analysis one may conclude that the students who perform well in the PFL study use a balanced strategy of initiating their monitoring processes by asking queries and then following them up by asking for explanations (to check on the reasoning mechanisms) or reading the resources further (to check on the correctness of the answer).

4.3. Behavior Patterns by Group

We were also interested in knowing if the different Betty’s Brain conditions influenced the students’ behavior patterns. Like before, we hypothesized that the metacognitive support for the SRL group in the main study would result in these students using activity patterns linked to high performance more frequently than the ITS and LBT groups. On the other hand, the ITS group would show more frequent use of the low performance activity patterns (see [2], [8]). We used an ANOVA to check for significant differences behaviors between the groups (see Table 2). The ANOVA was followed by post-hoc analysis using Tukey’s HSD to establish pairwise differences between groups [15]. This table is not included for the sake of brevity. Pairwise differences at the p < 0.05 level are marked in bold, and those significant at the p < 0.1 level are marked in italics.

<table>
<thead>
<tr>
<th>Behavior</th>
<th>F(2, 51)</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQ_RA_EM</td>
<td>2.554</td>
<td>0.088</td>
</tr>
<tr>
<td>EM_AQ_RA</td>
<td>16.925</td>
<td>&lt; 0.001*&lt;sup&gt;a,b&lt;/sup&gt;</td>
</tr>
<tr>
<td>AQ_RA</td>
<td>3.490</td>
<td>0.038</td>
</tr>
<tr>
<td>AQ_EM</td>
<td>1.829</td>
<td>0.171</td>
</tr>
<tr>
<td>EM_RQ_EM</td>
<td>8.345</td>
<td>0.001&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>RQ_EM_RQ</td>
<td>8.656</td>
<td>0.001&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>RQ_EM</td>
<td>7.111</td>
<td>0.002&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

* SRL students performed behavior significantly more than ITS students (p<0.05)
* LBT students performed behavior significantly more than ITS students (p<0.05)
* SRL students performed behavior significantly less than ITS students (p<0.05)

The results show significant differences between the SRL and ITS groups for three of the behaviors (one high performing behavior: EM_AQ_RA, and two low performing: RQ_EM_RQ and EM_RQ_EM). The only significant difference between ITS and LBT is the EM_AQ_RA pattern. If one relaxes the significance level to p < 0.1, five patterns (bold + italicized) show significant differences between the SRL and ITS groups, five of the behavior patterns are different between the ITS and LBT groups, and there is one behavior difference between the SRL and LBT groups (EM_AQ_RA). This analysis seems to support the fact that the SRL group with metacognitive support used more high performing behavior patterns to support learning than the other two groups and the ITS group used more of the low perform-
ing behavior patterns than the other two groups. The LBT group was in between. Table 5 shows the main study concept map scores for each group. It is clear that the SRL students produced better concept maps (correct concepts + links) than the ITS and LBT groups. The differences in concept map scores are statistically significant.

4.4. PFL Behaviors and PFL performance

Next, we examined activity patterns in the PFL task linked to high and low performance. The mean correlation value for these patterns with the transfer study map was 0.061 (SD = 0.199). Using M ± 2SD as the cutoff, we defined the behaviors indicative of high and low performance.

A comparison of Tables 1 and 3 shows a definite shift in the high performing activity patterns. In the main study, there was significant use of RA and AQ as well as AQ_RE. On the other hand, the top four patterns in Table 3, revolve around the use of EM and AQ. At first glance, one may wonder if this implies a shift to the suboptimal Quiz-Edit-Quiz strategy by the high performing group. On further reflection we realized that the high performing students had gained a good understanding of the reasoning process through self-monitoring by asking queries and explanations in the main study, and they could now directly apply this to the new concept maps without using the scaffolds provided in the system (AQ, RE, and CE). It is also reassuring to see that RA_EM_RQ remained a significant activity pattern, which implies the students were still demonstrating information seeking strategies in the new domain by accessing the resources before editing their maps. EM_RQ_RE is an unusual pattern, but we ignored it because its frequency of use was very low compared to the other patterns in the table. In the transfer study, there were few intermediate questions, so students had to do a lot more reasoning on their own to generate the right maps.

