Interactive Concept Maps and Learning Outcomes in Guru

Natalie Person[1], Andrew Olney[2], Sidney D'Mello[3], and Blair Lehman[2]

[1]Department of Psychology, Rhodes College, Memphis, TN 38152
[2]Institute for Intelligent Systems, University of Memphis, Memphis, TN 38152
[3]Departments of Psychology and Computer Science, University of Notre Dame, South Bend, IN 46556
person@rhodes.edu, [aolney‖balehman]@memphis.edu, sdmello@nd.edu

Abstract

Concept maps are frequently used in K 12 educational settings. The purpose of this study is to determine whether students' performance on interactive concept map tasks in Guru, an intelligent tutoring system, is related to immediate and delayed learning outcomes. Guru is a dialogue based system for high school biology that intersperses concept map tasks within the tutorial dialogue. Results indicated that when students first attempt to complete concept maps, time spent on the maps may be a good indicator of their understanding, whereas the errors they make on their second attempts with the maps may be an indicator of the knowledge they are lacking. This pattern of results was observed for one cycle of testing, but not replicated in a second cycle. Differences in the findings for the two testing cycles are most likely due to topic variations.

Introduction

We are currently developing an intelligent tutoring system for high school biology. Our system, Guru, is a step based system that provides students multiple opportunities to engage with the material. VanLehn (2011) contrasts answer based systems and step based ITSs in regards to the granularity of the interactions that occur. Answer based systems (e.g., most Computer Assisted Instruction (CAI), Computer Based Training (CBT), and Computer Assisted Learning (CAL) systems) typically pose problems and evaluate students' final answers to those problems; the student does all of the reasoning. Step based systems, however, provide students multiple reasoning opportunities either through tutorial dialogue or through prompt based interactions as students work through problems. In a comprehensive analysis that examined the effectiveness of answer based systems, step based ITSs, and human tutoring, all compared to no tutoring learning controls, VanLehn(2011) concluded that the effect sizes of human tutoring are not as large as those that continue to be reported in the literature (i.e., $d = 2.0$, see Bloom, 1984).

Instead, the effect sizes for human tutoring are much lower ($d = .79$) and step based systems ($d = .76$) are almost as effective as human tutoring. The effect size for answer based systems is .31 (Kulik & Kulik, 1991).

Guru would be classified as a step based system in that students participate in tutoring dialogues and are provided multiple opportunities to learn about topics. To our knowledge, Guru will be the first comprehensive ITS for high school biology in that it will cover all topics typically included in a basic biology course. The focus of this paper is to examine student's performance on the interactive concept maps. Specifically, we want to examine whether students' concept map behaviors are related to learning outcome measures.

Guru: An ITS for Biology

Guru is primarily a dialogue based system that is based on the tactics, actions, and dialogue moves of expert human tutors (Olney, Graesser, & Person, 2010). The pedagogical and motivational strategies of Guru are informed by a detailed computational model of expert human tutoring. The computational model includes various levels of granularity from tutorial modes (e.g., lecturing, modeling, scaffolding), to collaborative patterns of dialogue moves within individual modes (e.g., information elicitation, information transmission), to individual dialogue moves (e.g., direct instruction, positive feedback, prompt), to the language, facial expression, intonation, and gestures of expert tutors (Cade et al. 2008; D'Mello et al., 2010; Lehman et al., 2008, Person et al., 2007; Williams et al., 2010). In addition to the tutorial dialogue, there are designated points in the tutoring session where students do other interactive and knowledge assessment activities (e.g., concept maps and Cloze tasks). According to most theories of learning, interactivity generally leads to robust learning (Chi, 2009).

We are currently developing 150 biology topics, all of which are aligned with the Tennessee Biology I Curriculum Standards. The biology content for each tutoring topic varies in length and difficulty, and depending on the student's ability, completing one topic can take 15 to 40 minutes. A typical Guru tutoring session is structured as follows: Preview (brief introduction to the topic), Collaborative Lecture, Summary, Concept Maps I, Scaffolding I, Concept Maps II, Scaffolding II, and Cloze Task.

