Data mining and education

Kenneth R. Koedinger, Sidney D'Mello, Elizabeth A. McLaughlin, Zachary A. Pardos and Carolyn P. Rosé

Abstract

An emerging field of educational data mining (EDM) is building on and contributing to a wide variety of disciplines through analysis of data coming from many kinds of educational technologies. EDM researchers are addressing questions of cognition, metacognition, motivation, affect, language, social discourse, etc. using data from intelligent tutoring systems, massive open online courses, educational games and simulations, and discussion forums. The data include detailed action and timing logs of student interactions in user interfaces such as graded responses to questions or essays, steps in rich problem solving environments, games or simulations, discussion forum posts, or chat dialogs. They might also include external sensors such as eye tracking, facial expression, body movement, etc. We review how EDM has addressed the research questions that surround the psychology of learning with an emphasis on assessment, transfer of learning and model discovery, the role of affect, motivation and metacognition on learning, and analysis of language data and collaborative learning. For example, we discuss 1) how different statistical assessment methods were used in a data mining competition to improve prediction of student responses to intelligent tutor tasks, 2) how better cognitive models can be discovered from data and used to improve instruction, 3) how data-driven models of student affect can be used to focus discussion in a dialog-based tutoring system, and 4) how machine learning techniques applied to discussion data can be used to produce automated agents that support student learning as they collaborate in a chat room or discussion board.

1. Introduction

Educational data mining (EDM) is an exciting and rapidly growing new area that combines multiple disciplines toward understanding how students learn and toward creating better support for such learning. Participating disciplines (and subdisciplines) include cognitive science, computer science (human-computer interaction, machine learning, artificial intelligence), cognitive psychology, education (psychometrics, educational psychology, learning sciences), and statistics. The data that makes educational data mining possible is coming from an increasing variety of sources and is being used to address a variety of questions about both the psychology of human learning and how best to evaluate and improve student learning (see Figure 1).

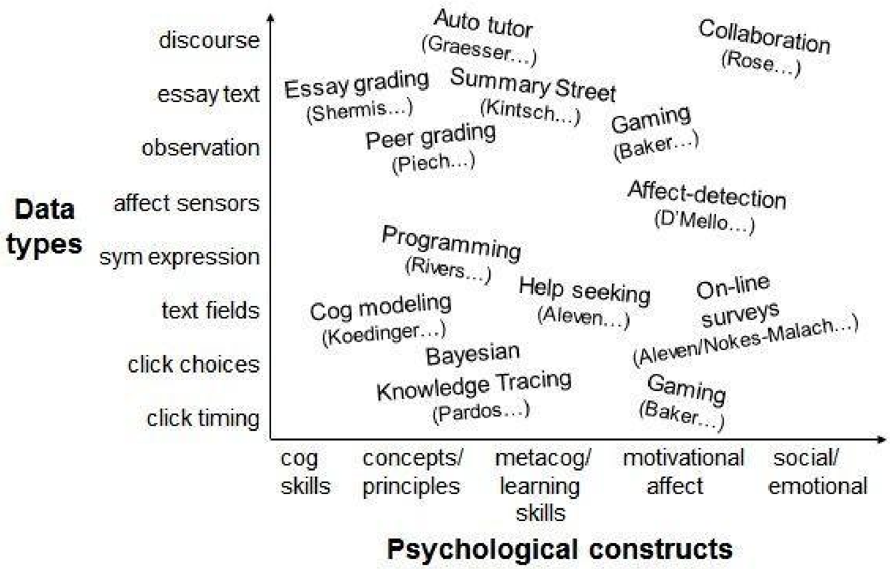

Figure 1. The data that makes educational data mining possible is produced from educational technologies that produce a wide variety of data types (the vertical axis) and is being used to explore a wide variety of psychological constructs relevant to learning (the horizontal axis). Different research paradigms and projects have emerged, exemplified in the content of the figure, and are discussed in the paper.

The vertical axis of Figure 1 illustrates the variety of educational data types that come from different sources. Starting at the bottom are simpler data types including menu- based choices with their timing coming from mouse clicks, such as multiple choice questions in MOOCs and other online courses, educational games and simulations, and simple tutoring systems. Moving up the vertical axis, more complex problem -solving environments and intelligent tutoring systems produce data types that include short text inputs and symbolic expressions. Researchers are also building affect sensors and doing classroom observations of student engagement and emotional states. At the top of Figure 1 are more complex forms for student writing or sometimes speaking,[1-2] including student answers to essay questions[3] or their dialogues within chat rooms or discussion boards.

The horizontal axis of Figure 1 illustrates the spectrum of psychological constructs that educational data miners have been exploring with these kinds of data. These constructs include tacit cognitive skills, explicit concepts and principles, metacognitive skills and self-regulatory learning strategies, student affect and motivations, skills for social discourse and argument, and socio-emotional dispositions and strategies.

Within Figure 1, we illustrate a sampling of research projects/paradigms that have used different types of data to advance understanding of different psychological constructs. Cognitive skills and principles have been explored using many different types of data. Data on student correctness of their click choices or their text or symbolic input over opportunities to practice has been explored in models of learning, especially Bayesian Knowledge Tracing (e.g.,[4-6]). That same kind of data has been used to evaluate cognitive models and aid discovery of data-driven cognitive model improvements (e.g.,[7,8]). Complex symbolic problem solutions (e.g., computer programs) are being analyzed to understand changes in students' skills and strategies over time (e.g.,[9,10]). Peer grading of open-ended writing (e.g.,[11]) and interface design (e.g.,[12]) provides the basis for mining both the quality of student responses and the quality of their ability to evaluate them. Data corpora of written essays and answers, from students and experts, have been used to create automated methods for essay grading (e.g.,[13]) and dialogue-based intelligent tutors[1,14-15] or dialogue support for collaborative learning .[16]

Moving to the right in Figure 1, student metacognition and self-regulatory learning has been explored using student data from intelligent tutor interaction[17,18] and from multimedia interaction.[19] Such data has been also used along with formal classroom observation data to drive machine-learned detectors of forms of student disengagement.[20,21] This data has been further augmented with real-time surveys[22] and with affect sensors.[23] Addressing issues of social interaction, researchers have analyzed student collaboration data (e.g.,[24]) including student entries in MOOC discussion boards.[25-28]

Another dimension of relevance, not shown in Figure 1, is the time course of observations reflected in the data. Click choice and timing data observations tend to occur in the range of 100s of milliseconds to 10s of seconds. Text fields and symbolic expression data tend to take longer 10s to 100s of seconds. Essay and discourse data points may take many minutes to produce. As discussed elsewhere,[29,30] the complexity of psychological constructs (roughly, left to right in Figure 1) tend to occur at a correspondingly increasing time scale, with the exception of affective physiological data, which can operate in the millisecond range.

Educational Data Mining is defined by Baker and Yacef[31] as "an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in." Other empirical methods, besides EDM, for driving discoveries relevant to education include experimentation, surveys, and design research. EDM is set apart primarily by the fact that wide use of educational technology is producing volumes of ecologically valid data on student learning in a context in which experiments and surveys can be embedded. It is similar to design research with a focus on data from natural settings, but rather than being qualitative, it is different (complementary) in its focus on quantitative analysis using machine learning and statistical methods. It is similar to experimentation and survey methods in that those methods can be employed in the context of educational technologies, but is broader in that it also includes analysis of naturally-occurring student interaction data either without or in conjunction with experimental manipulation or embedded surveys.

Research fields are grounded in the community of engaged researchers and in the conferences and journals in which they communicate. EDM is no exception. The primary conferences are Educational Data Mining (EDM), which began in 2008, and Learning Analytics and Knowledge (LAK), which began in 2011. Each has peer reviewed, published proceedings. The International Journal of Educational Data Mining began in 2009 and the Journal of Learning Analytics began in 2014. Elements of EDM research have a longer history as revealed by deep analyses of learner data in publications in the proceedings of the Intelligent Tutoring Systems conference (established in 1988), the Artificial Intelligence in Education conference (established 1997), and the Cognitive Science conference (established 1979). More recently, an expansion of interest in EDM has been revealed through workshops and papers in other conferences, both pre-existing (Neural Information Processing Systems, and Knowledge, Discovery, and Data Mining) and new (Learning at Scale).

This article is organized around different kinds of research questions regarding the psychology of learning that educational data mining research has been addressing. Section 2 discusses the status of current EDM research especially on questions of assessment, transfer, motivation, and language. Section 3 provides related challenges and future research opportunities.

2. Educational Data Mining Addresses Research Questions About the Psychology of Learning

Educational Data Mining has addressed a variety of research questions regarding the psychology of learning. The following five sections discuss assessment of cognition, learning, and achievement (Section 2.1), transfer of learning and discovery of cognitive models (2.2), affect, motivation, and metacognition (2.3), language and discourse analytics (2.4), and other applications of educational data mining (2.5). In each section, we summarize both technical and empirical research that is sometimes descriptive, yielding new insights about the nature of learning, and sometimes prescriptive, indicating methods for yielding better student learning outcomes. In particular, we illustrate connections from efforts to use data to build models that reflect insights on learning to efforts to "close-the-loop" by translating those models or insights into systems and testing whether they produce better learning outcomes. A desirable sequence in a program of research (which often involves multiple projects/papers and may involve different teams) starts with data mining yielding new statistical models of that data. Next an adaptive system is built using the resulting statistical models. Sometimes that model is used directly, for example, as in a detector of student disengagement behavior that prompts students to re-engage[32] or a language interpretation system that provides students feedback in response to their typed entries[33-37] or provides recommendations for how to engage more productively.[38] In other cases, the model resulting from data mining may be interpreted for insights that can be employed to an adaptive system design (e.g.,[39,40]). In both cases, a desirable final step is to "close-the-loop" by running an experiment (an "A/B test") comparing the original adaptive system to one with the data-driven adaptive features.

2.1 Assessment, growth modeling of learning, and prediction of achievement

Educational technologies provide rich data for understanding and developing assessment of students' skills, concepts, mental models, beliefs, and learning trajectories. Collectively, these aspects of human intelligence can be referred to as a "knowledge base" with elements of this knowledge base referred to as "knowledge components" or KCs. Koedinger, Corbett and Perfetti[41] define a KC as "an acquired unit of cognitive function or structure that can be inferred from performance on a set of related tasks."

