Research Statement

Noboru Matsuda
August 10, 2014

Methodological Philosophy

My primary research interest is to advance a transformative theory of effective learning technology and its applications in order to advance cognitive theories of learning and teaching. I am particularly interested in artificial intelligence technologies for students to learn, teachers to teach, and researchers to understand how people both learn and fail to learn. My scholarly expertise thus spans education, learning science, cognitive science, and computer science.

To advance the theory of learning and teaching, I find it particularly important to collect *process* data from studies conducted in authentic classroom settings—known as in vivo studies—that shows detailed interactions between students and the technology. Analyzing process data along with the learning *outcome* data (which usually show test scores and questionnaire responses) often reveals how and why students learn and (more importantly) fail to learn. I apply qualitative and quantitative methods as well as educational data mining techniques to address specific research questions.

To advance a transformative theory of learning technologies, I apply data-driven, iterative design engineering. When designing an innovative learning technology, I start from understanding students and teachers. I apply user-centered methods commonly used in the field of human-computer interaction such as cognitive task analysis, storyboarding, multi-level rapid prototyping, Wizard of Oz, etc. I then test the developed technologies in authentic classroom settings as mentioned above. I use knowledge gained from the data to address research questions for developing effective educational technology and also to identify strengths and weaknesses of already-developed technologies.

Research Accomplishments

While the technologies that I use for my research are generally domain independent, my particular research focus is on mathematics education (which was my primary major in the undergraduate and master's programs). This section describes a few systems that I have developed over the last ten years.

Teaching Geometry Theorem Proving with Construction—a PhD project

My PhD dissertation, under the supervision of Dr. Kurt VanLehn, focused on the science of learning for Euclidian geometry proofs with construction. I developed an intelligent tutoring system as a tool to theorize a better instructional strategy to teach how to write a proof when it requires drawing additional lines as a necessary component of the proof.

I developed an automated theorem-proving system, called GRAMY, that is capable of deriving a human-readable proof with construction (Matsuda & VanLehn, 2004). Before my work, there was no efficient algorithm for theorem proving with construction *suitable for teaching*. Prior techniques were either too ad-hoc or barely comprehensible for students. I solved this problem by inventing a new knowledge representation in which geometric axioms represented as the first order logic are coupled with their diagrammatic configurations.

I then integrated GRAMY into an intelligent tutoring system that teaches theorem proving with construction. I conducted a lab study to test which of the two extreme proof strategies better facilitate students' learning—forward chaining or backward chaining (Matsuda & VanLehn, 2005). Results showed that forward chaining is a better strategy, partly because students frequently have trouble figuring out the premises of the goal to be proven in backward chaining. *This work provided significant knowledge about how to build an effective educational system to teach theorem proving with construction*.

The SimStudent Project – Building Cognitive Tutors by teaching a synthetic student

In 2004, I started a new research project for my postdoctoral research training at the Human-Computer Interaction Institute at Carnegie Mellon University under the supervision of Drs. Ken

Koedinger (Cognitive Science expert) and William Cohen (Machine Learning expert). The goal of the project was to develop an innovative technology for authoring cognitive tutors that allows authors to build an expert model of a cognitive tutor by interactively *tutoring* a machine-learning agent how to solve problems (Matsuda, Cohen, & Koedinger, 2005; Matsuda, Cohen, Sewall, Lacerda, & Koedinger, 2008). The agent that I developed is called SimStudent. SimStudent functions as an intelligent building block for an existing suite of authoring tools called Cognitive Tutor Authoring Tools (CTAT) (Koedinger, Aleven, & Heffernan, 2003). Using CTAT, an author first creates a tutoring interface for the target cognitive tutor. The author then tutors SimStudent using the tutoring interface just created. SimStudent learns a set of production rules to perform the target task that becomes the expert model of the cognitive tutor.

In an evaluation study, we compared the effectiveness of two authoring strategies for creating a cognitive tutor for algebra equations. In the *Authoring by Tutoring* strategy, SimStudent learned skills by being interactively tutored (a model of learning through guided problem solving); in the *Authoring by Demonstration* strategy, SimStudent learned skills by passively observing solutions demonstrated by the author (a model of learning from worked-out examples). The results showed that (1) Authoring by Tutoring is 2.8 times faster than Authoring by Demonstration to author an expert model, and (2) the accuracy of the expert model is better with Authoring by Tutoring due to explicit feedback on errors that the author provides during interactive tutoring (Matsuda, Cohen, & Koedinger, in press). These results suggest that *the SimStudent technology potentially can have a substantial impact on facilitating authoring cognitive tutors, which in turn can have a significant influence on the dissemination of cognitive tutors*.