<table>
<thead>
<tr>
<th>High correlation</th>
<th>Low correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity Pattern</td>
<td>Correlation Value</td>
</tr>
<tr>
<td>RQ_EM_RQ</td>
<td>0.584</td>
</tr>
<tr>
<td>RQ</td>
<td>0.575</td>
</tr>
<tr>
<td>EM_RQ</td>
<td>0.547</td>
</tr>
<tr>
<td>EM_RQ_EM</td>
<td>0.530</td>
</tr>
<tr>
<td>EM_RQ_RE</td>
<td>0.481</td>
</tr>
<tr>
<td>RA_EM_RQ</td>
<td>0.480</td>
</tr>
</tbody>
</table>

We believe that the observed shift demonstrates a progression from relying on self-other monitoring in the LBT and SRL groups to self-reliance and self-monitoring (cf. to [6]). Preliminary analysis of students’ query usage by session shows that the shift in behavior occurs as a continuum over time in the main study and not as two discrete points from the main to PFL study (Fig. 2). This shift is particularly noticeable for the SRL condition. A more detailed causal analysis across time is required to establish the nature of the shift, but the correlational analysis provides a plausible reason for the decrease in query use from the main to the PFL study.
There is also an observed shift in the activity patterns used in the main and transfer studies for the low performers. These students had difficulties understanding the reasoning mechanisms in the main study, so they resorted to suboptimal strategies to build their concept maps. In the transfer study, AQ was used frequently in the low performing activity patterns, implying an attempt to gain a better understanding of the reasoning mechanisms using monitoring processes. However, unlike the high performing behaviors in the main study, the transfer study activity patterns do not combine AQ with information seeking (RA) or monitoring (RE and CE) activities. The high performing activity patterns from the main study EM_AQ_RA and AQ_RA_EM still show positive correlations of 0.252 and 0.188, respectively, in the transfer study, which implies a stronger association with the high performing behaviors. This implies that at best, this positive shift for the low performers was still in the early phases of learning and practicing metacognitive strategies, as opposed to demonstrating internalization (understanding) and self-reliant behavior like the high performers.

Like before, we were interested in seeing if the three different treatments in the main study\(^3\) influenced the transfer study activity patterns. Table 4 lists the results of the ANOVA followed by Tukey’s HSD for post hoc analysis of significant pair-wise differences. Only one activity pattern, RQ, showed post hoc differences between pairs of groups at the p < 0.05 level. As we have discussed earlier, the high performing students had a good understanding of the reasoning mechanisms, and they did not need to use the AQ and RE features to monitor and debug their concept maps. It is quite likely that a number of these students were from the SRL group (based on the results in Table 2 and 5). Therefore, their observed behaviors mainly involve reading resources, editing the map, and taking the quiz to see if their generated map is correct. Though their monitoring behaviors are not directly observed, their overall strategies were successful in producing higher quality maps. Again, detailed causal and temporal (i.e., student progression over time) analysis of this data has to be performed to gain a better understanding of the effect of the different conditions on learning performance.

<table>
<thead>
<tr>
<th>Behavior</th>
<th>F(2, 51)</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ_EM_RQ</td>
<td>0.680</td>
<td>.512</td>
</tr>
<tr>
<td>RQ</td>
<td>4.149</td>
<td>.022(^a,b)</td>
</tr>
<tr>
<td>EM_RQ</td>
<td>2.613</td>
<td>.084</td>
</tr>
</tbody>
</table>

\(^3\) All three groups used the same barebones LBT system in the transfer study. Students were told if their quiz answers were right or wrong. They did not receive directed or metacognitive feedback.
We reiterate that overall the SRL group showed significantly better performance in the main and transfer study concept map generation. The LBT performance was in between the SRL and ITS performance. Table 5 summarizes these as concept map scores at the end of the main and transfer studies. In the main study, the better scores may be attributed to the explicit metacognitive feedback provided to the SRL students. For the transfer study, this seems to reinforce our hypothesis that students in the SRL group (and possibly the LBT group) internalized the reasoning mechanism and monitoring strategies, and, therefore, produced better concept maps even though our log files did not capture these behaviors.