Two aspects of assessing these knowledge components have defined themselves in educational data mining research. The first being the statistical model, a mathematical abstraction of student behavior measurements, and the second, the cognitive model, which involves the mapping of knowledge components to items or tasks. This knowledge-to-task mapping provides a way to convert a symbolic cognitive model, which a cognitive scientist might produce, into a form that can be used, along with the statistical model, to make predictions about student performance. This specialized form of a cognitive model is referred to as a knowledge component model (in educational data mining), a Q matrix (in Psychometrics), or a student model (in AI in Education). The statistical model and the cognitive model are inexorably linked, conceptually, when referring to a predictive model of student learning. However, the mathematics for quantifying learning (the form of the statistical model) has been an active research area in and of itself, separable from the question of identifying an empirically justified cognitive model, which has been in equal measure represented in the literature.

2.1.1 Student performance prediction showcase

An Educational Data Mining competition in 2010[1] provided a forum to compare statistical models of student performance and learning from EDM to models from the broader machine learning community. The competition scored the models on their ability to forecast the correctness of student responses to multi-step problems within Intelligent Tutoring Systems for junior high mathematics. Table 1 depicts a simplified example of the student event log dataset that was provided by the competition. Each row in the dataset represented a student action or a "step" toward solving a math problem given by a Cognitive Tutor, which was the source of the competition's data. The modeling task was to use past information, such as a students' responses related to a particular skill or knowledge component (KC), in order to predict the probability of correct responses on future problem steps. There were 24 million rows worth of past information and 8 million rows to be predicted. Selected meta-information about the row being predicted was omitted, such as time to response and number of attempts, but information such as the student, problem, and knowledge component associated with the problem step remained. Participants were scored on the average error

[1] http://pslcdatashop.web.cmu.edu/KDDCup/

between their predictions of correctness and the actual correctness.

Table 1. Simplified sample of data used in the KDD Cup competition.

Sequence ID	**Student ID**	**Problem ID**	**Step ID**	**Answer**	**KC**
1	S01	WATERING VEGGIES	WATERED AREA-1	Incorrect	Circle-Area
2	S01	WATERING VEGGIES	WATERED AREA-2	Correct	Circle-Area
3	S01	WATERING VEGGIES	TOTAL AREA	Correct	Rectangle-Area
4	S01	MAKING CANS	POG AREA Q3	Correct	Circle-Area

The most successful approach to the prediction task featured a combination of Hidden Markov Models, random decision trees, and logistic regression on millions of machine generated features of student interaction.[42] The second best overall prediction was achieved with a collaborative filtering approach[43] much like the methods used in Netflix movie rating predictions. In this method orthogonal matrices of latent factors of students and questions are machine learned and multiplied together to recover a matrix of predicted student responses to steps. Standard matrix factorization methods do not take into account time (and thus ignore learning); however, expanding the dimensions to include time, using tensors, has resulted in significant predictive gains in subsequent work.[44] Finally, the fourth place finisher (3rd place did not disclose their methods) featured random decision trees and a student individualized Bayesian Network model,[45] a computational form of cognitive diagnostic models that leverages inference techniques and optimizations from the broader family of probabilistic graphical models.[46] All top participants combined their featured statistical model with other methods in an ensemble in order to maximize predictive performance. Using ensembles of methods instead of a single best method produced a 10% gain in prediction accuracy when applied to a different 8th grade math tutoring platform.[6]

While the 2010 competition demonstrates how Educational Data Mining research explores the boundaries of predictive models, the field has focused even greater attention on the science of learning, seeking explanation for knowledge gain and performance through the careful and deliberate investigation of learning data.

2.1.1 Bayesian Knowledge Tracing based models of learning

Bayesian Networks provide a computational approach to modeling student learning by elegantly capturing the dynamic of knowledge probabilistically. The base Bayesian Networks model of reference, called Knowledge Tracing,[4] has roots in Cognitive Science. Atkinson and Paulson[47] laid the foundation for the model of instruction and Anderson[48] formalized a continuous increase in the activation level of procedural knowledge with practice, which is mapped to a probability in Knowledge Tracing. The standard Bayesian Knowledge Tracing (BKT) model, shown in Figure 2, can be described as a Hidden Markov Model (HMM) with such a model being defined for each KC. The acquisition of knowledge is defined in the binary latent nodes while the correctness of responses to questions is defined in the observables.

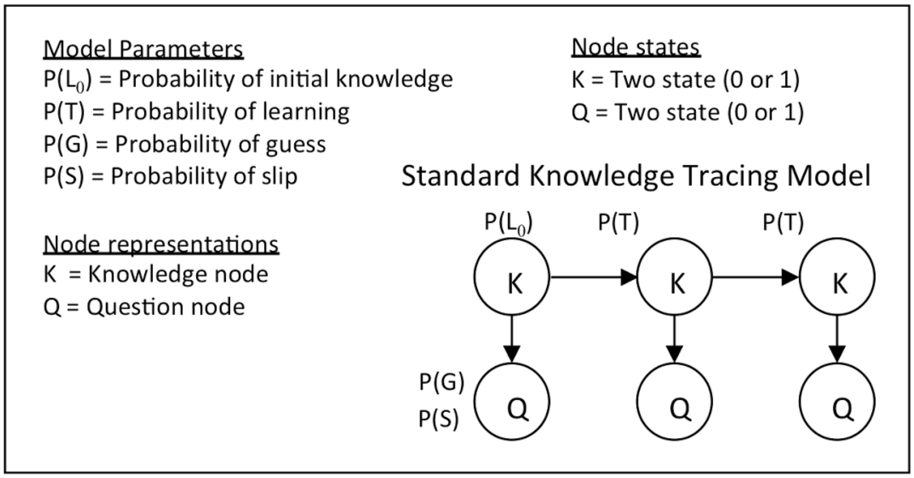

Figure 2. Representation of the Standard BKT model with parameter and node descriptions.

Research on and expansion of the BKT model has been fueled by advances in statistical parameter learning and by the use of the model in practice inside the popular Cognitive Tutor(c) suite of ITSs[49] where BKT is used to determine when a student has mastered a particular KC and is thus no longer asked to answer steps consisting only of mastered KCs. The standard BKT model has four parameters per knowledge component: 1) a single point estimate for prior knowledge, 2) a learning probability between opportunities to practice, 3) a guessing probability for when a student is correct without knowing, and 4) a slip probability for when a student is incorrect despite knowing. Pardos and Heffernan[5] introduced student individualization to the BKT model through modifying the conditional graph structure of the model. Specifically, they found that by using a bimodal prior, bootstrapped on the student's first response, prediction improved on 30 of 42 datasets (avg. correlation increase from .17 to .30). This particularly individualized prior approach allowed for students to be in one of two priors, adding only a single parameter per KC over the standard four parameter per KC model. Performance prediction is often used as a proxy for assessment quality in EDM models with error and accuracy metrics signifying the goodness of a model. While these metrics are a convenient way to quantitatively compare models, they can be unsatisfying in explaining the real world impact of prediction improvement on an individual student. Addressing this, Lee and Brunskill[50] evaluated the impact of model individualization on the average number of under and over practice attempts compared to a non-individualized model. They concluded that individualized BKT can better ensure that struggling students get the practice they need and that better students are not wasting time with extensive busy work (i.e., about 20% of students would be given *twice* the practice opportunities if standard BKT is used than they would be if individualized BKT is used). Individualization in student modeling is further explored by Yudelson, Koedinger, and Gordon[51] and in section 3.1.

2.1.2 Logistic regression based models of growth

Logistic regression models are another family of statistical models, which can be compared with the BKT family of models (see Table 1). Logistic regression models extend a history of elaborations on Item Response Theory including the addition of a knowledge component model (or "Q- matrix") in the Linear Logistic Test Model (LLTM).[52] Further additions were made to model learning across tests[53] and within an intelligent tutor.[54] These learning models introduced a growth term that models learning by including an estimated increment in predicted success on a KC for each time a student gets practice on an item (or problem step) that needs that KC, as indicated by the Q matrix. This model, later called the Additive Factors Model (AFM), was picked up in efforts to evaluate and search for alternative cognitive models.[55] Other efforts have explored logistic regression variations including modeling learning with a separate count of correct versus incorrect practice attempts of a knowledge component (PFA),[56] modeling student variations in learning rate (e.g.,[57]), modeling different learning rates depending on the nature of instruction such as feedback practice versus seeing an example (IFM).[58]

2.1.3 Comparison of model approaches

As a way of summarizing, Table 2 indicates features of an incomplete, but representative set of statistical models that have used been used to assess student proficiency and, in most cases, student learning. It illustrates features of two families of statistical models, logistic regression and Bayesian Knowledge Tracing, by indicating the way the models parameterize student and task domain attributes. As indicated in the Student parameter column, most of the models have a parameter value for each student that indicates that student's overall general proficiency. Notably Bayesian Knowledge Tracing (row 7) does not. However, approaches, such as iBKT, have extended BKT to include a student proficiency parameter (row 8).

In the terms used by Wilson and De Boeck,[52] the models that separately represent difficulty of each task are "descriptive" whereas the models that represent fewer latent knowledge components across multiple tasks are "explanatory" in that they provide an account for task performance correlations (same KCs) and performance differences (different KCs). In the Task parameters column, we see that most of these models include knowledge components as latent variables for assessing task difficulty. Some models, in contrast, have a parameter for each unique task or item, particularly IRT (row 1) and PFA (row 4).

Older psychometric models, such as IRT 3PL (row 1) and LLTM (row 2) do not model learning (see the Learning parameters column) but have useful features (e.g., a guess parameter) that could be incorporated into a learning model. All the BKT models and the newer logistic regression models model learning by including estimates of learning for each knowledge component. These models use knowledge components as the basis for explaining learning, though it is possible to have a descriptive model of learning (not shown in Table 1) that would have a single learning parameter across all tasks, as in a faculty theory of transfer (cf.,[59]).

Table 2. Various statistical models and parameters that have been used to assess student proficiency.

