Title
The Antecedents of Moments of Learning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 37(0)

Authors
Moore, Gregory R
Baker, Ryan S
Gowda, Sujith M

Publication Date
2015

Peer reviewed

The Antecedents of Moments of Learning

Gregory R. Moore (grm13@my.fsu.edu)
College of Education, Florida State University
1114 W. Call Street, Tallahassee, FL 32306, USA

Ryan S. Baker (baker2@exchange.tc.comlumbia.edu)
Teachers College, Columbia University
525 West 120th St., New York, NY 10027, USA

Sujith M. Gowda (mgsujith@gmail.com)
Metacog, Inc.
55 Linden St., Worcester, MA 01609, USA

Abstract

In this paper, we study the antecedents of moments of particularly successful learning while students use a Cognitive Tutor for geometry. Students used the Cognitive Tutor as part of their regular classroom activities and data was collected automatically. Learning moments were operationalized as when the probability that the student just learned was extremely high, as determined by a probabilistic model: the moment-by-moment learning model. The results indicate that while self-explanation is weakly predictive of learning moments, contextual guessing and several other factors are even better predictors of learning moments. These results suggest that unexpected events in student behavior may be good predictors of changes in knowledge.

Keywords: Moment-by-Moment Learning; Intelligent Tutoring System; Educational Data Mining; Robust Learning

In the process of learning a skill, a learner goes from not knowing the skill, and being unable to demonstrate it, to knowing the skill in a fashion that allows them to demonstrate it. The development of a skill can occur in several fashions; in particular, learning can occur rapidly or gradually over time. In some cases, learning takes the form of a sudden insight, or a "eureka" moment, where the learner gains understanding of a concept in a brief moment. The question of how insight occurs during learning has been an enduring question in Cognitive Science (as discussed in Chu & MacGregor, 2011). There has been considerable research on insight, across decades and in recent years. Much of this research has involved insight problems, which are designed to be solved in a moment of insight after sustained effort (Schooler, Ohlsson, & Brooks, 1993). These problems typically are highly difficult, require a single insight, and have only one correct answer. Insight problems can be a useful instrument to study insight in a controlled, replicable fashion. They have allowed researchers to learn a considerable amount about insight, such as the incompatibility between verbalizing thoughts and solving insight problems (Schooler, Ohlsson, & Brooks, 1993), the ways that external stimuli can facilitate insight (Slepian, Weisbuch, Rutchick, Newman, & Ambady, 2010), and how increased cognitive load can disrupt insight (De Dreu, Nijstad, Baas, Wolsink, & Roskes, 2012).

A criticism of this literature, however, is that insight problems in laboratory settings may not be representative of how “eureka moments” manifest in authentic learning situations (Bowden, Jung-Beeman, Fleck, & Kounios, 2005). The focus on laboratory research on insight problems allows for greater control and facilitates research, given that eureka moments are relatively rare during real-world learning situations. However, if the properties of insight in authentic learning are different—if real-world insight involves problems substantially different than “insight problems”, and if problem-solving manifests differently in real-world contexts, where help and various types of learning support are often available—then the findings of laboratory insight research may not translate to understanding real-world insight during learning (Bowden et al., 2005).

Therefore, it is important for research to examine insight in real-world environments. In this work, we begin to address this need by attempting to examine insight in an intelligent tutoring system—an authentic learning environment. Insight is a difficult construct to measure. In this paper, we operationalize insight as the probability that the student just learned, according to a Bayesian Model that detects sudden shifts from incorrect to correct performance. This operationalization is open to question as a measure of insight, as it is less straightforward than traditional laboratory measures of insight. It may capture the culmination of a student’s thinking that leads to a qualitative and rapid change in performance, rather than the true “eureka” experience. However, this measure has the benefit of being feasible to use to study the phenomenon of insight (or simply moments of rapid learning) in authentic learning contexts and tasks.

We base this work on two recent developments that have made it more feasible to study insight in real-world learning environments. First, the increasing availability of very large data sets from online learning environments, in particular intelligent tutoring systems that reify each of the steps of solving a specific problem (Koedinger & Corbett, 2006), allow us to find many examples of moments of rapid learning. Second, the recent advent of models that attempt to explicitly identify how much learning is occurring moment-by-moment (Baker, Goldstein, & Heffernan, 2011) provides a new opportunity to identify situations where unusually rapid learning occurred and study what differentiated these situations from other situations where less learning occurred. In addition, the longitudinal and intensive nature of this data allows us to not just study what was occurring in those moments of enhanced learning, but also what occurred in the moments leading up to them. As such, for the present study, we combined these two resources to try to better understand what factors precede and are associated with insight.