Collaborative Lecture. Guru lectures are designed to cover predetermined important facts about a topic and are modeled after the lecture styles of expert human tutors. Cade et al. (2008) found that expert human tutors do indeed provide brief interactive lectures when they introduce a new topic to students. These tutoring lectures differ from typical classroom lectures in that students do contribute to some extent. That is, the tutor may ask students simple concept completion questions (e.g., Enzymes are a type of what?), verification questions (e.g., Is connective tissue made up of proteins?), or comprehension gauging questions (e.g., Is this making sense so far?) to ensure the students are paying attention and are engaged with the material. Over the course of a Guru lecture, the tutor covers all of the important facts that the student needs to understand a topic. Two examples of bulleted facts for the topic *Protein Function* are *"Proteins help cells communicate with each other"* and "Antibodies are a type of protein."

During the lectures, the tutor also uses detailed images to elaborate the facts. The images are displayed on a large workspace in front of the tutoring agent (see Figure 1).

Figure 1. Guru interface.

Guru is unique from other dialogue based systems that use images in that Guru's image annotations are carefully integrated with the tutor's dialogue moves.

Summary. After the lecture, the student generates a summary of the material that was discussed in the lecture. Student generated summaries are pedagogically advantageous in that they facilitate organization and retention of the material (Graesser et al., 1995). An enhanced version of Latent Semantic Analysis that includes the span method (Graesser et al., 2005; Hu et al., 2003) is used to compare the student summary to the bulleted facts. Coverage thresholds for each of the bulleted facts are used to determine which information will require additional instruction during the Concept Maps and Scaffolding phases.

Concept Maps. Concept maps are automatically generated from the bulleted facts, and students complete a concept map for each bulleted fact they either omitted or fail to fully cover in their summaries. Students are provided a skeleton map, which means that some of the nodes and/or edges are provided for them; more detail will be provided about these kinds of maps and how they are generated in a subsequent section of the paper. Correct answers with no distractors are provided in the *Concepts* and *Links* boxes on the left of the interface (see Figure 2). Students click on the empty cells and type their answers. Students receive feedback on their answers; correct answers are highlighted in blue, incorrect answers in red.
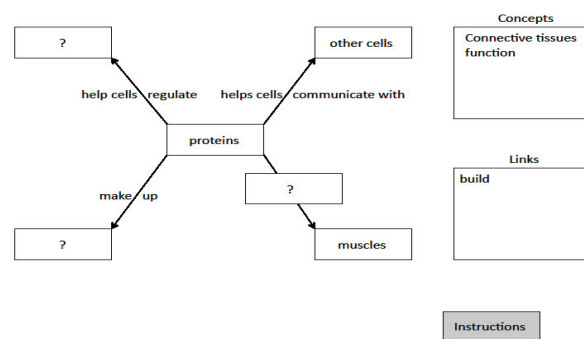

Figure 2. Guru Concept Map Task

Scaffolding. After the student completes all of the concept maps, Guru resumes the tutoring session with dialogue based scaffolding. The scaffolding dialogue covers all of the bulleted facts that were either omitted or not fully elaborated in the student's summary. Currently, Guru adheres to a *Direct Instruction → Prompt → Feedback → Verification Question → Feedback* dialogue cycle to help students learn each important fact.

In an analysis of 40 sessions of expert human tutors, Cade et al. (2008) found that 51% of all the dialogue occurred in Lecture and Scaffolding modes. A Guru tutoring session is structured to resemble those patterns.

As mentioned previously, success on the concept maps does not constitute coverage of the bulleted facts. The bulleted facts are only considered covered when students can provide correct answers to the scaffolding *prompts*.

Students repeat another Concept Map and Scaffolding cycle (i.e., Concept Map II and Scaffolding II) for all bulleted facts that were not covered in Scaffolding I.

Cloze Task. The Guru session wraps up with an interactive Cloze task. Cloze tasks are activities the require students to supply missing concepts from a passage (Taylor, 1953). The Guru passages are the "ideal" summaries for each topic; they include all of the bulleted facts and the important concepts in the concept maps. Although there are multiple methods for deleting concepts from a Cloze passage (Alderson, 1979), we opted for targeted deletion. Specifically, we deleted important biology concepts that were also included in the concept maps. Jongsma (1980) reported that targeted deletion is an effective method for helping students learn new vocabulary. Cloze tasks can either be *open*, where students have to recall the missing concepts, or *closed*, where students have to select the correct concept from a list that contains distractor items. Currently, Guru employs an *open* Cloze that forces students to recall key concepts.