			Parameters					
Row	Name[1]	Statistical Family	Global Student	Task	Learning	Performance	Prior Success/ Failure	Sample References
1	IRT 3PL	Logistic regression	Proficiency (1/stu)	Item	none(0)	Guess(1)	no	Wilson&DeBoeck,2004
2	LLTM	Logistic regression	Proficiency (1/stu)	KC	none(0)	none(0)	no	Wilson&DeBoeck,2004
3	AFM	Logistic regression	Proficiency (1/stu)	KC	KC (=KC#)	none(0)	no	Spada & Magaw, 1985; Cen, 2009
4	PFA	Logistic regression	none (0)	Item	KC (=2*KC#)	none(0)	yes	Pavlik et al., 2009
5	IFM	Logistic regression	Proficiency (1/stu)	KC	KC (=3*KC#)	none(0)	yes	Chi et al., 2011
6	CFM	Logistic regression	Proficiency (1/stu)	KC multi-skill	KC (=KC#) multi-skill	none(0)	no	Cen et al., 2008; Cen, 2009
7	BKT	Bayesian Knowledge Tracing	none(0)	KC	KC (=2*KC#)	Guess&Slip (2)	yes	Corbett & Anderson, 1995
8	iBKT	Bayesian Knowledge Tracing	Proficiency (1/stu)	KC	KC (=2*KC#)	Guess&Slip (2)	yes	Pardos et al., 2010; Lee et al., 2012; Yudelson et al., 2013
9	cBKT	Bayesian Knowledge Tracing	none(0)	KC multi-skill	KC (=2*KC#) multi-skill	Guess&Slip (2)	yes	Koedinger et al., 2011

1 IRT3PL is Item Response Theory 3 Parameter Logistic Model; LLTM is Linear Logistic Test Model; AFM is Additive Factors Model; PFA is Performance Factors Analysis; IFM is Instructional Factors Analysis Model; CFM is Conjunctive Factor Model; BKT is Bayesian Knowledge Tracing; iBKT is Individualized BKT; cBKT is Conjunctive BKT

The Performance parameters column shows how the BKT models distinguish themselves from logistic regression models by including explicit parameters to estimate the amount of student guessing or slipping. Logistic regression models of learning have not included performance parameters, but as illustrated by IRT 3PL (row 1), it is possible to incorporate such parameters into a logistic regression model.

The final Prior Success/Failure parameters column illustrates the difference in models that have explicit differentiation of past student task success or task failure. These models are more dynamic in their ability to adjust to student and knowledge specific variations in performance. The success of these models over alternative models that do not make this distinction[52,53,60,61] suggests there are student by knowledge component interactions in learning rate.

2.2 Transfer of learning and discovery of cognitive models from data

Most of the statistical models in the prior section assume a cognitive model of student knowledge or skill. But how are these cognitive models created and how can they be evaluated for their empirical validity? Data mining provides an answer. Further, data-driven cognitive model improvements can be used to design novel instruction that yields better student learning and transfer.

The general version of this question of how cognitive models are discovered is addressed by techniques for cognitive task analysis (CTA),[62,63] particularly qualitative methods such as interviews of experts or think alouds of expert or student problem solving. CTA is a proven and relevant method for improving student performance and learning through instructional redesign (e.g., medical[64]; military[65];

mathematics[66]; aviation industry[67]). However, CTA methods are costly both in time and effort, generally involve a small number of experts or cases, and are mainly powered by human action.

Thus, new quantitative approaches to cognitive task analysis that use educational data and promise improved efficiency have emerged. The key idea is that a good cognitive model of student knowledge should be able to predict differences in task difficulty and/or in how learning transfers from task to task better than an inferior cognitive model.

One such method, called difficulty factors assessment (DFA), goes beyond experts' intuitions by employing a knowledge decomposition process that uses real student log data to identify the most problematic elements of a given task. The basic idea in DFA is that when one task is much harder than a closely related task, the difference implies a knowledge demand (at least one "knowledge component") of the harder task that is not present in the easier one. In other words, one way to empirically evaluate the quality of a cognitive model is to test whether it can be used to accurately predict task difficulty (cf.,[68-72]). Cognitive models have been developed using this technique in domains including algebraic symbolization,[73,74] geometry,[75] scatterplots[76] and story problem solving.[77,78]

To further accelerate and scale the process of improving cognitive models, automated techniques have been developed that make analysis more efficient (i.e., search a much larger space in a reasonable amount of time) and more effective (i.e., by reducing the probability of human error). Furthermore, these directed search techniques maintain skill labels from the student model unlike many of the prominent machine learning techniques that have been applied to educational data that result in unlabeled, difficult or impossible to interpret models. The models created by an automated process (e.g., Learning Factors Analysis, LFA) are interpretable and have led to real instructional improvements.[79,41]

Early automated efforts in model discovery involved human generation of model improvements followed by machine testing (i.e., model comparison) using a variety of established data mining metrics (e.g., Akaike information criterion (AIC), Bayesian information criterion (BIC), and cross validation (CV).[79] More sophisticated automated techniques have been developed such as Learning Factors Analysis (LFA),[55] Rule Space,[80] Knowledge Spaces,[81] and matrix factorization.[82,83]

Table 3. Example Q matrix, P matrix and the resulting Q' matrix from a split on Negative Result.

	Q		P		Q' split [Q,Negative Result]		
Problem Step	**Multiply**	**Subtract**	**Negative Result**	**Order of Op**	**Multiply**	**Sub-Positive**	**Sub-Negative**
2*8-30=> 16-30	1	0	0	0	1	0	0
16-30=> -14	0	1	1	0	0	0	1
30-2=> 30-16	1	0	0	1	1	0	0
30-16=> 14	0	1	0	0	0	1	0

In LFA, a search algorithm automatically searches through a space of knowledge component (KC) models represented as Q-matrices[80,84] to uncover a new model that (best) predicts student-learning data. The input to LFA includes human-coded candidate factors (called the P-matrix) that may (or may not) affect student performance and learning. Table 3 shows a simple illustration of the mapping of problem steps to Q and P matrices and how a new Q-matrix is created (see Q') by applying a split operator to a factor in the Q-matrix (e.g., Subtract) using a factor from the P-matrix (e.g., Negative Result). Additional new factors are created as a consequence of the split (e.g., Sub-Positive and Sub-Negative) that make different (and better) predictions about task performance and transfer.

Koedinger, McLaughlin and Stamper,[7] applied a form of the LFA algorithm to 11 datasets from different technologies (e.g., intelligent tutors, games) in different domains (e.g., mathematics, language). In all the datasets, a machine generated cognitive model was more predictive (according to AIC, BIC and CV) of student learning than the models that had been created by human analysts. Having a general empirical method for discovering and evaluating alternative cognitive models of student learning is an important contribution of EDM to cognitive science. One specific example of an LFA discovery comes from its application to data from student learning in cognitive tutor geometry unit on finding area. The original cognitive model in that tutor distinguished (had separate knowledge components), for each area formula, whether the formula is applied in the usual "forward" direction (to find the area given linear measures) or "backward" (given the area, find a missing linear measure). LFA revealed that this distinction was mostly unnecessary – backward application was no harder and there is no lack of transfer from practice on forward applications. However, the distinction is critical for the circle area formula. Finding the radius given the area of a circle is much harder than finding the area given the radius and there is, at best, only partial transfer from forward application practice to backward application performance.

This empirical method for comparing alternative cognitive model proposals has also been used to demonstrate that automatically generated cognitive models, produced by a computational model of human learning called SimStudent, are often better than models built by hand.[85] For example, when trained on algebra equations, SimStudent learned separate knowledge components (production rules) for removing the coefficient in problems such as "5x = 10", where the coefficient is a number, from problems such as "-x = 10", where the coefficient is implicit. The hand built model did not make this distinction, but the learning curve data support it as a real difference in cognitive processing and transfer. Students make twice as many errors when the coefficient is implicit rather than explicit and practice on the explicit problem type does not transfer to performance on the implicit problem type.

Making better predictions about student learning is a first step toward both improving understanding of student cognition underlying learning transfer and improving learning outcomes. A second step is interpretation of data mining results in terms of underlying cognitive mechanisms. For example, consider the data mining discovery, mentioned above, that backward applications of area formulas are harder than forward applications for circle and square area but not for other figures. A cognitive interpretation is that students have trouble learning when to use the square root operation (i.e., to find s in A=s^2 or r in A=pi*r^2 given A). Such cognitive interpretation not only suggests a scope of generalization and transfer that can be tested (cf.,[86]), it also critical to translating a data mining result into a prescription for instructional design. In this case, it suggests practice tasks, examples, and verbal instruction that precisely target good decision making about when to use the square root.

A final step toward both better understanding of transfer and better student learning is to run a "close-the-loop" experiment that compares student learning from instruction that is redesigned based on the data mining insight with learning from the original instruction. For example, Koedinger, Stamper, McLaughlin and Nixon[40] showed how a data mining discovery revealing hidden planning skills could be translated into a redesign of an intelligent math tutor that produced more efficient and effective student learning.

2.3 Affect, motivation, and metacognition

It is widely acknowledged that affect, motivation, and metacognition indirectly influence learning by modulating cognition in striking ways.[87-89] In line with this, EDM researchers have been studying these processes as they unfold over the course of learning with technology.[90] For example, researchers are interested in analyzing log traces to uncover events in a learning session that precede affective states like confusion, frustration, and boredom. They might also analyze physiological data to develop models that can automatically detect these states from machine-readable signals, such as facial features, body movements, and electrodermal activity. Similarly, researchers might be interested in real-time modeling of log traces in order to make inferences about students' motivational orientations (e.g., approach vs. avoidance) or their meta-cognitive states, evidenced via the use of self-regulated learning strategies (e.g., re-reading vs. self-explanation). In general, EDM approaches to model affect, motivation, and metacognition follow two main trajectories: (a) use of EDM methods to computationally model affect, motivation, and metacognition and (b) embedding these models into learning environments to afford dynamic intervention (e.g., re-engaging bored students).

EDM researchers typically use supervised learning to model the multi-componential, multi-temporal, and dynamic nature of affect, motivation, and metacognition during learning. Supervised learning attempts to solve the following problem. Given some *features* (e.g., facial features, eye blinks, interaction patterns) with corresponding *labels* (e.g., confused or frustrated) initially provided by humans, can we learn to *model* the relationship between the features and the labels so as to automatically provide the correct labels on future unseen data (i.e., features from new students without corresponding labels)? In other words,after a training period, the computer now automatically provides assessments of the students' mental states. Supervised learning methods have been used for modeling of students' affective, motivational, and metacognitive states as elaborated in the examples below.