To do this, we first distilled a range of features of the data for situations where unusually rapid learning occurred in an online learning environment, as well as for the situations and student actions preceding those situations. Though our approach was a bottom-up data mining approach (cf. Baker & Yacef, 2009), we distilled these features with specific candidate hypotheses in mind.

One particularly important candidate hypothesis involved self-explanation. Self-explanation is a self-directed, constructive activity that occurs when a student generates explanations during learning (Conati & VanLehn, 1999;Hausmann, Nokes, VanLehn, & Gershman, 2009). Self-explanation can involve attempting to understand worked examples (Conati & VanLehn, 1999; Shih, Koedinger, & Scheines, 2008), or attempting to understand feedback (Baker, Gowda, & Corbett, 2011). While self-explaining instructional content, students develop an understanding of complex phenomena, actively construct knowledge, and make knowledge personally meaningful (Jordan, Makatchev, & VanLehn, 2003; Roy & Chi, 2005). Self-explanation has been shown to promote deeper processing and more robust learning (Hausmann, Nokes, VanLehn, & Gershman, 2009; Roy & Chi, 2005). Self-explanation's positive effects arise in part because it can expose a student's misconceptions about a concept (Roy & Chi, 2005) and the gaps in the student's knowledge (VanLehn & Jones, 1993). We believe that some of the way that self-explanation may promote robust learning in real-world situations is through promoting "eureka" moments.

Several other candidate hypotheses were also considered. These hypotheses are in line with past evidence from the cognitive and learning sciences that suggest that these specific factors are associated with positive learning outcomes in online learning settings. In particular, we examined the relationships between learning moments and receiving "bug" messages (which inform a student if they have a common misconception), and between learning moments and utilizing online help systems. Each of these experiences (and how students react to them) has been previously shown to be associated with robust learning, and insights are one way that this might occur (Baker, Gowda, & Corbett, 2011).

Finally, we also examined contextual guessing behaviors. Guessing is not mentioned in the literature as being associated with robust learning, but we felt that it was worth examining because it is a behavior that only occurs before learning moments. Additionally, guessing is measured as part of many knowledge modeling frameworks, such Bayesian Knowledge Tracing (described below), but is typically unexamined.

Method

Learning Environment

We studied learning moments within the context of Cognitive Tutor Geometry (CTG), a computer learning environment that promotes learning by doing, currently used by tens of thousands of students a year (Koedinger & Corbett, 2006). In CTG, students individually solve mathematics problems, which are broken down into the series of steps needed to solve them. As a student works through a problem, a running cognitive model assesses whether the student's answers map to correct understanding or to a known misconception (Anderson, Corbett, Koedinger, & Pelletier, 1995). If the student inputs an incorrect answer, the answer turns red. If the student's answer also indicates a known misconception (called a "bug"), the student is given a message about their error.

An important feature of CTG is that students need to input both an answer and a justification for that answer, in the form of a geometric principle. Students can enter their justification either by typing the name of the geometric principle next to their answer or by choosing the geometric principle from a Glossary, which contains a list of theorems and definitions that are relevant to the lesson as well as illustrations and short examples demonstrating those theorems and definitions (Aleven & Koedinger, 2002). In addition to being used for justifying problems steps, the Glossary also acts as a reference for students to use to help them solve the problems (Aleven & Koedinger, 2002).

CTG also has context-sensitive multi-step hints, which are tailored to the exact problem step the student is working on. A student who requests a hint first receives a conceptual hint, and can then request further hints, which become more and more specific until the student is given the answer.

As students work through problems in a specific curricular area, the system uses Bayesian Knowledge-Tracing (Corbett & Anderson, 1995), or BKT, to estimate which skills the student knows and which skills the student is having difficulty with. BKT is a commonly-used student modeling algorithm that infers the probability of student knowledge at a given time based on the student's history of correct and incorrect answers and help requests. BKT also empirically determines the probability that the student got the answer correct without having the necessary knowledge (called the guess probability) and the probability that the student got the answer incorrect even though they had the knowledge (called the slip probability). CTG then uses these estimates to give each student problems that are relevant to the skills that he or she is having difficulty with.

CTG material is structured into independent lessons that each cover a set of related skills and concepts, such as parallel and perpendicular lines, similarity, congruence, volume and surface areas, and vectors. Year-long courses are composed of sequences of lessons, where later lessons build upon knowledge from previous lessons. Log files are automatically collected while the students use the software over the course of the year.