<table>
<thead>
<tr>
<th>Group</th>
<th>Main Concept Map Score mean (sd)</th>
<th>PFL Concept Map Score mean (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITS</td>
<td>22.83 (5.3)</td>
<td>22.65 (13.7)</td>
</tr>
<tr>
<td>SRL</td>
<td>31.58 (6.6)&lt;sup&gt;a&lt;/sup&gt;&lt;sup&gt;b&lt;/sup&gt;</td>
<td>32.56 (9.9)&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>LBT</td>
<td>25.65 (6.5)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>31.81 (12.0)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

- **a** – SRL performed behavior significantly more than ITS (p<0.05)
- **b** – LBT performed behavior significantly more than ITS (p<0.05)
- **c** – significantly greater than ITS (p<0.10)

5. Conclusions

The results of this study provide evidence that metacognitive support promotes more effective learning of domain content. For the high performers, a clear shift in behaviors was observed from the main to the PFL study. These students used a balanced strategy that combined information seeking and self-monitoring in the main study. They demonstrated better understanding and self-reliance; therefore, there was less use of the scaffolds provided for monitoring in the PFL study. The information seeking behavior to learn new domain content was a dominant activity pattern across both studies. A shift was also observed for low-performing students, from the classic suboptimal Quiz-Edit-Quiz strategy to more use of the query mechanism (monitoring), but the strategies they used did not progress enough to where they combined monitoring with effective information seeking behavior (RA) to learn the new domain, or to use the explanation mechanism (RE) to monitor their own performance when building their concept maps. More of the low performers came from the ITS and LBT groups, which did not receive metacognitive feedback during the learning phase.
We believe a more in-depth analysis of both student behaviors and additional performance metrics or assessments will more clearly reveal the underlying differences and the cause for these differences. Also, examining how these behaviors form and evolve over time may lead to a better understanding of the differences between groups and learners.

Lastly, this work has important implications toward the development of more intelligent learning environments. If the goal of these environments is to make students better prepared for future learning (PFL) on their own, it is important to design these environments so that they facilitate the shift from observations to emulation to internalization and self-reliance. One way to achieve this is to build in adaptive metacognition (e.g., [16]) into these learning environments. We hope more detailed causal analysis of the performance and behavior data will provide us with the information needed to start designing these environments.

Acknowledgments

This work has been supported by a Dept. of ED IES grant #R305H060089 and NSF REESE Award #0633856.

References

Learners’ Use of Various Types of Representations during Self-Regulated Learning and Externally-Regulated Learning Episodes

Amy WITHERSPOON, Roger AZEVEDO, and Shanna BAKER
University of Memphis, Department of Psychology, Institute for Intelligent Systems, Cognition and Technology Lab
Memphis, TN, USA, 38152 {awthrspn@mail.psyc.memphis.edu}

Abstract. This study empirically examines students’ use of various representations (text-only, text and diagram, videos, and externally constructed representations) while learning about a complex science topic using hypermedia. Eighty-two undergraduate students were randomly assigned to one of two tutoring conditions: self-regulated learning (SRL) or externally-regulated learning (ERL). Participants in the self-regulated learning condition used a commercially-based hypermedia environment to learn about the circulatory system on their own, while participants in the externally-regulated learning condition also used the hypermedia environment, but were given prompts and feedback from a human tutor during the session to facilitate their self-regulatory behavior. Each participant’s interactions with the hypermedia environment were video-recorded and were used to investigate any differences between the SRL and ERL groups’ use of types of representations. Results from process data (think-aloud) indicate that students in the ERL condition spent significantly more time producing externally constructed representations, i.e. their own drawings or notes on the material, than students in the SRL condition. Students in the SRL condition spent significantly more time reading text-only than students in the ERL condition. Correlational analyses indicate that students who spent more time reading text-only scored lower on posttest measures of learning and those who spent more time constructing external representations scored higher on measures of learning. Implications for the design of a computer-based learning environment intended to foster students’ effective integration of multiple representations during learning with hypermedia are discussed.