Affect modeling aims to study and computationally model affective states that are known to occur during learning and that indirectly influence learning by modulating cognitive processes. For example, Pardos, Baker, San Pedro, and Gowda[91] developed an automated affect model for the ASSISTments mathematics ITS.[92] Their model discriminated among various affective states (confusion, boredom, etc.) based on the digital trace of students' interactions stored in log files (e.g., use of hints, performance on problems). Their model successfully predicted performance on a state standardized test[91] as well as college enrollment several years later,[93] thereby demonstrating the utility of EDM-based affect models to predict important academic outcomes. In addition to an offline analysis of existing data, researchers have also explored real-time models of affect. One example is Affective AutoTutor, a dialog-based ITS for computer literacy that automatically models students' confusion, frustration, and boredom in real-time by monitoring facial expressions, body movements, and contextual cues.[35] The model is then used to dynamically adapt the tutorial dialog in a manner that is responsive to the sensed states (e.g., empathetic and motivational dialog moves when frustration is detected). A randomized controlled trial revealed learning benefits for students who interacted with the Affective AutoTutor compared to a non-affective version, but only for low-domain knowledge students.[94] Other EDM systems for affective modeling and dynamic intervention have been developed in the domains of Physics,[95] microbiology,[96] mathematics,[97,98] and many others – see review by D'Mello, Blanchard, Baker, Ocumpaugh, and Brawner.[99]

Meaningful learning requires motivation or the intention to learn. Demotivated students will likely disengage from a learning task entirely or engage in shallow levels of processing like skimming or re-reading. On the other hand, motivated students are expected to engage at deeper levels of processing, such as inference generation and effortful deliberation. Of course, contemporary theories of motivation go beyond simply differentiating between motivated and demotivated students by considering different aspects of student motivation orientations (e.g., mastery vs. performance and approach vs. avoidance).[88,100] However, these theories largely conceptualize motivation as stable traits rather than as dynamic processes that unfold over a learning session.[101] Taking a somewhat different approach, researchers have been applying EDM techniques to model the dynamic nature of motivational processes and have also developed interventions to re-engage demotivated students.[102] For example, M-Ecolab or Motivation Ecolab is a motivational version of Ecolab-II, a system that teaches concepts pertaining to food chains and food webs to 5th grade students.[103] M-Ecolab models students' motivation based on how they interact with the system (e.g., use of help resources, answer correctness) and intervenes by providing motivational feedback and cognitive scaffolding. A preliminary comparison of M-Ecolab to its non-motivationally supportive counterpart yielded increased motivation but no measured learning improvements.[104] Other pertinent studies have utilized EDM techniques to model motivation-related constructs, such as self-efficacy[105] and disengagement.[106] However, when compared to modeling of affect, motivation modeling is still in its infancy, so there are likely to be more advances in the next few years.

Both affect and cognition are under the watchful eye of metacognitive processes (thinking about thinking).[87] According to Dunlosy, Serra and Baker[107] monitoring-control framework, students continually monitor multiple aspects of the learning process (e.g., ease-of-learning, feelings-of-knowing, quality of information-sources) and use this information to adjust or control their learning activities (e.g., deciding what to study, when to study, how long to study). These self-regulatory behaviors provide critical insight into the metacognitive processes that underlie learning. Students might engage in productive strategies like self-explaining[108] or self-correcting errors.[109] They might also use less beneficial activities like failing to utilize help utilities despite making repeated errors or abusing these utilities to get quick answers.[110] In line with this, researchers have been applying EDM techniques to model self-regulatory behaviors in order to encourage more productive strategies while simultaneously discouraging less productive ones.[111] For example, Baker et al.[112] applied supervised classification methods to detect and respond to instances of "gaming the system" – succeeding by attempting to exploit systematic properties of the system (abusing hints and other help resources). Roll, Aleven, McLaren and Koedinger[18] applied EDM techniques to analyze "help seeking skills" and to leverage these insights to develop adaptive tutorial interventions to help students improve these skills. Finally, MetaTutor[113] is an ITS specifically designed to model and scaffold students' use of effective self-regulated learning strategies and relies heavily on EDM methods to achieve this goal. An important research question emerging from this overall line of research concerns the extent to which machine- and student- estimates of self-regulatory strategies align and if this information can be used to correct misconceptions and biases in students' perceptions of self-regulatory strategy use.[114]

EDM projects in this area contribute to cognitive science by producing insights into the interplay between student affect and cognitive states. For example, the use of sequential data mining techniques to study how affective states arise and influence student behaviors while solving analytically reasoning problems (e.g.,[115]) adds an affective perspective to theories of problem solving. Similarly, EDM approaches to studying self-regulation (motivation and metacognitive) are guided in contemporary learning theories, so patterns discovered can be used to systematically test these theories and revise them as needed.

Another set of insights comes from work on the interplay between student affect and cognitive states. Educational data mining techniques have revealed how affective states arise and influence student behaviors during complex problem solving. For example, confusion, which is often perceived as being a negative affective state, can positively impact learning if accompanied by effortful cognitive activities directed towards confusion resolution.[116]

To summarize, EDM techniques have had a profound influence on computationally modeling the complex, elusive, and evasive constructs of affect, motivation, and metacognition. They allow the researcher to go beyond the norm of simply capturing a few static snapshots of these processes with self-report questionnaires. Instead, they afford the construction and use of rich behavior-informed dynamic models that are inherently coupled within the learning context. The diversity and richness of research can be appreciated by consulting recent reviews and edited compilations in this area.[87,99,117-119]

2.4 Language data and collaborative learning support

Language technologies and machine learning have been applied to analysis of verbal data in education relevant fields for nearly half a century. Major successes in this area enable educational applications such as automated essay scoring,[13] tutorial dialogue,[3,120,121] computer supported collaborative learning environments with context sensitive dynamic support,[33,122,123,] and, most recently, prediction of student likelihood to drop out in Massive Open Online Courses (MOOCs).[27,28] Across all of these application areas, a key consideration is the manner in which raw text or speech is transformed into a set of features that can be processed using machine learning. Researchers have explored a wide range of approaches. The most simplistic are "bag of word" approaches that represent texts as the set of words that appear at least once in the text.[124] The most complex approaches extract features from full linguistic structural analyses.[3] A consistent finding is that representations motivated by theoretical frameworks from linguistics and psychology show particular.[125,28,126,3] The impact on student learning is achieved through the ability to detect and adapt to individual student.[127,122, 33]

One of the earliest applications was automated essay scoring,[128] which has recently experienced a resurgence of interest.[13] The earliest approaches used simple models, like regression, and simple features, such as counting average sentence length, number of long words, and length of essay. These approaches were highly successful in terms of reliability of assignment of numeric scores (e.g.,[129]), however they were criticized for lack of validity in their usage of evidence for assessment. Later approaches used techniques akin to factor analysis such as latent semantic analysis[130] or Latent Dirichlet Allocation[131] to incorporate approximations of content based assessments. Other keyword based language analysis approaches such as CohMetrix[132] have been used for assessment of student writing along multiple dimensions, including such factors as cognitive complexity, narrativity, and cohesion. In highly causal domains, approaches that build in some level of syntactic structural analysis have shown benefits.[3] In science education, success with assessment of open ended responses has been achieved with LightSIDE,[133,36] a freely available suite of software tools supporting use of text mining technology by non-experts. Applications such as Why2-Atlas[121] and Summary Street[37] use these automated assessments to offer detailed feedback to students on their writing. An in depth discussion of reliability and validity trade-offs between alternative approaches to automated text analysis has been published in prior work.[24]

In the past decade, applications of machine learning have been applied to the problem of assessment of learning processes in discussion. This problem is referred to as automatic collaborative learning process analysis.[24] Automatic analysis of collaborative processes has value for real time assessment during collaborative learning, for dynamically triggering supportive interventions in the midst of collaborative learning sessions, and for facilitating efficient analysis of collaborative learning processes at a grand scale. This dynamic approach has been demonstrated to be more effective than an otherwise equivalent static approach to support.[122] Early work in automated collaborative learning process analysis focused on text-based interactions and click stream data.[24,134-137] Early work towards analysis of collaborative processes from speech has begun to emerge as.[126,138]

One aspect of collaborative discussion processes that has been a focus in this area of research is Transactivity.[139] This research is a good example of how insights about the relevant processes from a psychological perspective informs computational work. The concept of Transactivity originally grows out of a Piagetian theory of learning where this conversational behavior is said to reflect a balance of perceived power within an interaction.[140] Transactive contributions make reasoning public and relate expressions of reasoning to previously contributed instances of reasoning in the conversation. It is argued to reflect engagement in interactions that may provide opportunities for cognitive conflict. Research in the area of sociolinguistics suggests that the underlying power relations might be detectable through analysis of shifts in speech style within a conversation, where the speakers shift to speaking more similarly to one another over time. This convergence phenomenon is known as speech style accommodation. It could be expected, then, that linguistic accommodation would predict the occurrence of Transactivity, and therefore a representation for language that represents evidence of such language usage shifts should be useful for predicting occurrence of Transactivity. This hypothesis has been confirmed through empirical investigation.[126] Consistent with this work, it has also been demonstrated that in a variety of efforts to automatically identify Transactive conversational contributions in conversational data, the more successful ones were those in which one or more features were include that represents similarity (in terms of word usage or topic coverage) between the language contributed by different speakers within a.[24,141]

EDM projects in this area contribute to cognitive science models of student learning by providing insights into how social processes can impact cognitive processes, in part by manipulating how safe or attractive opportunities for cognitive engagement appear to students. For example, Howley, Mayfield and Rosé[142] found in a secondary analysis of an earlier study[143] that expressed aggression within collaborative groups resulted in students who were the targets of aggression engaging in less functional help seeking patterns. This less functional help seeking pattern was associated with significantly less learning.