Participants

The data set used in this research comes from the LearnLab DataShop data repository (Koedinger, Stamper, Leber, & Skogsholm, 2013). Data was collected from 102 students at a high school in rural Western Pennsylvania. The students used CTG across the course of the entire school year, approximately two days a week, as part of their regular mathematics curriculum. Students in this school are 98% Caucasian. While this is typical for rural schools in this region, it is higher than the state average (73% Caucasian). There are approximately 16 students per teacher in the school, which is about the same as the state average (15 students per teacher). Additionally, 28% of students in the school qualified for free or reduced lunch, which is slightly less than the state average (33%). In this school, 69% of students were rated proficient or higher on the math section of the PSSA standardized exam, which is approximately equal to the state average (72%). The students were approximately balanced in terms of gender.

Students made 683,285 total transactions with the system (a transaction is defined as any action that the student makes, such as attempting to enter a problem step or asking for help), within 509,854 total problem steps, for an average of 1.34 transactions per problem step. There was an average of 6698.87 transactions per student, an average of 10845.79 transactions per lesson across all students, and an average of 106.33 transactions per lesson per student. There was an average of 4998.57 problem steps per student, an average of 8092.92 problem steps per lesson across all students, and an average of 79.34 problem steps per lesson per student.

Measuring Moment-by-Moment Learning

We computed the probability that a student learned in a specific problem step using the moment-by-moment learning model, also referred to as P(***J***), the probability that the student ***J***ust learned (Baker, Goldstein, & Heffernan, 2011). A high P(***J***) value indicates that there was a high probability that the student learned during the associated problem step. The full mathematical equations for the P(***J***) model are given in Baker, Goldstein, & Heffernan (2011), but we summarize the process here.

The calculation of P(***J***) builds upon BKT and is a two-step process. First, we generate an initial value for each problem step that represents the probability that the student learned a knowledge component or skill on that specific problem step. The assignment of these values is based on the idea that learning is indicated when a student does not know a skill at one point, but then starts performing correctly afterwards. These initial probabilities are generated using a combination of predictions of current student knowledge from BKT and data on future correctness, integrated using Bayes' Theorem. Thus, the calculation uses evidence from both past and future data to assess the probability that learning occurred at a specific time.

Second, these initial probabilities are then used as inputs to a model that infers the probability of learning at a specific problem step based only on past data. This model uses a broader feature set (e.g., response time, use of help, the type of interface widget, and the student's problem-solving history with the tutor), but uses no data from the future. In this way, we create a model that can be used either at run-time or retrospectively to assess the probability that a knowledge component is learned at a specific practice opportunity. This process also "smoothes" model predictions, reducing the degree to which extreme probability values are obtained by chance. This prediction smoothing is useful because it makes the predictions more stable and reliable and, in turn, allows us to examine the predictions more closely.

Data Analysis

To examine insight, we compared two sub-sets of the data – the data associated with the top 1% of P(***J***) values, treated as rapid learning moments, and the data associated with the remaining 99% of P(***J***) values, treated as non-rapid learning moments. This 99/1 split is a somewhat arbitrary designation; it is hard to say if this is too liberal or too conservative. It is possible that not all P(***J***) values in the top 1% are indicative of learning moments. However, moments in the top 1% are definitely more likely to be rapid learning moments than those in the top 50%, for instance.

In order to examine the predictors of these moments, we looked at the preceding problem step on the same skill for each rapid and non-rapid learning moment. Depending on the design of the lesson, the antecedent problem step of the same skill could have immediately preceded the moment or been separated from the moment by several minutes, or even a few days. Each antecedent problem step consisted of one or more student actions, such as asking for help or inputting a response. To create features (discussed below) at the grain-size of problem steps, we averaged the data across all student actions in each individual problem step to create a single value per feature per problem step.

The top 1% of P(***J***) values was determined using all 509,854 P(J) measurements in the data set. However, not all problem steps had an antecedent problem step. Problem steps were only included in the analysis if they had an antecedent problem step on the same skill. This produced a set of 3996 problem steps with a P(***J***) in the top 1% and a comparison set of 467701 problem steps with a lower P(***J***).