Keywords: Self-regulated learning, hypermedia, human tutoring, multiple representations, computer-based learning environments

Introduction

Students often learn about complex science topics using multimedia and hypermedia, which involve various sources of information, including diagrammatic representations, textual representations, formulaic representations, and animations or videos. These learners usually have the opportunity not only to view these various forms of representations, but also to construct their own pictorial or textual interpretations of the information conveyed in learning environments through externally constructed representations (i.e. drawing and taking notes). The role these different representations play in the learning process has been theorized and studied by various scientists in past years [1,2,3,4,5]. In addition, models on the way in which these multiple external representations are integrated by the learner to form internal representations have been developed, including Mayer’s [6] Cognitive Theory of Multimedia Learning (CTML).
and Schnitz’s [7] Integrated Model of Text and Picture Comprehension (ITPC), which have both been informed by Baddeley’s working memory theory [8], Paivio’s dual coding theory [9], and Chandler and Sweller’s [10] cognitive load theory.

In order for learners to achieve the potential learning outcomes while learning using hypermedia and multimedia, they must translate between various sources of information, including diagrams, text, and videos with narration. Although it has been assumed that multiple representations of information will always provide students with greater opportunity to realize this potential, previous research has shown, that, in fact, students do not always perform better when using text and diagrams [11,12,13]. In an attempt to explain why certain properties of multiple representations lead to greater learning outcomes, Mayer [6] and Schnitz [7] developed models of the way students integrate information from multiple representations and sensory channels into internal representations and mental models.

1. Theoretical Models of Learning with Multimedia

Both Mayer [6] and Schnitz [7] developed their models with three assumptions. The first assumption, based on Paivio’s [9] dual coding theory, is that humans have visual channels and auditory channels, for processing visual and auditory information, respectively. The next assumption is that each channel of information has a limited capacity for processing, based on Baddeley’s [8] working memory theory and Chandler & Sweller’s [10] cognitive load theory. Finally, both models assume that humans, in actively attending to important information and organizing the selected information into internal representations, are active learners.

Mayer’s [6] Cognitive Theory of Multimedia Learning (CTML) encompasses much more than simply the integration of multiple representations. However, we will only discuss this aspect of the theory. In Mayer’s CTML, incoming information from a multimedia presentation first enters sensory memory according to its modality. For example, words can enter sensory memory either through the eyes (visual modality) or the ears (auditory modality), depending of the presentation mode. Pictures necessarily enter sensory memory through the eyes. Next, words and images from sensory memory that are deemed important are selected to move forward to working memory. Working memory operates at two distinct levels: 1) raw information entering WM from the senses, and 2) constructed knowledge in WM. The raw information in working memory is comprised of words selected from auditory sensory memory and outputted as a word sound base in verbal working memory and images selected from visual sensory memory and outputted as a visual image base in visual working memory. In order for these words and images to enter working memory, a learner must first attend to and select relevant images and words to proceed. After selected words and images enter working memory, the working memory system organizes the selected words into a verbal model and the selected images into a pictorial model. Finally, the verbal and pictorial internal representations are integrated with one another and with prior knowledge.

Schnitz’s [7] Integrated Model of Text and Picture Comprehension (ITPC) is similar to Mayer’s CTML, with interesting differences. First, in addition to text entering the auditory sensory register, in the ITPC, images, called sound images, also register in the auditory sensory memory. Next, both the visual and auditory sensory registries forward both words and images to visual and auditory working memory, respectively. In other words, the visual sensory registry can select words and images to
proceed to visual working memory through the visual channel and the auditory sensory
registry can select words and images to proceed to auditory working memory. Finally,
verbal information in both visual working memory and auditory working memory
proceeds to the propositional representation level in working memory and pictorial
information in both visual and auditory working memory proceeds to the mental model
level in working memory. Propositional representations and mental models in the
ITPC are equivalent to the CTML’s verbal model and pictorial model, respectively.
Also, Schnotz refers to information from long term memory, which is integrated with
the information from working memory, as cognitive schemata, rather than prior
knowledge. The process of the eventual integration of the verbal and pictorial model
into long-term memory is one which has received little elaboration in the theoretical
models.

Ainsworth [1] argues that multiple representations play three major roles in
learning. First, they play complementary roles, in supporting complementary processes
and providing complementary information to the learner. Second, they constrain
possible interpretations on the part of the learner by familiarity or by inherent
properties. Finally, multiple representations aid learners in constructing a deeper
understanding of material by supporting abstraction, by promoting generalization to
novel situations, and by demonstrating relations among representations.