Most recently, machine learning has been used in a MOOC context to detect properties of student participation that might signal likelihood of dropping out of the course. The goal is to identify students who are in particular need of support so that the limited instructor time can be invested where it is most needed. This form of automated assessment has been used to detect expressed motivation and cognitive engagement,[28] student attitudes towards course affordances and tools,[27] relationship formation and relationship loss.[26] It has also been used to detect emergent subcommunities in discussion forums that differ with respect to content focus.[144,145] In each case, the validity of these measures has been assessed by measuring the extent to which differential measurements predict attrition over time in the associated courses. Simplistic applications of sentiment analysis make significant predictions about dropout in some MOOCs,[27] however the pattern is not consistent across MOOCs. A careful qualitative analysis demonstrates that coarse grained approaches to sentiment analysis pick up on different kinds of signals depending on the content focus of the course, and thus interpretation of such patterns must be treated with care.

With respect to student motivation and cognitive engagement, approximations have been made either using unsupervised probabilistic graphical modeling techniques[144] or using supervised learning over carefully hand labeled data rated on a likert scale from highly motivated to highly unmotivated, and using linguistically inspired features related to cognitive engagement.[28] In both cases, the story is the same, namely, dips in detected motivation predict higher likelihood of dropout at the next time point. However, a more accurate prediction results from the supervised method using linguistically motivated features.

2.5 Other areas of educational data mining

Many other questions have been explored with educational data mining. We give brief examples of some of these areas. Some researchers have demonstrated promise for using learning data to identify instructional policies or pedagogical tactics predicted to optimize learning. One ideal source for such analysis is from data where some instructional choice is randomized. For example, Chi, VanLehn, Litman, and Jordan[146] had students use a Physics intelligent tutoring system where for each solution step in a problem given to a student, the system would randomly either show the student an example of how to solve that step or ask the student to perform the step on his or her own. They used this data to train Markov Decision Process (MDP) policies and in a follow-up experiment, they showed that students learned more from an MDP policy trained to maximize student learning gains than from one trained to minimize student learning gains. Other researchers have begun to extend these efforts using Partially Observable MDP (POMDP) planning (e.g.,[147-149]).

Some researchers have demonstrated how models trained on educational technology interaction data can predict long-term standardized test results (e.g.,[150]) or college enrollment (e.g.,[93]). Others have built models to predict when students will get stuck in a course[151] or drop out (e.g.,[152]). Other areas of interest include employing recommender systems in education (e.g.,[153]), social network analysis (e.g.,[154]), and peer grading.[9,12,11]

3. Current challenges & future opportunities

3.1 Future work: assessment, growth modeling of learning, and prediction of achievement

Table 2 provides a jumping off point for suggestions for future research in educational data mining. While some work has approached the question of how best to model student performance when multiple skills or strategies (KCs) are needed or possible (rows 6 and 9), more work is needed in this area. In particular, researchers should explore consequences of adding learning parameters to multi-skill assessment approaches (cf.,[155]).

Missing combinations of features in Table 2 suggests new research. For example, the only logistic regression family model that has performance parameters is IRT 3PL, however that model does not have learning parameters. Creating a logistic regression model that has both learning and performance parameters is quite feasible but, to our knowledge, has not been done. It is also an open issue whether other kinds of parameters are potentially productive in modeling. For example, combining multiple parameters per student, as in multidimensional IRT, along with learning parameters has not been sufficiently explored.

Models based on Bayesian Networks take advantage of strong statistical inference techniques and abundant computational resources in order to maximize fit to the data while striving to converge to pedagogically informative parameters maximize fit to the data. However, the error gradients of these models, which guide the learned parameter values towards convergence, can be complex and non-convex leading to local optima that may describe the data with high accuracy but have an explanation that is not educationally plausible.[156] This problem of model identifiability is especially pronounced in canonical model forms, such as the Hidden Markov Model that Bayesian Knowledge Tracing is based on. This necessitates parameter constraining heuristics such as bounding to plausible regions or biasing the starting parameter values. More complex models that take into account information outside of KC and response sequence have demonstrated improved predictive accuracy[157,158] as have models accounting for individual learning differences[57,5,51,159] and difference in learning by type of help seen in an ITS[160,161] and in a MOOC.[162] Still, some argue for parsimony opting to instead simplify models down to a form where they may be analytically solved.[163] Striking a balance between complexity, interpretability, and validity remains a challenge in mainstreaming model-based discovery and formative assessment.

3.2 Future work: Transfer and cognitive model discovery

Perhaps the most important need for future work in the area of cognitive model discovery is for further examples of close-the-loop experiments that demonstrate how cognitive models with greater predictive accuracy can be used to improve student learning (cf.,[7,40]).

Empirical comparison of cognitive models of transfer of learning has been done mostly using the Additive Factors Model (AFM) as the statistical model. It is unknown whether the results of such comparisons would change if other statistical models were used (e.g., Bayesian Knowledge Tracing). While substantial differences (e.g., cognitive model A is much better than B with AFM but B is much better than A for BKT) seem unlikely, the issue deserves attention especially given that BKT is more widely used in fielded intelligent tutoring systems.

Given that most of the datasets that have been used for cognitive model discovery in particular, and EDM in general, are naturally occurring data sets, there are related open questions of interest and importance. For example, estimations of learning and item difficulty can be confounded by sampling issues. A lack of randomization of items in a curriculum might be responsible for performance variance due to learning being attributed to item parameters (cf.,[59]). For example, an item that is consistently completed at the end of a unit is more likely to benefit from transfer of learning from earlier items than an item that is randomly distributed throughout the unit. Here the future research issue may be less about the statistical modeling framework and more about identifying or creating data sets that include more random variation of task ordering.

3.3 Future Work: Scope and scale of models of affect, motivation, and metacognition

We discussed how EDM techniques have become essential tools for researchers interested in modeling affect, motivation, and metacognition in digital learning contexts. The field has made much progress relative to its infancy, yet there is much more to be done. One open area pertains to the constructs themselves. Researchers have focused on individual components of each construct, despite the constructs being multi-componential in nature. For example, engagement is a complex meta-construct with affective, cognitive, and behavioral components,[164] yet researchers mainly model its individual sub-components. At a broader level, complex learning inherently involves an interplay between cognition, affect, motivation, and metacognition, so it is important to consider unified models that capture the interdependencies and unfolding dynamics of these processes. For example, a learner with performance-avoidance orientations (motivation) might experience anxiety (affect) due to fear of failure, which leads to rumination on the consequences of failure (metacognition), thereby consuming working memory resources (cognition), ultimately resulting in failure. Models capable of capturing multiple links in this chain of events, without being overly diffuse, would represent significant progress. Another research goal involves devising models that scale. While, cognitive models have enjoyed widespread implementation in intelligent tutoring systems used by tens of thousands of students each day, the same cannot be said for models of affect, motivation, and metacognition. The challenge of scaling up has been difficult due to the use of supervised machine learning techniques, which require labeled data to infer relationships between observable behaviors and latent mental states. Labeled data can be collected in small-scale research studies, where log-files, videos, and other artifacts can be meticulously annotated, but this approach does not scale to the hundreds of thousands of students who learn from Massive Open Online Courses (MOOCs) and other online resources each day. Alternate modeling techniques, such as semi-supervised or unsupervised approaches might be needed to resolve the scalability issue. Thus, expanding the scope and scale of the models reflect two of the several potential avenues for research.

3.4 Future Work: Language data and collaborative learning support

As more and more emphasis is placed on scaling up work on computer supported instruction, we turn to application of the frameworks and technologies developed so far in contexts such as Massive Open Online Courses. In this context, social interaction occurs within online environments such as threaded discussion, twitter, blog sites, Facebook study groups, and small group breakout discussions, sometimes in chat, and sometimes in computer mediated video. Some work has already successfully produced automated analyses of discussion forum data in MOOCs.[165,145,26-29] This recent work has interesting similarities and differences with past research on student dialogues. For example, discussions in the threaded forums in MOOCs is much less focused and task oriented than the typical chat logs from computer supported collaborative learning activities. While some work applying automated collaborative learning process analysis to audio data has been done,[126] this is still largely an open problem, and even less work has been done so far on automated analysis of video. While some work on threaded discussion data has been done,[166,167] much more work in the computer supported collaborative learning literature has focused on analysis of chat.[168,169] The language phenomena that have been studied in a chat context will look different in other modes of communication, and therefore work must be done to adapt approaches and frameworks developed for one mode of communication to others. MOOCs also bring with them the opportunity to apply the analysis to different problems. For example, addressing the problem of attrition was not a major focus of work in computer supported collaborative learning because it was not an issue in those environments, but is an extremely central concern in the context of MOOCs. Thus, the shift in contexts opens new opportunities for impact going forward.

4. Conclusion

We set out to describe the exciting and rapidly growing area of educational data mining. It is an area of interest as it touches upon basic research questions of how students learn and how learning can be modeled in a manner that is relevant to multiple disciplines within and beyond cognitive science. It is important as it contributes to the development of better human and technical support for more effective, efficient, and rewarding student learning. Educational data mining will play a central role in the anticipated "two revolutions in learning",[170] the boom in affordable and accessible online courses and the increased attention on learning science.

References

1. Litman, D. & Silliman, S. (2004). ITSPOKE: An Intelligent Tutoring Spoken Dialogue System.*Companion Proceedings of the Human Language Technology Conference: 4th Meeting of the North American Chapter of the Association for Computational Linguistics (HLT/NAACL)*, Boston, MA.

2. Johnson, W. L. & Valente, A. (Summer 2009). Tactical Language and Culture Training Systems: Using AI to Teach Foreign Languages and Cultures. *AI Magazine*, 30 (2), 72-83.

3. Rosé C. P., & VanLehn, K. (2005). An Evaluation of a Hybrid Language Understanding Approach for Robust Selection of Tutoring Goals, *International Journal of AI in Education 15(4).*

4. Corbett, A.T., Anderson, J.R. (1995). Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modeling and User-Adapted Interaction*, 4, 253-278.

5. Pardos, Z. A., & Heffernan, N. T. (2010a). Modeling individualization in a bayesian networks implementation of knowledge tracing. *In User Modeling, Adaptation, and Personalization*, (pp. 255-266). Springer Berlin Heidelberg.

6. Pardos, Z. A., Gowda, S. M., Baker, R. S.J.D., & Heffernan, N. T. (2012a). The Sum is Greater than the Parts: Ensembling Models of Student Knowledge in Educational Software. *ACM SIGKDD Explorations*, 13(2).