In order to better understand the situations in which insights occur, we used features of the antecedent problem steps of rapid and non-rapid learning moments to develop a set of prediction models that attempt to infer whether a problem step will be a rapid learning moment. Specifically, we built a set of step regression models (linear regression with a step function; not the same as step-wise regression), using RapidMiner 4.6 (Mierswa et al., 2006). Step regression models are a method for predicting binary data. In this case, we used them to predict whether an antecedent problem step preceded a rapid learning moment or not. Step regression models postulate that there are sharp disjunctions between the values of a variable. They have been successful in many educational data mining problems, and seem particularly appropriate in this case, as we are trying to infer a sharp disjunction in student learning and performance. In this study, we created one model per potential feature in order to understand the range of features that predict insight.

Potential Predictors of Insight

We distilled a set of features that were potential predictors of insight from the logs of students' interactions with the Cognitive Tutor. These features were quantitative or binary descriptors of key aspects of each problem step and were hypothesized to be associated with the construct of interest, insight. As discussed above, these features were computed using data from the problem step preceding each rapid or non-rapid learning moment.

One of the candidate features we examined was self-explanation. This type of large-scale log data is analyzed retrospectively. Therefore, it was not possible to directly measure whether students were engaging in self-explanation. Instead, we adopted the operationalization used by Baker, Gowda, and Corbett (2011). They suggested looking for when students pause after receiving a bug message or pause after asking for help. Previous research suggests that long pauses in these situations may indicate self-explanation (Shih, Koedinger, & Scheines, 2008). We specifically looked for pauses that were at least 10 seconds long. The cutoff of 10 seconds was chosen because this amount of time indicates that the learner was probably doing something other than just making the next action in the system. These pauses can contain other behaviors, such as off-task behavior (typically 80 seconds or longer – Baker, 2007) or talking to the teacher, but are likely to contain a substantial proportion of self-explanation behavior. Eighteen other features were distilled as well, such as guessing behaviors and the number of actions it took the student to achieve a correct answer, the latter of which may indicate that students are making many mistakes and/or are asking for a lot of help. All of the features distilled represent theoretically-justified hypotheses for factors that may lead to learning moments. Furthermore, these features all represent unique, though potentially correlated, actions and occurrences within the Cognitive Tutor. While a description of all of the features is out of the scope of this paper, six are listed in Table 1 to highlight the most relevant findings.

Metrics Used

We evaluated each of the models using cross-validation. In cross-validation, models are repeatedly built on a subset of the data, and tested on an unseen subset. In this analysis, we cross-validated at the student-level (e.g. the same student was not represented in both the training and test folds), using 6 folds. Cross-validation is an alternative to statistical significance testing that is theoretically equivalent to the Bayesian Information Criterion (BIC) (Raftery, 1995).

The goodness of each model was determined using A', a metric mathematically identical to the Wilcoxon statistic and to AUC, the "Area Under [the ROC] Curve" (Hanley & McNeil, 1982). A' is the probability that if a detector compares a problem step preceding a rapid learning moment to a problem step that is not, it will correctly identify which is which. A model with an A' of 0.5 performs at chance and a model with an A' of 1.0 performs perfectly. In this study, A' was calculated using custom code that can be found at http://www.columbia.edu/~rsb2162/computeAPrime.zip. This custom code avoids the computational errors that are seen in A' implementations that compute the integral of the curve. Cohen's Kappa (1960) is another goodness metric that is often used for models of this type. However, due to the extreme imbalance between the number of cases in the comparison groups, it was not appropriate for this data.

Along with the A' values, we also calculated the means and standard deviations of each feature for each group. These values, in general, represent the approximate proportion of the times that the action associated with the feature occurred. However, because the unit of analysis is problem steps and not all individual actions are treated equally, labeling these as proportions is not quite accurate. Additionally, some means and standard deviations, such as the number of actions in a problem step, represented average counts instead of proportions.

Results

In line with our initial hypothesis, students about to have a moment of rapid learning were more likely to self-explain bug messages and hints (M = 0.120, SD = 0.248) than students not about to have a moment of rapid learning (M = 0.018, SD = 0.107). However, self-explanation was only weakly predictive of rapid learning moments (A' = 0.578). Other features were more predictive.

Contextual guessing, calculated using the model from Baker, Corbett, & Aleven (2008) and defined as having a high probability of getting an answer correct due to guessing rather than knowing the skill, was the strongest predictor of rapid learning moments (*A'* = 0.709). Students who were about to have a moment of rapid learning were more likely to contextually guess (*M* = 0.100, *SD* = 0.128) than students not about to have a moment of rapid learning (*M* = 0.022, *SD* = 0.071). This indicates that guessing may help students learn when they do not understand a skill. Alternatively, it may indicate that students appear to guess when they have developed an understanding that is partially correct and only succeed intermittently.