2. Previous Empirical Research on Multiple Representations

Various studies have examined how multiple external representations (presented and
constructed) affect learning outcomes using different manipulations to text, diagrams,
instructions, etc. Mayer, for example, has demonstrated that learners acquire more
knowledge when learning from both text and diagrams, when the two representations
are both informationally relevant [14] presented using temporal contiguity and spatial
contiguity [15,16], non-redundant [14], and presented using both auditory and visual
modalities [17]. Although Mayer has provided much evidence supporting these
multimedia effects, known as the coherence effect, temporal and spatial contiguity
effect, redundancy effect, and split-attention effect, respectively, these learning
sessions were very short (average learning time from above-cited experiments was 120
seconds) and process data (think-aloud protocols) were not collected during the
learning sessions to examine what learning activities the students engaged in while
viewing the presentations. In addition, Mayer’s studies involve pre-recorded
presentations which do not give the learner any control over navigation. As of this
writing, the role of integration of multiple representations in learning with hypermedia
remains unclear.

3. Previous Research on Self-Regulated and Externally-Regulated Learning with
Hypermedia

In order to investigate how students learn about complex science topics using
hypermedia environments, Azevedo and colleagues [18,19,20] have been conducting
experiments on students’ use of self-regulated learning processes during learning
sessions, by collecting think-aloud protocols on each learner. These experiments have
demonstrated that learners acquire deeper understanding of material when they engage
in active learning by setting goals for their learning sessions, monitoring their emerging
understanding throughout the learning sessions, and enact effective learning strategies,
such as coordination of informational sources, selection of new informational source,
summarization, inference generation, hypothesizing, and knowledge elaboration [18,19,20]. They have also demonstrated that adaptive scaffolding [20,21], self-regulated learning training [18], and human tutoring [22,23] aid students in developing more sophisticated mental models at posttest. In regards to the role of multiple representations in self-regulated learning, one study [24] showed that students who spent less time on text-only learned more from pretest to posttest. This suggests that in order for students to gain a deeper understanding of complex science topics, they should visit different representations of the same material within a hypermedia environment.

Several of the learning activities Azevedo and colleagues investigate are relevant to learning with multiple representations. According to Mayer’s [6] and Schnotz’s [7] models of multimedia learning, in order for integration of information from multiple representations to occur, learners must actively integrate the incoming information with long-term memory, indicating that prior knowledge activation is an important activity. Also, students must be aware of the relevance of different representations in order to make appropriate choices about the content they should access (content evaluation). When text and diagram are spatially contiguous, students should engage in coordination of informational sources, by referencing the diagram when necessary while reading the accompanying text. If learners are constructing their own external representations of the material, they will be engaging in taking notes or drawing. Finally, if students are constructing more external representations, one would expect these students to read notes more often as well and possibly review the notes before the end of the learning session.

This paper investigates how access to a human tutor can affect the way learners use various representations and the amount of time these learners spend constructing their own external representations during learning about a complex science topic. The research questions for this study were: 1) How does access to a human tutor affect the amount of time learners spend in different representations of the circulatory system during learning with hypermedia?; and 2) Is there a relationship between amount of time in different representations and learning outcomes?

4. Method

4.1 Participants.

Participants were 82 undergraduate non-biology majors from a large public mid-Atlantic university in the United States. The mean age of the participants was 21 years.


Paper and pencil materials for the experiment included a participant informed consent form, participant demographic questionnaire, and identical circulatory system pretest
and posttest. The circulatory system pretest and posttest were identical to those used by Azevedo and colleagues (e.g. [25]) and included a matching task, a labeling task, and a blood flow diagram task. In the matching task, participants were asked to match 13 circulatory system components to short definitions of the parts. In the labeling task, participants labeled 14 parts of the heart without the use of a word bank. In the blood flow diagram, participants were asked to fill in the order of components of the circulatory system in blood flow (beginning and ending with the superior and inferior vena cava), using a word bank.

4.3 Hypermedia Learning Environment (HLE).

During the learning sessions, all participants interacted with a commercially-based hypermedia learning environment to learn about the circulatory system. The main relevant articles, which were indicated to the participants during a training phase, were ‘circulatory system’, ‘blood’, and ‘heart’. These three articles contained 16,900 words, 18 sections, 107 hyperlinks, and 35 illustrations. All of the features of the system, including the search functions, hyperlinks, table of contents, multiple representations (e.g. pictures, videos, etc.) were available to the participants and they were allowed to navigate freely within the environment to any article or representation.