7. Koedinger, K. R., McLaughlin, E. A., & Stamper, J. C. (2012). Automated cognitive model improvement. Yacef, K., Zaïane, O., Hershkovitz, H., Yudelson, M., and Stamper, J. (eds.) *Proceedings of the 5th International Conference on Educational Data Mining*, pp. 17-24. Chania, Greece. [Best Paper Award]

8. Martin, B., Mitrovic, T., Mathan, S., & Koedinger, K.R. (2011). Evaluating and improving adaptive educational systems with learning curves. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research (UMUAI), 21*(3), 249-283. [2011 James Chen Annual Award for Best UMUAI Paper]

9. Piech, C., Sahami, M., Koller, D., Cooper, S. & Blikstein, P. (2012). Modeling How Students Learn To Program. *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education*, Raleigh, USA. 2012.

10. Rivers, K & Koedinger, K. R. (in press). Automating Hint Generation with Solution Space Path Construction. To appear in *Proceedings of the International Conference on Intelligent Tutoring Systems*, Lecture Notes in Computer Science.

11. Patchan, M. M., Hawk, B. H., Stevens, C. A., & Schunn, C. D. (2013). The effects of skill diversity on commenting and revisions. *Instructional Science, 41*(2), 381-405

12. Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., & Koller, D. (2013). Tuned Models of Peer Assessment in MOOCs. In *Proceedings of the 6th International Conference on Educational Data Mining, Memphis, Tennessee. 2013.*

13. Shermis, M. & Hammer, B. (2012). Contrasting state-of-the-art automated scoring of essays: Analysis, Annual National Council in Measurement in Education Meeting, pp 14-16.

14. Evens, M., Brandle, S., Chang, R., Freedman, R., Glass, M., Lee, Y., Shim, L., Woo, C., Zhang, Y., Zhou, Y., Michael, J. & Rovick, A. (2001). CIRCSIM-Tutor: An Intelligent Tutoring System Using Natural Language Dialogue. Proc. MAICS 2001, Oxford, OH, 16-23.

15. Chi, M., Jordan, P. & VanLehn, K.(in press). When is tutorial dialogue more effective than step-based tutoring? Intelligent Tutoring Systems (ITS 2014).

16. Kumar, R. & Rosé, C. P. (in press). Triggering Effective Social Support for Online Groups. ACM Transactions on Interactive Intelligent Systems.

17. Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. International Journal of Artificial Intelligence in Education, 16, 101-128.

18. Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2007). Designing for metacognition—applying cognitive tutor principles to the tutoring of help seeking. Metacognition and Learning, 2(2-3), 125-140.

19. Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. Educational Psychologist, 40(4), 199209.

20. 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.

21. Beck, J. (2005). Engagement tracing: using response times to model student disengagement. In Proceedings of the 12th International Conference on Artificial Intelligence in Education (AIED 2005), pp. 88–95.

22. Ocumpaugh, J., Baker, R.S.J.d., Gaudino, S., Labrum, M.J., & Dezendorf, T. (in press). Field Observations of Engagement in Reasoning Mind. To appear in Proceedings of the 16th International Conference on Artificial Intelligence and Education. Page, E. B. (1966). The imminence of grading essays by computer. Phi Delta Kappan, 48, 238–243.

23. D'Mello, S. K., Craig, S. D., Gholson, B., Franklin, S., Picard, R.,& Graesser, A. C.(2005). Integrating affect sensors in an intelligent tutoring system. In Affective Interactions: The Computer in the Affective Loop Workshop at 2005 International conference on Intelligent User Interfaces (pp. 7-13) New York: AMC Press.

24. Rosé, C. P., Wang, Y.C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F., (2008). Analyzing Collaborative Learning Processes Automatically: Exploiting the Advances of Computational Linguistics in Computer-Supported Collaborative Learning. International Journal of Computer Supported Collaborative Learning 3(3), 237-271.

25. Yang, D., Sinha, T., Adamson, D., & Rosé, C. P. (2013). Turn on, Tune in, Drop out: Anticipating student dropouts in Massive Open Online Courses, *NIPS Data-Driven Education Workshop.*

26. Yang, D., Wen, M., Rosé, C. P. (2014). Peer Influence on Attrition in Massively Open Online Courses, *Proceedings of Educational Data Mining*

27. Wen, M., Yang, D., & Rosé, C. P. (2014a). Sentiment Analysis in MOOC Discussion Forums: What does it tell us? *Proceedings of Educational Data Mining.*

28. Wen, M., Yang, D., Rosé, D. (2014b). Linguistic Reflections of Student Engagement in Massive OpenOnline Courses, in *Proceedings of the International Conference on Weblogs and Social Media.*

29. Nathan, M. J. & Alibali, M. W. (2010). Learning Sciences. Wiley Interdisciplinary Reviews – *Cognitive Science, 1 (May/June), 329-345.*

30. Newell, A. (1990). Unified theories of cognition. Cambridge, MA: Harvard University Press.

31. 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.

32. Baker, R.S. (2005). Designing Intelligent Tutors That Adapt to When Students Game the System.Doctoral Dissertation. Carnegie Mellon University.

33. Adamson, D., Dyke, G., Jang, H. J., Rosé, C. P. (2014). Towards an Agile Approach to Adapting Dynamic Collaboration Support to Student Needs, *International Journal of AI in Education* 24(1), pp91-121.

34. Dyke, G., Adamson, A., Howley, I., & Rosé, C. P. (2013). Enhancing Scientific Reasoning and Discussion with Conversational Agents, *IEEE Transactions on Learning Technologies* 6(3), special issue on Science Teaching, pp 240-247.

35. D'Mello, S., Jackson, G., Craig, S., Morgan, B., Chipman, P., White, H., Person, N., Kort, B., El Kaliouby, R., Picard, R., & Graesser, A. (2008). AutoTutor detects and responds to learners affective and cognitive states. Paper presented at the *Proceedings of the Workshop on Emotional and Cognitive Issues in ITS held in conjunction with the Ninth International Conference on Intelligent Tutoring Systems*, Montreal, Canada.

36. Mayfield, E. & Rosé, C. P. (2013). LightSIDE: Open Source Machine Learning for Text Accessible to Non-Experts, Invited chapter in the *Handbook of Automated Essay Grading,* Routledge Academic Press.

37. Wade-Stein, D. & Kintsch, E. (2004). Summary Street: Interactive computer support for writing, *Cognition and Instruction* 22(3), pp 333-362.

38. Yang, D., Adamson, D., Rosé, C. P. (under review-b). Question Recommendation with Constraints for Massive Open Online Courses, submitted to the Recommender Systems

Conference.

39. Koedinger, K.R. & McLaughlin, E.A. (2010). Seeing language learning inside the math: Cognitive analysis yields transfer. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. (pp. 471-476.) Austin, TX: Cognitive Science Society.

40. Koedinger, K. R., Stamper, J. C., McLaughlin, E. A., & Nixon, T. (2013). Using data-driven discovery of better student models to improve student learning. *The 16th International Conference on Artificial Intelligence in Education (AIED2013).*

41. Koedinger, K.R., Corbett, A.C., & Perfetti, C. (2012). The Knowledge-Learning-Instruction (KLI) framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36* (5).

42. Yu, H-F., Lo, H-Y., Hsieh, H-P., Lou, J-K., Mckenzie, T.G., Chou, J-W., et al., (2010). Feature Engineering and Classifier Ensemble. *Proc. of the KDD Cup 2010 Workshop, 1--16.*

43. Toscher, A., Jahrer, M. (2010). Collaborative Filtering Applied to Educational Data Mining. *Proc. of the KDD Cup 2010 Workshop,* 17--28.

44. Nguyen, T-N, Drumond,L., Horváth, T., Schmidt-Thieme, L. (2011). Multi-Relational Factorization Models for Predicting Student Performance. *Proceedings of the KDD 2011 Workshop on Knowledge Discovery in Educational Data* (KDDinED 2011). Held as part of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.

45. Pardos, Z. A., & Heffernan, N.T. (2010b). Using HMMs and bagged decision trees to leverage rich features of user and skill. *Proc. of the KDD Cup 2010 Workshop*, pp. 28---39.

46. Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

47. Atkinson, R. C. & Paulson, J. A. (1972). An approach to the psychology of instruction. *Psychological Bulletin*, 78, 49-61.

48. Anderson, J. R. (1993). *Rules of the Mind*. Hillsdale, NJ: Erlbaum.

49. Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, 14 (2):249-255.

50. Lee, J. I. & Brunskill, E. (2012). The Impact on Individualizing Student Models on Necessary Practice Opportunities. *In Proceedings of the 5th International Conference on Educational Data Mining*, 118-125. Chania, Greece.

51. Yudelson, M., Koedinger, K.R., Gordon, G. (2013) Individualized Bayesian Knowledge Tracing Models. In: Lane, H.C., and Yacef, K., Mostow, J., Pavlik, P.I. (eds.) *Proceedings of 16th International Conference on Artificial Intelligence in Education (AIED 2013),* Memphis, TN. LNCS vol. 7926, (pp. 171-180). Springer-Verlag Berlin Heidelberg (2013).

52. Wilson, M., & De Boeck, P. (2004). Descriptive and explanatory item response models. In P. De Boeck, & M. Wilson, (Eds.) *Explanatory item response models: A generalized linear andnonlinear approach*. New York: Springer-Verlag.

53. Spada, H., & McGaw, B. (1985). The assessment of learning effects with linear logistic test models. In S.E. Embretson (Ed.), *Test design: Developments in Psychology and Psychometrics* (pp. 169-193). New York: Academic Press.

54. Draney, K., Wilson, M., & Pirolli, P. (1996). Measuring learning in LISP: An application of the random coefficients multinomial logit model. In, G. Engelhard & M. Wilson (Eds.), *Objective measurement III: Theory into practice*. Norwood, NJ: Ablex.

55. Cen, H., Koedinger, K. R., & Junker, B. (2006). Learning Factors Analysis: A general method for cognitive model evaluation and improvement. In M. Ikeda, K.D. Ashley, T.- W. Chan (Eds.) *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, pp. 164-175. Berlin: Springer-Verlag.

56. Pavlik, P. I., Cen, H., & Koedinger, K. R. (2009). Performance factors analysis: a new alternative to knowledge tracing. In *Proceeding of the 2009 conference on Artificial Intelligence in Education*. IOS Press, 531-538.