Table 1: A' values for a subset of the features

Feature	A'
Low Probability of Knowing Before Answering and High Probability of Guessing	0.709
Probability of Knowing Before Answering	0.706
Number of Actions in the Problem Step	0.639
Receiving a Bug Message	0.626
Time > 10 Seconds and Previous Action Help or Bug	0.578
Asking for Help	0.539

The probability of knowing the skill before answering (*A'* = 0.706) was also more predictive of rapid learning moments that self-explanation. Students who had a lower probability of knowing the skill before completing an action were more likely to have a moment of rapid learning (*M* = 0.674, *SD* = 0.357) than those with a higher probability of knowing the skill before completing an action (*M* = 0.889, *SD* = 0.250). This makes sense, as a student cannot have a learning moment if they already know the skill.

As hypothesized, another feature associated with rapid learning moments was receiving a bug message, though this feature was only weakly associated (*A'* = 0.584). Students about to have a rapid learning moment were more likely to receive a bug message (*M* = 0.168, *SD* = 0.312) than students not about to have a rapid learning moment (*M* = 0.053, *SD* = 0.192). This suggests that the feedback present in the bug messages helped the students learn the skill – a positive impact for that aspect of the Cognitive Tutor's design. It is surprising that bug messages were not more predictive of learning moments though. A more detailed examination of bug messages may clarify these results.

However, contrary to our hypothesis, asking for help was not very predictive of rapid learning moments (*A'* = 0.539). Given that help seeking behavior is commonly considered to be good for learning, this is a surprising result. It may indicate that the help being given was only intermittently useful or that students were abusing the help. However, this result requires a more thorough investigation before we can make any conclusions with confidence.

Finally, the number of actions it took a student to get the correct answer to a problem step was also predictive of rapid learning moments (*A'* = 0.639). Students about to have a rapid learning moment tended to make more attempts before getting the correct answer (*M* = 2.199, *SD* = 2.376) than students who were not about to have a rapid learning moment (*M* = 1.347, *SD* = 1.557). This implies that persisting in working on a difficult problem is associated with moments of rapid learning.

Discussion and Conclusions

In this research, we looked to understand when insight occurs within real-world learning contexts by studying large quantities of log files from students using an Intelligent Tutoring System. Specifically, we used the probability that a student had just learned as an indicator of whether insight occurred and compared rapid learning moments (i.e., insights) to non-rapid learning moments (i.e., non-insights) in terms of a variety of features. It is our hope that this research is a first step towards being able to accurately study "eureka" moments in authentic learning environments.

In line with this, these results should be seen as opening up new hypotheses rather than conclusively confirming them. As is true of all measures developed using data mining and knowledge engineering, our operationalizations are imperfect. For the purposes of this study, we have drawn from previous literature to operationalize these constructs as accurately as possible. However, it is difficult to verify the degree to which our operationalizations fully capture these constructs.

Despite this limitation, clear findings emerge from this analysis. We initially hypothesized that self-explanation would lead to rapid learning moments, and this hypothesis was weakly supported by the results. Other successful predictors of moments of rapid learning included the probability of knowing the skill, the number of actions it took to complete the problem step, and receiving a bug message.

However, contextual guessing was the most strongly associated with rapid learning moments. This is interesting because guessing is typically assumed to be associated with behaviors that have negative effects on learning, such as gaming the system (Baker, Corbett, Roll, & Koedinger, 2008). Guessing is an event that occurs unexpectedly. This may mean that unexpected events are good indicators of rapid changes in knowledge. Alternatively, it might mean that the differences between unexpected events and rapid changes in knowledge are hard to discriminate.

For these reasons, future work should focus on clarifying the relationships between rapid learning moments and their antecedents at a finer grain size, especially examining unexpected events such as contextual guessing. Future work should also further examine how closely P(***J***) values relate to insight. One way to approach this may be to distill features that have been shown to be associated with insight (e.g., not verbalizing thoughts, minimized cognitive load) and to see how these features associate with P(***J***) values. In this way, we can better understand the factors that lead students to experience insight, better understand how to design online learning to facilitate learning moments, and help fulfill Anderson's (1993) vision for intelligent tutoring systems as both a way to transform education and a platform for Cognitive Science research.

Acknowledgments

This research was supported by the Pittsburgh Science and Learning Center, NSF award number SBE-0836012.