I was co-PI on this project which was supported by the National Science Foundation: Building Cognitive Tutors with Programming by Demonstration: When Simulated Students help Cognitive Modeling and Educational Studies. William W. Cohen (PI), Kenneth R. Koedinger, and Noboru Matsuda. September 15, 2005 to August 31, 2009. Award No. REC-0537198. Advanced Learning Technologies (ALT). $499,473.

Beyond Authoring by Tutoring – SimStudent as a computational model of learning

SimStudent is a computational model of cognitive skill acquisition from worked-out examples and learning through tutored problem solving. It is both domain general and domain specific model of learning. Using SimStudent technology, researchers can conduct tightly controlled simulation studies to advance theories of learning.

As an example of studying a domain-general theory of learning, I studied the effectiveness of two learning strategies: learning by *tutored problem solving* and learning from *worked-out examples*. The results showed that tutored problem solving led to better learning than learning from worked-out examples in terms of the accuracy of skills learned. The primary advantage of the tutored problem solving was the error detection and correction that is available only when the tutor provides feedback on incorrect skill applications (Matsuda, Cohen, et al., 2008). *These findings provide a theoretical account of the benefit of learning by problem solving over learning from worked-out examples*. As for a contribution to the machine-learning literature, *these findings provide theoretical justifications for the advantage of program execution and feedback for interactive machine learning*.

As an example of studying a domain-specific theory of learning, I studied how students make induction errors in Algebra that lead to learning incorrect skills (Matsuda, Epstein, Cohen, & Koedinger, 2008; Matsuda, Lee, Cohen, & Koedinger, 2009). We hypothesized that induction errors occur due to shallow, perceptually-grounded prerequisite knowledge, where operations are carried out upon algebraic symbols without taking their meanings into account; e.g., reading the '3' in '3*x*' as a number before a letter instead of a coefficient of a variable term. To test this hypothesis, SimStudent's prerequisite knowledge was controlled in a simulation study. The results showed that when the shallow perceptually-grounded prerequisite knowledge is given to SimStudent, the accuracy of learned skills decreased and the speed of learning slowed down significantly. Also, during tutoring, only the SimStudent with perceptually-grounded knowledge made the errors that students commonly make (e.g., subtracting 4 from 3x-4=5). *These results provide a theoretical account for the relationship between conceptual prerequisites and learning to solve problems in Algebra* (e.g., Booth & Koedinger, 2008).

These studies were partly supported by the Pittsburgh Science of Learning Center for the project entitled, "Towards a Theory of Learning Errors: Application of a Synthetic Student to Model How Students Learn Errors," for which I was a Principal Investigator in collaboration with William W. Cohen and Kenneth R. Koedinger (from September 1, 2008 to August 31, 2009 for $72,335).

Studying the Effect of Learning by Teaching – SimStudent as a teachable agent

Despite ample evidence showing that students learn when they teach (Roscoe & Chi, 2008), little is known about how and why students learn by teaching others. Thus, I launched a new project to advance a cognitive and social theory of learning by teaching using SimStudent as a *teachable agent*. I developed an online game-like learning environment where students learn algebraic equations by interactively teaching SimStudent. The particular methodological strength of this project is the collection of detailed *process data* showing interactions between the students and SimStudent. The process data, when combined with *outcome data* (e.g., test scores and questionnaires), allow us to understand how students learn (and fail to learn) by teaching the synthetic peer. The SimStudent technology also allows us to conduct tightly controlled studies in authentic classroom settings as shown below.

I have conducted five classroom (*in-vivo*) studies so far to test specific hypotheses. We have published two journal articles and a dozen peer-reviewed conference papers on this line of research. The most important findings include the following: (a) The SimStudent technology is adequately robust to use in authentic classroom settings (Matsuda et al., 2011). (b) Through teaching SimStudent, students improve their skills of solving equations (Matsuda, Yarzebinski, Keiser, Raizada, Stylianides, et al., 2012). (c) Answering a tutee's questions can facilitate tutor learning when students actually make elaborate answers (Matsuda, Cohen, et al., 2012). (d) Students often tutor SimStudent incorrectly and inappropriately without realizing they are making mistakes. A student's (and SimStudent's) prior knowledge has a significant influence on tutor learning (Matsuda, Yarzebinski, Keiser, Raizada, William, et al., 2013). (e) The student's and SimStudent's learning are highly correlated. When SimStudent commits to shallow learning (due to induction errors), the student's learning is also subpar (Matsuda, Yarzebinski, Keiser, Raizada, William, et al., 2012). (f) Adding a competitive game show to the learning by teaching environment, in which pairs of SimStudents tutored by students beforehand compete against each other by solving challenging problems, increases students' intrinsic motivation (i.e., engagement in tutoring). Some students also showed increased extrinsic motivation (i.e., winning the game), but the extrinsic motivation did not hinder tutor learning. A careful alignment between game-show and learning goals is a key component to facilitate students' learning (Matsuda, Yarzebinski, Keiser, Raizada, Stylianides, et al., 2013).