Concept Maps in Biology

Concept maps (a.k.a. knowledge maps) are node link assemblies that represent concepts, ideas, or processes. Concept mapping is a common instructional activity in science education (Fisher, Wandersee, & Moody, 2000; Novak, 1990). Since 1985, over 500 research articles have highlighted the application of concept maps in educational settings; most of these articles have been published since 1997 (Nesbit & Adesope, 2006). In fact, one would be hard pressed to find a modern middle or high school science textbook that does not include concept maps for practice and end of chapter learning activities. Previous research suggests that concept mapping is pedagogically effective in many contexts. In a meta analysis of 55 studies, Nesbit and Adesope (2006) found large learning gains for students creating concept maps ($d = .82$) and medium learning gains for students studying concept maps ($d = .37$) when compared to activities such as reading texts, attending lectures, and participating in class discussions. Creating concept maps from memory and studying pre authored concept maps represent two ends of a scaffolding continuum, whereas the cognitive effort lies entirely with the student when the map has to be created from memory and primarily with the author when a complete map is given to a student to study.

At an intermediate level on the continuum are "expert skeleton concept maps." Expert skeleton concept maps are underspecified maps that are missing edges, nodes, and/or labels for edges and nodes. By varying the amount of missing information, expert skeleton maps can provide more or less scaffolding to the student. Novak and Canas (2006) have suggested that expert skeleton maps would be most beneficial for students with lower prior domain knowledge.

Guru utilizes expert skeleton maps, and the overwhelming majority of students who will use Guru can be categorized as low domain knowledge students. We are aware that other systems incorporate concept map tasks to help students learn science (Biswas et al., 2005; Evens et al., 2001). To our knowledge, however, no studies until this one have examined the usefulness of *skeleton* maps with low knowledge students or have attempted to relate performance on skeleton maps to learning outcomes.

Previous work has investigated the extraction of expert skeleton concept map exercises from textbooks (Olney, Cade, & Williams, 2011). Olney et al. extracted concept maps by defining a set of pedagogically relevant key terms, an ontology for linking terms together, and a set of rules for mapping semantic parses to concept map triples. A triple is a start node, relation, and an end node; an example triple is *golgi body* `is-a` *organelle*. There were, however, two notable drawbacks to this approach that have implications for using this strategy to generate concept maps dynamically in tutoring sessions. The first is that the computer generated maps tended to be sparse, with approximately 3.5 times fewer links than the expert human maps (Olney et al., 2011). Thus the computer generated maps may function as suitable expert skeleton maps in a classroom setting where a teacher can grade the final map, but the lack of links makes it difficult for automated grading of student filled in maps. Secondly, the approach depends on having a large corpus to generate from because the algorithm may fail to extract every triple from every sentence. Instead the algorithm maintains high precision (the maps returned are correct) by sacrificing recall (not all maps are found).

The approach we used to generate concept maps for Guru addresses both of these concerns. First, our approach guaranteed that a concept map triple would be extracted from every sentence. We accomplished this by removing constraints on keywords and the rules for linking nodes together. Secondly, to address the potential increase in errors resulting from relaxing constraints, domain experts reviewed the extracted maps to ensure quality. This human review step parallels prior work done in building treebanks, which has found that it is more efficient for experts to check candidate structures than to create those structures themselves (Marcus et al., 1993).

Concept Map Generation Algorithm. We began by extracting triples from the bulleted facts in the Guru lectures. Each bulleted fact was segmented into three regions: start node, edge, and end node. To accomplish this segmentation, we first parsed each sentence using a syntactic dependency parser. We then applied the following rules to segment each parse into triples using only syntactic information (e.g. part of speech for each

word and the syntactic structure). The start node of the triple is the subject of the sentence minus initial determiners. Edges are the main verb, minus any helping verbs, and include prepositions and other material that occur before the final noun phrase. For example, in "The mitochondrion is an organelle in eukaryotic cells," "mitochondrion" would be the start node, and "is an organelle in" would be the edge. The remainder of this sentence would be used as the end node, e.g. "eukaryotic cells." As is clear from this example, simple sentences are straightforward, but more complex sentences require alternative strategies. For example, subjunctive and relative clauses that occur at the beginning of the sentence are moved to the end of the sentence and become the end node; likewise if they occur at the end of the sentence they are not moved and remain the end node. Space constraints prevent a full accounting of the rules used to extract the triples.