57. Rafferty, A. N. & Yudelson, M (2007). Applying learning factors analysis to build stereotypic student models. In Luckin, R., Koedinger, K. R. & Greer, J. (Eds.). *Proceedings of 13th International Conference on Artificial Intelligence in Education (AIED2007*), 697-698. Amsterdam, IOS Press.

58. Chi, M., Koedinger, K.R., Gordon, G., Jordan, P., & VanLehn, K. (2011). Instructional Factors Analysis: A cognitive model for multiple instructional interventions. In *Proceedings of the Fourth International Conference on Educational Data Mining*, pp. 61-70.

59. Koedinger, K.R., Yudelson, M., & Pavlik, P. (submitted). Testing theories of transfer using error rate learning curves.

60. Cen, H. (2009). *Generalized Learning Factors Analysis: Improving Cognitive Models with MachineLearning*. (Doctoral dissertation).

61. Cen, H., Koedinger, K. R., Junker, B. (2008). Comparing two IRT models for conjunctive skills. In B.Woolf et al. (Eds): ITS 2008, *Proceedings of the 9th International Conference of Intelligent Tutoring Systems, pp. 796-798.* Springer-Verlag Berlin Heidelberg.

62. Clark, R. E. (2014). Cognitive Task Analysis for Expert-Based Instruction in Healthcare. In J. Michael Spector, J. M. Merrill, M. D. Elen, J. and Bishop, M. J. (eds.). *Handbook of Research on Educational Communications and Technology,* 4th Edition, pp. 541-551. Springer: New York.

63. Lee, R. L. (2003). Cognitive task analysis: A meta-analysis of comparative studies. Unpublished doctoral dissertation, University of Southern California, Los Angeles.

64. Velmahos, G. C., Toutouzas, K. G., Sillin, L. F., Chan, L., Clark, R. E., Theodorou, D., & Maupin, F. (2004). Cognitive task analysis for teaching technical skills in an inanimate surgical skills laboratory. *The American Journal of Surgery*, *18*, 114-119.

65. Schaafstal, A., Schraagen, J. M., & Van Berlo, M. (2000). Cognitive task analysis and innovation of training: The case of structured troubleshoot-ing. *Human Factors*, 42,75-86.

66. Merrill, M. D. (2002). A pebble-in-the-pond model for instructional design. *Performance Improvement*,41(7), 39–44.

67. Seamster, T.L., Redding, R.E., Cannon, J.R., Ryder, J.M. & Purcell, J.A. (1993). Cognitive task analysis of expertise in air traffic control. *The International Journal of Aviation Psychology*, 3, 257-283.

68. Koedinger, K.R., & MacLaren, B.A. (2002). Developing a pedagogical domain theory of early algebra problem solving. CMU-HCII Tech Report 02-100.

69. Polk, T. A., VanLehn, K. & Kalp, D. (1995). *ASPM2: Progress toward the analysis of symbolic parameter models*.

70. Ritter, F. E. (1992). A methodology and software environment for testing process models' sequential predictions with protocols. Doctoral Dissertation, Department of Psychology, Carnegie Mellon University.

71. Ritter, F. E., & Larkin, J. H. (1994). Developing process models as summaries of HCI action sequences. *Human-Computer Interaction*, 9, 345-383.

72. Salvucci, D. D. (1999). Mapping eye movements to cognitive processes (Tech. Rep. No. CMU-CS-99-131). Doctoral Dissertation, Department of Computer Science, Carnegie Mellon University. (A thesis summary and the full thesis are available at http://www.cs.cmu.edu/~dario/TH99/

73. Heffernan, N. & Koedinger, K. R. (1998). A developmental model for algebra symbolization: The results of a difficulty factors assessment. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, (pp. 484-489). Hillsdale, NJ: Erlbaum.

74. Heffernan, N. T. (2001). Intelligent Tutoring Systems have Forgotten the Tutor: Adding a Cognitive Model of Human Tutors. Doctoral Dissertation. Advisors: Kenneth R. Koedinger, John R. Anderson.

75. Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26(2).*

76. Baker R.S., Corbett A.T., Koedinger K.R., Schneider, M.P. (2003). A Formative Evaluation of a Tutor for Scatterplot Generation: Evidence on Difficulty Factors. *Proceedings of the Conference on Artificial Intelligence in Education*, 107-115.

77. Koedinger, K. R. & Nathan, M. J. (2004). The real story behind story problems: Effects of

representations on quantitative reasoning. *The Journal of the Learning Sciences, 13* (2), 129-164.

78. 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,* (pp. 61-78). Cambridge University Press.

79. Stamper, J. & Koedinger, K.R. (2011). Human-machine student model discovery and improvement using data. In J. Kay, S. Bull & G. Biswas (Eds.), *Proceedings of the 15th International Conference on Artificial Intelligence in Education*, pp. 353-360. Berlin: Springer.

80. Tatsuoka, K.K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement,* 20, 345-354.

81. Villano, M. (1992). Probabilistic student models: Bayesian belief networks and knowledge space theory. *Proceedings of the Second International Conference on Intelligent Tutoring Systems*. New York: Springer-Verlag, Lecture Notes in Computer Science.

82. Desmarais, M. C. (2011). Mapping question items to skills with non-negative matrix factorization. *ACM KDD- Explorations*, 13(2), pp. 30-36.

83. Desmarais, M. C. and Naceur, R. (2013). A Matrix Factorization Method for Mapping Items to Skills and for Enhancing Expert-Based Q-matrices. *In Proceedings of the 16th Conference on Artificial Intelligence in Education (AIED2013)*, pp. 441-450. Memphis, TN.

84. Barnes, T. (2005). The q-matrix method: Mining student response data for knowledge. *Proceedings of the AAAI-2005 Workshop on Educational Data Mining*, July 9-13, 2005, Pittsburgh, PA.

85. Li, N., Stampfer, E., Cohen, W., & Koedinger, K.R. (2013). General and efficient cognitive model discovery using a simulated student. In M. Knauff, N. Sebanz, M. Pauen, I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. (pp. 894-9) Austin, TX: Cognitive Science Society.

86. Liu, R., Koedinger, K. R., & McLaughlin, E. A. (2014). Interpreting Model Discovery and Testing Generalization to a New Dataset. *The 7th International Conference on Educational Data Mining (EDM 2014).*

87. Azevedo, R., & Aleven, V. (Eds.). (2013). *International handbook of metacognition and learning technologies* (Vol. New York): Springer.

88. Elliot, A., & McGregor, H. (2001). A 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*(3), 501-519. doi: 10.1037//0022-3514.80.3.501

89. Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed., pp. 153-189). Mahwah: Lawrence Erlbaum Associates.

90. Winne, P., & Baker, R. (2013). The potentials of educational data mining for researching

metacognition, motivation and self-regulated learning. *Journal of Educational Data Mining, 5*(1), 1-8.

91. Pardos, Z.A., Baker, R.S.J.d., San Pedro, M.O.C.Z., Gowda, S.M. (2014) Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End of Year Learning Outcomes. *Journal of Learning Analytics*, 1(1), 107–128.

92. Razzaq, L., Feng, M., Nuzzo-Jones, G., Heffernan, N. T., Koedinger, K. R., Junker, B., Ritter, S., Knight, A., Aniszczyk, C., & Choksey, S. (2005). The Assistment project: Blending assessment and assisting. In C. Loi & G. McCalla (Eds.), *Proceedings of the 12th International Conference on Artificial Intelligence In Education* (pp. 555-562). Amsterdam: IOS Press.

93. San Pedro, M., Baker, R. S., Bowers, A. J., & Heffernan, N. T. (2013). Predicting College Enrollment from Student Interaction with an Intelligent Tutoring System in Middle School. In S. D'Mello, R. Calvo & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining (EDM 2013)*. 177-184: International Educational Data Mining Society.

94. D’Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., Perkins, L., & Graesser, A. (2010). A time for emoting: When affect-sensitivity is and isn’t effective at promoting deep learning. In J. Kay& V. Aleven (Eds.), *Proceedings of the 10th International Conference on Intelligent Tutoring Systems* (pp. 245-254). Berlin / Heidelberg: Springer.

95. Forbes-Riley, K., & Litman, D. J. (2011). Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor. *Speech Communication, 53*(9-10), 1115-1136. doi:10.1016/j.specom.2011.02.006

96. Sabourin, J., Mott, B., & Lester, J. (2011). Modeling learner affect with theoretically grounded dynamic bayesian networks. In S. D'Mello, A. Graesser, B. Schuller & J. Martin (Eds.), *Proceedings of the Fourth International Conference on Affective Computing and Intelligent Interaction* (pp. 286-295). Berlin Heidelberg: Springer-Verlag.

97. Arroyo, I., Woolf, B., Cooper, D., Burleson, W., Muldner, K., & Christopherson, R. (2009). Emotion sensors go to school. In V. Dimitrova, R. Mizoguchi, B. Du Boulay & A. Graesser (Eds.), *Proceedings of the 14th International Conference on Artificial Intelligence In Education* (pp. 17-24). Amsterdam: IOS Press.

98. Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction, 19*(3), 267-303.

99. D'Mello, S., Blanchard, N., Baker, R., Ocumpaugh, J., & Brawner, K. (in press). I Feel Your Pain: A Selective Review of Affect-Sensitive Instructional Strategies. In R. Sottilare, A. Graesser, X. Hu & B. Goldberg (Eds.), *Design Recommendations for Adaptive Intelligent Tutoring Systems: Adaptive Instructional Strategies (Volume 2)*: Army Research Laboratory.

100. Meyer, D., & Turner, J. (2006). Re-conceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review, 18*(4), 377-390. doi: 10.1007/s10648-006-9032-1

101. Bernacki, M. L., Nokes-Malach, T. J., & Aleven, V. (2013). Fine-grained assessment of

motivation over long periods of learning with an intelligent tutoring system: Methodology, advantages, and preliminary results. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 629-644). New York: Springer.

102. du Boulay, B. (2011). Towards a motivationally-intelligent pedagogy: How should an intelligent tutor respond to the unmotivated or the demotivated? In R. Calvo & S. D'Mello (Eds.), *New Perspectives on Affect and Learning Technologies* (pp. 41-52). New York: Springer.