I have been serving the Principal Investigator on the three grants on this line of research:

(1) Learning by Teaching a Synthetic Peer: Investigating the effect of tutor scaffolding for tutor learning. Noboru Matsuda (PI), Kenneth R. Koedinger, William W. Cohen, and Gabriel Stylianides. October 1, 2013 to September 30, 2016. Award No. 1252440. Research on Education and Learning (REAL). $1,487,349.

(2) Learning by Teaching a Synthetic Student: Using SimStudent to Study the Effect of Tutor Learning. Noboru Matsuda (PI), Kenneth R. Koedinger, William W. Cohen, and Gabriel Stylianides. August 1, 2009 to July 31, 2013. NSF/REESE. Award No. 0910176. $524,439.

(3) Learning by Teaching a Synthetic Student: Using SimStudent to Study the Effect of Tutor Learning. Noboru Matsuda (PI), Kenneth R. Koedinger, William W. Cohen, and Gabriel Stylianides. June 1, 2009 to May 31, 2013. US Department of Education/IES. Award No. R305A090519. $1,413,273.

Evidence-based Online Course Engineering

Since the summer of 2012, I have collaborated with Candace Thille (currently at Stanford University) and Norman Bier at Open Learning Initiative (OLI) at CMU. The goal of the project is to investigate the evidence-based online course engineering to improve the quality of OLI courses (effectiveness, efficiency, and enjoyment) by using online learning data to provide constructive feedback to course developers and instructors. We have conducted contextual inquiry to understand how courseware developers interact with the current tools, followed by detailed interviewing and paper prototyping.

Based on the initial user-centered design research, we proposed to build an **I**ntegrated **D**evelopment **E**nvironment with **A**nalytics (IDEA) for the OLI course developers that provide a WYSIWIG courseware authoring system. The feedback from IDEA will provide assistance with the design and improvement of course activities and interactive instruction. This research seeks to enable course improvements using a data-driven approach by developing methods to measure the effects of course redesign. I will apply machine-learning techniques to data-mine meaningful information from students' learning logs to provide feedback to courseware developers.

The proposal has been recently awarded with myself as a principal investigator: National Science Foundation, Research on Education and Learning (DIR). Data-Driven Methods to Improve Student Learning from Online Courses. Noboru Matsuda (PI), Norman Bier, John Stamper, and Ken Koedinger. August 1, 2014 to July 31, 2016. Award No. 1418244. $504,740.

Future Directions

With the SimStudent technology, I plan to further explore cognitive theories of learning in the STEM field for the next five years. The SimStudent technology has a wide variety of project opportunities to advance theories of learning as well as the knowledge of how to build effective e-learning systems. Currently, I have three low-hanging proposal ideas:

(1) *Exploring the impact of the tutee's question asking to solicit better tutor learning* – It is known that when the tutee asks "good" questions that solicit the tutor's deep reasoning, the tutor learns more (Roscoe & Chi, 2008). A current version of SimStudent has the capability to occasionally ask "why" questions, but it does not actually interpret the student tutor's response to SimStudent. To address this issue, I propose to use an AI technology for natural language processing—SIDE; developed by Carolyn Rose at the Language Technology Institute at Carnegie Mellon University—that will allow SimStudent to understand natural language (i.e., students' response) so that it can provide a follow-up question for students' shallow input (e.g., "I don't understand what you mean by remove the number. Would you elaborate?").

(2) *Building a hybrid-learning environment for learning by tutoring and cognitive tutoring* – Our previous studies show that the accuracy and appropriateness of tutoring is a key factor in successful learning by teaching. Students often tutor SimStudent incorrectly and inappropriately without realizing their mistakes. Students' prior knowledge therefore has a significant influence on tutor learning. In this proposed study, I will develop a new learning environment in which students can freely switch between learning with a cognitive tutor (to prepare for teaching) and learning by teaching SimStudent.

(3) *Learning algebra symbolization with SimStudent as a learning companion* – As a step towards a future expansion of the SimStudent technology, I plan to adapt the SimStudent technology to learn to solve story problems with a particular focus on symbolization. In this context, SimStudent will act as a learning companion that collaboratively solves story problems with students. SimStudent would therefore help students learn symbolization by asking them to make sense of their reasoning and solutions.