References

Aleven, V. & Koedinger K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26*, 147-179.

Anderson, J. R. (1993). *Rules of the mind*. Mahwah, NJ: Erlbaum

Anderson, J.R., Corbett, A.T., Koedinger, K.R., & Pelletier, R. (1995). Cognitive Tutors: Lessons learned. *Journal of the Learning Sciences, 4*(2), 167-207.

Baker, R.S.J.d. (2007) Modeling and understanding students' off-task behavior in intelligent tutoring systems. *Proceedings of ACM CHI 2007: Computer-Human Interaction* (pp. 1059-1068).

Baker, R.S.J.d., Corbett, A.T., & Aleven, V. (2008) More Accurate Student Modeling Through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, 406-415.

Baker, R.S.J.d, Corbett, A.T., Roll, I., & Koedinger, K. R. (2008). Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction, 18*(3), 287-314.

Baker, R.S.J.d., Goldstein, A.B., & Heffernan, N.T. (2011). Detecting learning moment-by-moment. *Int'l Journal of Artificial Intelligence in Education*, *21* (1-2), 5-25.

Baker, R.S.J.d., Gowda, S.M., & Corbett, A.T. (2011). Automatically detecting a student's preparation for future learning: Help use is key. *Proc. of the 4th Int'l Conference on Educational Data Mining*, 179-188.

Baker, R.S.J.d., & Yacef, K. (2009) The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*(1), 3-17.

Bowden, E. M., Jung-Beeman, M., Fleck, J., & Kounios, J. (2005). New approaches to demystifying insight. *TRENDS in Cognitive Science, 9*(7), 322-328.

Chu, Y., & MacGregor, J. N. (2011). Human performance on insight problem solving: A review. *The Journal of Problem Solving*, *3*(2), 119-150.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46.

Conati, C. & VanLehn, K. (1999). A student model to assess self-explanation while learning from examples. *Proc. of Seventh Int'l Conference on User Modeling.*

Corbett, A.T. & Anderson, J.R (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, *4*, 253-278.

De Dreu, C. K., Nijstad, B. A., Baas, M., Wolsink, I., & Roskes M. (2012). Working memory benefits creative insight, musical improvisation, and original ideation through maintained task-focused attention. *Personality and Social Psychology Bulletin, 38*(5), 656-669.

Hanley, J., & McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*, 29-36.

Hausmann, R. G. M., Nokes, T. J., VanLehn, K., & Gershman, S. (2009). The design of self-explanation prompts: The fit hypothesis. *Proc. of the Thirty-First Conference of the Cognitive Science Society,* 2626-2631.

Jordan, P., Makatchev, M., & VanLehn, K. (2003). Abductive theorem proving for analyzing student explanations. *Proc. of the Int'l Conf. on Artificial Intelligence in Education* (pp. 73-80).

Koedinger. K. R. & Corbett, A. T. (2006). Cognitive Tutors: Technology bringing learning science to the classroom. In K. Sawyer (Ed.) *The Cambridge Handbook of the Learning Sciences*. Cambridge University Press.

Koedinger, K. R., Stamper, J. C., Leber, B., & Skogsholm, A. (2013) LearnLab's DataShop: A data repository and analytics tool set for cognitive science. *Topics in Cognitive Science, 5*, 668-669.

Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006). YALE: Rapid prototyping for complex data mining tasks. *Proc. of the 12th Int'l Conf on Knowledge Discovery and Data Mining,* 935-940.

Raftery, A. E. (1995). Bayesian model selection in social research (with discussion). *Sociological Methodology*, *25*, 111-195.

Roy, M. & Chi, M. T. H. (2005). The self-explanation principle in multimedia learning. In: Mayer, R.E. (Ed.), *The Cambridge handbook of multimedia learning*. Cambridge University Press, New York.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General, 122*(2), 166-183.

Shih, B., Koedinger, K., & Scheines, R. (2008). A response time model for bottom-out hints as worked examples. *Proc. of the First Educational Data Mining Conference.*

Slepian, M. L., Weisbuch, M., Rutchick, A. M., Newman, L. S., & Ambady, N. (2010). Shedding light on insight: Priming bright ideas. *Journal of Experimental Social Psychology*, *46*, 696-700.

VanLehn, K. & Jones, R. M. (1993). What mediates the self-explanation effect? Knowledge gaps, schema or analogies? *Proc. of the 15th Annual Conference of the Cognitive Science Society*, 1034 – 1039.