Domain experts reviewed the extracted triples and edited them as they saw fit. Some common edits involved making passive sentences active and/or making active sentences passive so that the resulting start node would match the start nodes of other triples. Radial concept maps like the one depicted in Figure 2 were generated from the triples and presented to students during the Guru session. Students were required to complete a concept map for each bulleted fact they omitted or fail to cover fully in their summaries.

Study Design

The Guru research team is currently partnering with an urban high school in Memphis, TN. The school is a publicly funded Title I charter school for the health sciences. Students who attend this school are recruited from No Child Left Behind high priority middle schools and are, for the most part, low achieving students. Thirty two tenth graders from the school volunteered to participate in the study. All students were enrolled in Biology I. Once a week during regular school hours, students were recruited from class (not biology) to participate in the study. It is worth noting that all students in the state of Tennessee must pass a state mandated Biology I *End of Course* exam to graduate.

The data reported in this paper are part of a larger efficacy study in which Guru was compared to a human tutoring condition and to a classroom control. We opted for a repeated measures design so that over the course of the multi week study all participating students would interact with both Guru and a human tutor multiple times. The tutoring topics (for both Guru and the human tutors) always lagged behind what the biology teacher covered in the classroom by one week. For example, if the teacher covered Topic A (e.g. *Biochemical Catalysts*) one week,

Guru and the human tutors would tutor on Topic A the following week. Students were assigned to conditions so that in a particular week they would receive either Guru or human tutoring. All students received classroom instruction each week.

All knowledge assessments were multiple choice tests consisting of items targeting shallow and deep knowledge, Shallow items targeted direct factual knowledge (e.g., What is mitosis?), whereas deep items required students to reason about the topic and make inferences (e.g., Why do cells need to complete mitosis?). All questions came from either previously administered standardized tests (e.g., *End of Course* exams) or were derived from the bulleted facts for each lecture. Pre and posttests with 12 items were administered at the beginning and end of each tutoring session for both the Guru and human sessions to assess *immediate* learning gains. Items were counterbalanced across pre and posttests by question type. Order of presentation for individual questions was randomized across students.

Delayed posttests with 48 items were given at the end of each design cycle. The delayed posttest consisted of both shallow and deep items. In addition, half of the items were previously seen by students (e.g., on the immediate posttest) and half the items were new. Order of presentation of individual items was randomized across students. A cycle was a two week unit in which students received Guru and human tutoring on two different topics.

As mentioned earlier, the analyses reported here are concerned with students' performance on the concept maps in the *Guru* sessions. In Cycle 1 students received Guru tutoring on one of two topics, *Biochemical Catalysts* or *Testing Biomolecules*. In Cycle 2, they received Guru tutoring on either *Protein Function* or *Facilitated Diffusion*. The analyses in the next section include data from these two cycles.

Data Analyses and Results

The purpose of the research presented here was to determine whether students' performance on concept map tasks in Guru tutoring sessions were related to learning outcomes measures. We tracked the number of errors that students made when attempting to complete the concept maps. *Error rates* were computed by dividing total number of errors by total number of attempts. We also measured how long it took students to successfully complete each map. *Error rates* and *Time per map* (average time spent on each map) were then correlated with immediate and delayed learning gain measures. Scores were proportionalized to range from 0 to 1 prior to computing the learning measures. Corrected learning gains (*CLG*), a normalized gain score that is often used to report learning

gain proportions, [(Posttest Pretest) / (1 Pretest)], were computed for both immediate and delayed learning measures for both cycles.

Pearson correlations were computed to examine the relationships between concept map performance variables and the learning outcome measures. The results from Cycles 1 and 2 are presented in Tables 1 and 2, respectively. Recall that within the course of one Guru tutoring session, the student completes two concept map tasks (Map Task I and Map Task II). It is worth noting that students did indeed learn from interacting with Guru (compared to a classroom control) and the effect sizes were medium to large in magnitude. Space limitations, however, preclude us from reporting those analyses in full.

Cycle 1 Results. For Map Task I, *Time per map* was positively correlated with both of the learning gain measures (see Table 1). That is, the more time spent on the first set of concept maps, the greater the learning gains. *Errors*, however, were negatively correlated with both learning measures for Map Task II and not related to any learning outcomes on Map Task I. Because of the exploratory nature of these analyses, we considered alphas of .15 marginally significant.