In addition to these potential projects, I am also interested in extending the computational model of learning. So far, SimStudent only models inductive learning. Human learning has other aspects including case-based reasoning, analogies, and stochastic learning. Learning from written text and self-regulated learning are also interesting possible extensions of using the SimStudent technology as an intelligent pedagogical agent.

References:

Matsuda, N., Cohen, W. W., & Koedinger, K. R. (2005). Applying Programming by Demonstration in an Intelligent Authoring Tool for Cognitive Tutors *AAAI Workshop on Human Comprehensible Machine Learning (Technical Report WS-05-04)* (pp. 1-8). Menlo Park, CA: AAAI association.

Matsuda, N., Cohen, W. W., & Koedinger, K. R. (in press). Teaching the Teacher: Tutoring SimStudent leads to more Effective Cognitive Tutor Authoring. *International Journal of Artificial Intelligence in Education*.

Matsuda, N., Cohen, W. W., Koedinger, K. R., Keiser, V., Raizada, R., Yarzebinski, E., . . . Stylianides, G. J. (2012). Studying the Effect of Tutor Learning using a Teachable Agent that asks the Student Tutor for Explanations. In M. Sugimoto, V. Aleven, Y. S. Chee & B. F. Manjon (Eds.), *Proceedings of the International Conference on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL 2012)* (pp. 25-32). Los Alamitos, CA: IEEE Computer Society.

Matsuda, N., Cohen, W. W., Sewall, J., Lacerda, G., & Koedinger, K. R. (2008). Why Tutored Problem Solving may be better than Example Study: Theoretical Implications from a Simulated-Student Study. In B. P. Woolf, E. Aimeur, R. Nkambou & S. Lajoie (Eds.), *Proceedings of the International Conference on Intelligent Tutoring Systems* (pp. 111-121). Heidelberg, Berlin: Springer.

Matsuda, N., Epstein, S., Cohen, W. W., & Koedinger, K. R. (2008). Towards a theory of learning errors: application of a synthetic student to analyze students errors *Proceedings of Japan National Conference on Information and Systems in Education* (pp. 362-363). Kumamoto, Japan.

Matsuda, N., Lee, A., Cohen, W. W., & Koedinger, K. R. (2009). A Computational Model of How Learner Errors Arise from Weak Prior Knowledge. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the Annual Conference of the Cognitive Science Society* (pp. 1288-1293). Austin, TX: Cognitive Science Society.

Matsuda, N., & VanLehn, K. (2004). GRAMY: A Geometry Theorem Prover Capable of Construction. *Journal of Automated Reasoning, 32*(1), 3-33.

Matsuda, N., & VanLehn, K. (2005). Advanced Geometry Tutor: An intelligent tutor that teaches proof-writing with construction. In C.-K. Looi, G. McCalla, B. Bredeweg & J. Breuker (Eds.), *Proceedings of The 12th International Conference on Artificial Intelligence in Education* (pp. 443-450). Amsterdam: IOS Press.

Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Stylianides, G. J., Cohen, W. W., & Koedinger, K. R. (2011). Learning by Teaching SimStudent – An Initial Classroom Baseline Study comparing with Cognitive Tutor. In G. Biswas & S. Bull (Eds.), *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 213-221). Berlin, Heidelberg: Springer.

Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Stylianides, G. J., & Koedinger, K. R. (2012). Motivational factors for learning by teaching: The effect of a competitive game show in a virtual peer-learning environment. In S. Cerri & W. Clancey (Eds.), *Proceedings of International Conference on Intelligent Tutoring Systems* (pp. 101-111). Heidelberg, Berlin: Springer-Verlag.

Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Stylianides, G. J., & Koedinger, K. R. (2013). Studying the Effect of a Competitive Game Show in a Learning by Teaching Environment. *International Journal of Artificial Intelligence in Education, 23*(1-4), 1-21 [invited for the special issue on the Best of ITS2012].

Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., William, W. C., Stylianides, G. J., & Koedinge, K. R. (2012). Shallow learning as a pathway for successful learning both for tutors and tutees. In N. Miyake, D. Peebles & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 731-736). Austin, TX: Cognitive Science Society.

Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., William, W. C., Stylianides, G. J., & Koedinge, K. R. (2013). Cognitive anatomy of tutor learning: Lessons learned with SimStudent. *Journal of Educational Psychology, 105*(4), 1152-1163. doi: 10.1037/a0031955