4.4. Procedure.

Each participant was tested individually in both conditions and participants in the ERL condition were tutored by an individual separate from the experimenter. Participants were randomly assigned to either the SRL (n = 41) or ERL condition (n = 41). Participants were given 20 minutes to complete the circulatory system pretest and immediately given the learning task by the experimenter. Participants in both the SRL group and the ERL group received the following instruction verbally from the experimenter and in writing on a sheet of paper that was available throughout the learning session:

You are being presented with a hypermedia learning environment, which contains textual information, static diagrams, and a digitized video clip of the circulatory system. We are trying to learn more about how students use hypermedia environments to learn about the circulatory system. Your task is to learn all you can about the circulatory system in 40 minutes. Make sure you learn about the different parts and their purpose, how they work both individually and together, and how they support the human body. We ask you to ‘think aloud’ continuously while you use the hypermedia environment to learn about the circulatory system. I’ll be here in case anything goes wrong with the computer or the equipment. Please remember that it is very important to say everything that you are thinking while you are working on this task.

Participants in the ERL condition, in addition to receiving this instruction, had access to a human tutor who scaffolded student’s self-regulated learning by:

(1) prompting participants to activate their prior knowledge (PKA);
(2) prompting participants to create plans and goals for their learning and to monitor the progress they were making toward the goals, and
prompting participants to deploy several key self-regulated learning strategies, including summarizing, coordination of informational sources, hypothesizing, drawing, and using mnemonics.

A tutoring script was used by the human tutor in the ERL condition to guide decision making in when prompts should be used and what kind of prompts to implement, given the current status of the learner. The script was created based on literature on human tutoring [26,27] and recent findings from empirical studies on SRL and hypermedia [18,19,21]. For more information about the tutoring script, please see [22].

4.5 Coding and scoring of product and process data.

This section describes the scoring procedure used for participants’ pretests and posttests as well as the procedure used for coding participants’ use of multiple representations.

Pretest and Posttest scoring procedure. The matching task was scored by giving either a 1 (for a correct match between the concept and its definition), or a 0 (for an incorrect match between concept and definition) on both pretest and posttest (range 0-13). The labeling task was scored by either giving a student a 1 (for a correctly labeled component of the heart), or a 0 (for an incorrectly labeled component of the heart) on both pretest and posttest (range 0-14). The blood flow diagram was scored by giving each student a 1 (for each correctly placed term) or a 0 (for each incorrectly placed term) on both pretest and posttest (range 0-8). The correct progression of the blood flow diagram was: 1) Right atrium, 2) right ventricle, 3) arteries/capillaries/veins or lungs, 4) lungs or arteries/capillaries/veins, 5) left atrium, 6) left ventricle, 7) arteries/capillaries/veins or body, and 8) body or arteries/capillaries/veins.

Use of multiple representations. Students’ use of the various types of representations were coded by viewing the videos of the learners’ interactions with the hypermedia environment. A time segment was coded as ‘text-only’ if the learner was reading text from any of the articles in the environment, with any diagrams appearing on the page occupying less than ten percent of the environment’s real estate. Any time the student was reading text or inspecting a diagram or picture, while diagrams or pictures occupied ten percent or more of the real estate, was coded as ‘text and diagram’. Any time the student was visiting the blood flow video/animation on either the ‘heart’ article or the ‘circulatory system’ article was coded as ‘video’, including times when the video was paused or being controlled by the learner. Finally, a time segment was coded as ‘externally constructed representation’ when the student was taking notes or drawing on paper provided by the experimenter.

5. Results

5.1 Question 1. How does access to a human tutor affect the amount of time learners spend in different representations of the circulatory system during learning with hypermedia? A one-way multivariate analysis of variance (MANOVA) was conducted to determine how access to a human tutor affected the amount of time learners spent in different representations of the circulatory system during learning with hypermedia. The condition had a significant effect on students’ use of various types of
representations, Wilks’s $\Lambda = .58$, $F(4,77) = 14.09$, $p < .001$. The multivariate $\eta^2$ based on Wilks’s $\Lambda$ was strong, .42.