Cycle 2 Results. The pattern of results for Cycle 1 did not replicate in Cycle 2. In fact, there were no significant correlations between the concept map variables in Cycle 2 and the learning outcome scores. We believe that these differences may have more to do with topic differences in the two cycles than anything else; these potential topic differences are addressed next.

Table 1. Correlations for Cycle 1

Learning Outcomes	Map Task I		Map Task II	
	Errors	**Time**	**Errors**	**Time**
Immediate CLG	.160	.264*	.409**	.045
Delayed CLG	.244	.410**	.267*	.033

**p < .05,*p < .15

Table 2. Correlations for Cycle 2

Learning Outcomes	Map Task I		Map Task II	
	Errors	**Time**	**Errors**	**Time**
Immediate CLG	.010	.077	.012	.193
Delayed CLG	.065	.056	.019	.133

**p < .05,*p < .15

Topic Differences

The different pattern of results for the two cycles may be due to differences in the topics and to their corresponding Guru lectures. Specifically, we have reason to believe that students found the Cycle 1 topics, *Biochemical Catalysts* and *Testing Biomolecules,* less difficult than the Cycle 2

topics, *Protein Function* and *Facilitated Diffusion*. We have some data to support this. First, students did significantly worse on the Cycle 2 pretest (M = .33 correct, SD = .16) compared to their performance on the Cycle 1 pretest (M = .49 correct, SD = .22), $t(31)$ = 3.50, p < .001. They also did worse on the posttest on Cycle 2 compared to the Cycle 1 posttest (M_{Cycle2} = .60 correct, SD = .19 ; M_{Cycle1} = .71 correct, SD = .23), $t(31)$ = 2.17, p < .05. The two topics in Cycle 1 tended to be more fact based, whereas the Cycle 2 lectures were more process based. We tend to think that students had more difficulty with Cycle 2 topics, although we need to conduct additional analyses to confirm this assertion.

Another difference between the two Cycles involved the lecture length. Cycle 1 lectures were considerably longer than those in Cycle 2, most likely because they included more (but less difficult) bulleted facts to cover. Specifically, the length for Cycle 1 lectures was 112 turns for *Biochemical Catalysts* and 109 turns for *Testing Biomolecules*. For Cycle 2, the average lecture length was 77 turns for *Protein Function* and 67 turns for *Facilitated Diffusion*. If the students received more instruction in the lectures on what we presume to be the easier topics, this may explain some of the patterns we observed. In the future, we may want to consider enhancing the collaborative lectures for the more difficult topics.

General Discussion

The results from Cycle 1, although preliminary, are somewhat promising in that we hope that we will eventually be able to use concept map measures to gauge students' understanding and adjust the tutorial instruction accordingly. Currently, a student's concept map performance does not alter the tutorial dialogue in any way. That is, Scaffolding I dialogue addresses every bulleted fact that students omit or gloss over in their summaries even if they successfully complete concept maps for particular bulleted facts. In Scaffolding II, Guru does not take students' performance on the second set of concept maps into account either. If subsequent analyses resemble Cycle 1, performance on the concept map tasks could be viewed as indicators of student understanding and should probably be used to update the student model.

For example, the time spent per map on students' first attempts with the maps may be a good gauge of their understanding. In the future, it might be worthwhile to monitor whether students are gaming the system (which could be determined with mouse clicks) or whether they are carefully weighing each option before they attempt to complete a map (which could be determined with response times). If we find that students are plowing through the first set of maps, the tutor could intervene and suggest that

they study the map and read each alternative before trying to complete the map.

Students' performance on the second set of maps should probably be evaluated differently. After all, when they receive the second set of maps, they have had three exposures to the material (Lecture, Concept Map I, and Scaffolding I). If students are making mistakes on the second set of maps, chances are they do not understand the material. The negative correlations between Map 2 performance and learning outcomes support this. Currently, students must undergo another round of scaffolding (Scaffolding II) after they complete the second set of maps. It might make more sense to alter Scaffolding II in a way that better targets their knowledge deficits.