Analyses of variance (ANOVA) on each dependent variable were conducted as follow-up tests to the MANOVA. The ANOVA on text-only was significant, $F(1,80) = 38.88$, $p < .001$, $\eta^2 = .33$, and also the ANOVA revealed significance on externally constructed representation, $F(1, 80) = 32.00$, $p < .001$, $\eta^2 = .29$. Students in the SRL condition spent over twice as much time reading text-only, ($M = 9.25$ mins.), than students in the ERL condition, ($M = 4.36$ mins.). Students in the ERL condition spent over twice as much time constructing their own external representations, ($M = 15.11$ mins.), than students in the SRL condition, ($M = 7.35$ mins.). See Table 1 for means and standard deviations of each type of representation for each condition. There were no other significant differences between conditions.

<table>
<thead>
<tr>
<th>Tutoring condition</th>
<th>Text-only M (SD)</th>
<th>Text+diagrams M (SD)</th>
<th>Video M (SD)</th>
<th>ECR M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-regulated learning (SRL)</td>
<td>9.25 (4.19)</td>
<td>20.08 (4.86)</td>
<td>2.54 (1.90)</td>
<td>7.35 (6.40)</td>
</tr>
<tr>
<td>Externally-regulated learning (ERL)</td>
<td>4.36 (2.78)</td>
<td>18.24 (4.86)</td>
<td>2.10 (0.89)</td>
<td>15.11 (6.01)</td>
</tr>
</tbody>
</table>

5.2 Question 2. Is there a relationship between amount of time in different representations and learning outcomes? Pearson’s correlation coefficients were conducted to determine if there was a relationship between the amount of time in different representations and learning outcomes. The correlational analyses revealed significant, negative correlations between amount of time spent on text-only and matching learning gains, $r = -.28$, $p < .05$; labeling learning gains, $r = -.35$, $p < .01$; and blood flow learning gains, $r = -.45$, $p < .001$. The correlational analyses also revealed significant, positive correlations between time spent on externally constructed representations and labeling learning gains, $r = .28$, $p < .05$; and blood flow learning gains, $r = .34$, $p < .01$. See Table 2 for Pearson’s correlations between time spent on each type of representation and each learning gain measure. There were no other significant correlations.

<table>
<thead>
<tr>
<th>Learning measure</th>
<th>Text-only Pearson’s R (sig.)</th>
<th>Text+diagrams Pearson’s R (sig.)</th>
<th>Video Pearson’s R (sig.)</th>
<th>ECR Pearson’s R (sig.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matching task</td>
<td>-.28 (.01)</td>
<td>.12 (.27)</td>
<td>.06 (.60)</td>
<td>.03 (.77)</td>
</tr>
<tr>
<td>Labeling task</td>
<td>-.35 (.001)</td>
<td>-.11 (.33)</td>
<td>.03 (.79)</td>
<td>.28 (.01)</td>
</tr>
<tr>
<td>Blood flow diagram</td>
<td>-.45 (.000)</td>
<td>-.08 (.46)</td>
<td>-.05 (.64)</td>
<td>.34 (.002)</td>
</tr>
</tbody>
</table>

6. Implications for the design of a computerized SRL tutor
The results from this study demonstrate that access to a human tutor does affect the way that students access different types of representations. Further, results indicate that learners who spend more time reading text (in this hypermedia environment) tend to score lower on learning outcomes and those who spend more time constructing their own external representations tend to score higher on learning measures. This seems to indicate that human tutors, as well computerized tutors designed to foster students’ learning of complex science topics, should attempt to guide students to spend less time simply reading information and more time attempting to construct their own external representations of this information, by fostering the integration of text and diagrams and construction of external representations. Future studies of the effects of various types of learning scaffolds on students’ use of multiple representations should include analyses on the various self-regulated learning activities that students are engaging in while using multiple representations. Researchers should attempt to converge findings on different manipulations to multiple representations that lead to greater learning outcomes [e.g. Mayer and colleagues’ multimedia learning research, 14,15,16,17] with the literature on students’ use of self-regulated learning processes [e.g. Azevedo and colleagues’ work, 18,23].

4. Acknowledgements

This research was supported by funding from the National Science Foundation (REC#0133346 and REC#0633918) awarded to the second author. The authors would like to thank Daniel Moos, Jeffrey Greene, Fielding Winters and Neil Hofman for assistance with data collection and transcribing the audio data.

References