The patterns of data for the two Cycles were quite different, and clearly we need to collect more data to determine whether the results from Cycle 1 will replicate in future tutoring studies. These data have certainly provided us with some initial insights as to how concept maps may be used diagnostically and for future analyses. Additionally, these data were analyzed at the subject level. It would be interesting to track the learning trajectory of each bulleted fact for each student from lecture to posttest. Such an analysis would reveal when (or if) learning occurs for a particular fact, and perhaps, which tutoring modality is best suited to teach that fact. After all, some knowledge may be best acquired in tutorial dialogue, whereas other kinds of knowledge may be acquired more easily by studying images or via interactive learning activities.

Acknowledgements

References

Alderson, J. C. (1979). The Cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly*, *13*(2), 219 227.

Biswas, G., Schwartz, D., Leelawong, K., & Vye. N. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, *19*, 363 392.

Bloom, B. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one to one tutoring. *Educational Researcher, 13*(6), 4 16.

Cade,W. L., Copeland, J. L., Person, N., & D'Mello, S. K. (2008). Dialogue modes in expert tutoring. In B. P.Woolf, E. Aimeur, R. Nkambou, & S. P. Lajoie (Eds.), *Intelligent tutoring systems: 9th International Conference, ITS2008* (pp. 470 479). Berlin, Germany: Springer.

Chi, M. T. H. (2009). Active Constructive Interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*, 73 105.

Evens, M. W., Brandle, S., Chang, R. C., Freedman, R., Glass, M., Lee, Y. H., Shim, L. S., Woo, C. W., Zhang, Y., Zhou, Y., Michael, J. A., & Rovick, A. A. (2001). CIRCSIM Tutor: An intelligent tutoring system using natural language dialogue. *Proceedings of the 12th Midwest AI and Cognitive Science Conference* (MAICS 2001), (pp. 16 23). Oxford, OH.

Fisher, K. M., Wandersee, J. H., & Moody, D. E. (2000). *Mapping biology knowledge*. Kluwer Academic Publishing.

Graesser, A. C., Hu, X., & McNamara, D. S. (2005). Computerized learning environments that incorporate research in discourse psychology, cognitive science, and computational linguistics. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications: Festschrift in Honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer* (pp. 183 194) Washington, DC: American Psychological Association.

Graesser, A., Person, N. K., & Magliano, J. (1995). Collaborative dialogue patterns in naturalistic one to one tutoring. *Applied Cognitive Psychology, 9*(6), 495 522.

Hu, X., Cai, Z., Louwerse, M. M., Olney, A. M., Penumatsa, P., & Graesser, A. C. (2003). A revised algorithm for latent semantic analysis. *Proceedings of the 2003 International Joint Conference on Artificial Intelligence* (pp. 1489 1491). San Francisco: Morgan Kaufmann.

Jongsma, E. (1980). Cloze instructional research: A second look. Newark, DE: International Reading Association. Urbana, IL.

Kulik, C., & Kulik, J. (1991). Effectiveness of computer based instruction: An updated analysis. *Computers in Human Behavior, 7,* 75 91.

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta analysis. *Review of Educational Research*, *76(3)*, 413 448.

Novak, J. D. & Canas, A. J. (2006). *The theory underlying concept maps and how to construct them*. *Technical report*, Institute for Human and Machine Cognition, January.

Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching, 27(10),* 937 49.

Olney, A. M., Cade, W. L., & Williams, C. (2011). Generating concept map exercises from textbooks. *Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications* (p. 111 119). Portland, Oregon: Association for Computational Linguistics.

Olney, A. M., Graesser, A. C., & Person, N. K. (2010). Tutorial dialog in natural language. *Advances in Intelligent Tutoring Systems* (Vol. 308, pp. 181 206). Springer Verlag.

Person, N. K., Lehman, B., & Ozbun, R. (2007). *Pedagogical and Motivational Dialogue Moves Used by Expert Tutors*. Paper presented at the 17th Annual Meeting of the Society for Text and Discourse, Glasgow, Scotland.

Taylor, W. L. (1953). Cloze procedure: a new tool for measuring readability. *Journalism Quarterly*, *30*, 415 453.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46(4),* 197 221.

Williams, B., Williams, C., Volgas, N., Yuan, B. & Person, N. K. (2010). Examining the role of gestures in expert tutoring. *Proceedings of the Tenth International Conference on Intelligent Tutoring Systems (pp.* 235 244). Berlin, Heidelberg: Springer Verlag.