103. Luckin, R., & du Boulay, B. (1999). Ecolab: The development and evaluation of a Vygotskian design framework. *International Journal of Artificial Intelligence In Education, 10*(2), 198-220.

104. Rebolledo-Mendez, G., Du Boulay, B., & Luckin, R. (2006). Motivating the learner: an empirical evaluation. In M. Ikeda, K. Ashlay & C. T-W (Eds.), *Proceedings of the 8th International Conference on Intelligent Tutoring Systems* (pp. 545-554). Berlin: Springer.

105. McQuiggan, S., Mott, B., & Lester, J. (2008). Modeling self-efficacy in intelligent tutoring systems: An inductive approach. [Article]. *User Modeling and User-Adapted Interaction, 18*(1-2), 81-123. doi:10.1007/s11257-007-9040-y

106. Cocea, M., & Weibelzahl, S. (2009). Log file analysis for disengagement detection in e-Learning environments. *User Modeling and User-Adapted Interaction, 19*(4), 341-385. doi: 10.1007/s11257-009-9065-5

107. Dunlosky, J., Serra, M. J., & Baker, J. M. (2007). Metamemory applied. In F.T .Durso, R.S. Nickerson, S.T. Dumais, S. Lewandowsky & T. J. Perfect (Eds.), *Handbook of applied cognition* (Vol. 2). New York: Wiley.

108. Chi, M., Deleeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439-477.

109. Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction, 9*(4), 329-389.

110. Aleven, V., & Koedinger, K. R. (2000). Limitations of student control: Do students know when they need help? In G. Gauthier, C. Frasson & K. VanLehn (Eds.), *Proceedings of the 5th International Conference on Intelligent Tutoring Systems, ITS 2000* (pp. 292-303). Berlin: Springer.

111. Koedinger, K. R., Aleven, V., Roll, I., & Baker, R. (2009). In vivo experiments on whether supporting metacognition in intelligent tutoring systems yields robust learning. In D. J. Hacker, J. Dunlosky & A. Graesser (Eds.), *Handbook of metacognition in education* (pp. 897-964). New York: Routledge.

112. Baker, R., Corbett, A., Koedinger, K., Evenson, S., Roll, I., Wagner, A., Naim, M., Raspat, J., Baker, D., & Beck, J. (2006). *Adapting to when students game an intelligent tutoring system.* Paper presented at the Intelligent Tutoring Systems Conference.

113. Azevedo, R., Witherspoon, A., Graesser, A., McNamara, D., Rus, V., Cai, Z., Lintean, M., & Siler, E. (2008). MetaTutor: An adaptive hypermedia system for training and fostering self-regulated learning about complex science topics. In R. Pirrone, R. Azevedo & G. Biswas (Eds.), *Papers from the AAAI Fall Symposium on Cognitive and Metacognitive Educational Systems* (pp. 14-19). Menlo Park, CA: AAAI Press.

114. Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417-444.

115. D'Mello, S. K., Person, N., & Lehman, B. A. (2009). Antecedent-Consequent Relationships and Cyclical Patterns between Affective States and Problem Solving Outcomes. In V. Dimitrova, R. Mizoguchi, B. du Boulay, & A. Graesser (Eds.), *Proceedings of 14th International Conference on Artificial Intelligence In Education*.(pp. 57-64). Amsterdam: IOS Press.

116. D'Mello, S. K. & Graesser, A. C. (2014). Confusion. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), International handbook of emotions in education (pp. 289-310) : Routledge: New York, NY.

117. Calvo, R., & D'Mello, S. (2011). *New perspectives on affect and learning technologies*. New York: Springer.

118. Hacker, D., Dunlosky, J., & Graesser, A. (Eds.). (2009). *Handbook of metacognition in education*. New York: Routledge.

119. Porayska-Pomsta, K., Mavrikis, M., D'Mello, S. K., Conati, C., & Baker, R. (2013). Knowledge Elicitation Methods for Affect Modelling in Education. *International Journal of Artificial Intelligence in Education, 22*, 107-140.

120. Graesser, A., VanLehn, K., Rosé, C., Jordan, P., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue, *AI Magazine* 22(4), pp39-52.

121. VanLehn, K., Graesser, A., Jackson, G. T., Jordan, P., Olney, A., & Rosé, C. P., (2007). Natural Language Tutoring: A comparison of human tutors, computer tutors, and text. *Cognitive Science* 31(1), 3-52.

122. Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., Robinson, A. (2007). Tutorial Dialogue as Adaptive Collaborative Learning Support, *Proceedings of the 2007 conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work,* pp 383-390.

123. Kumar, R. & Rosé, C. P. (2011). Architecture for building Conversational Agents that support Collaborative Learning. *IEEE Transactions on Learning Technologies,* 4(1), pp 21-34.

124. Joachims, T. (2002). *Learning to Classify Text Using Support Vector Machines: Methods, Theory, and Algorithms,* Kluwer.

125. Rosé, C. P. & Tovares, A. (in press). What Sociolinguistics and Machine Learning Have to Say to One Another about Interaction Analysis. In Resnick, L., Asterhan, C., Clarke, S. (Eds.) *Socializing Intelligence Through Academic Talk and Dialogue,* Washington, DC: American Educational

Research Association.

126. Gweon, G., Jain, M., Mc Donough, J., Raj, B., Rosé, C. P. (2013). Measuring Prevalence of Other-Oriented Transactive Contributions Using an Automated Measure of Speech Style Accommodation, *International Journal of Computer Supported Collaborative Learning*, 8(2), pp245-265.

127. Rosé, C. P., Jordan, P., Ringenberg, M., Siler, S., VanLehn, K., Weinstein, A. (2001). Interactive Conceptual Tutoring in Atlas-Andes, *Proceedings of the 10th International Conference on AI in Education*, pp 256-266, San Antonio, Texas.

128. Page, E. B. (1966). The imminence of grading essays by computer. *Phi Delta Kappan*, 48, 238–243.

129. Shermis, M. D., & Burstein, J. (Eds.). (2013). *Handbook of automated essay evaluation: Current applications and new directions*. New York, NY: Routledge.

130. Deerwester, S., Dumais, S., Furnas, G., Landauer, T., Harshman, R. (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science* 4(6), pp 391-407.

131. Blei, D., Ng, A. and Jordan, M. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, (3), 993-1022.

132. McNamara, D. S., & Graesser, A. C. (in press). Coh-Metrix: An Automated Tool for Theoretical and Applied Natural Language Processing, in P. M. McCarthy & C. Boonthum (Eds.), *Applied natural language processing: Identification, investigation, and resolution.* Hershey, PA: IGI Global.

133. Nehm, R., Ha, M., & Mayfield, E. (2012). Transforming biology assessment with machine learning: automated scoring of written evolutionary explanations, *Journal of Science Education and Technology* 21(1), pp 183-196.

134. Soller, A., & Lesgold, A. (2000). Modeling the Process of Collaborative Learning. *Proceedings of the International Workshop on New Technologies in Collaborative Learning*. Japan: Awaiji–Yumebutai.

135. Erkens, G. & Janssen, J. (2008). Automatic Coding of Dialogue Acts in Collaboration Protocols, *International Journal of Computer Supported Collaborative Learning* (3), pp. 447-470.

136. McLaren, B., Scheuer, O., De Laat, M., Hever, R., de Groot, R. & Rosé, C. P. (2007). Using Machine Learning Techniques to Analyze and Support Mediation of Student E-Discussions, *Proceedings of the 2007 conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*, pp 331-338.

137. Mu, J., Stegmann, K., Mayfield, E., Rosé, C. P., Fischer, F. (2012). The ACODEA Framework: Developing Segmentation and Classification Schemes for Fully Automatic Analysis of Online Discussions. *International Journal of Computer Supported Collaborative Learning 7(2)*, 285-305.

138. Gweon, G., Agarwal, P., Udani, M., Raj., B., Rosé, C. P.(2011). The Automatic Assessment of Knowledge Integration Processes in Project Teams, in *Proceedings of the 9th International Computer Supported Collaborative Learning Conference, Volume 1: Long Papers*, pp 462-469.

139. Berkowitz, M., & Gibbs, J. (1983). Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly*, 29, pp. 399-410.

140. de Lisi, R. & Golbeck, S. L. (1999). Implications of Piagetian Theory for Peer Learning, in O'Donnell & King (Eds.) *Cognitive Perspectives on Peer Learning*, Lawrence Erlbaum Associates: New Jersey.

141. Ai, H., Sionti, M., Wang, Y. C., Rosé, C. P. (2010). Finding Transactive Contributions in Whole Group Classroom Discussions, in *Proceedings of the 9th International Conference of the Learning Sciences, Volume 1: Full Papers,* pp 976-983.

142. Howley, I., Mayfield, E. & Rosé, C. P. (2013). Linguistic Analysis Methods for Studying Small Groups, in Cindy Hmelo-Silver, Angela O'Donnell, Carol Chan, & Clark Chin (Eds.) *International Handbook of Collaborative Learning*, Taylor and Francis, Inc.

143. Cui, Y., Chaudhuri, S., Kumar, R., Gweon, G., Rosé, C. P. (2009). Helping Agents in VMT. In G. Stahl (Ed.) *Studying Virtual Math Teams*, Springer CSCL Series, Springer.

144. Yang, D., Wen, M., Kumar, A., Xing, E., Rosé, C. P. (under review). Towards an Integration of Text and Graph Clustering Methods as a Lens for Studying Social Interaction in MOOCs, submitted to *The International Review of Research in Open and Distance Learning.*

145. Rosé, C. P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P. & Sherer, J. (2014). Social Factors that Contribute to Attrition in MOOCs. *Proceedings of the First ACM Conference on Learning @ Scale (poster).*

146. Chi, M., VanLehn, K., Litman, D. & Jordan, P. (2011). An Evaluation of Pedagogical Tutorial Tactics for a Natural Language Tutoring System: A Reinforcement Learning Approach. International *Journal of Artificial Intelligence in Education, 21*: 83-113.

147. Brunskill, E. & Russell, S. (2010). RAPID: A Reachable Anytime Planner for Imprecisely-sensed Domains. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI).* Catalina Island, California.

148. Brunskill, E., Garg, S., Tseng, C., Pal, J. & Findalter, L. (2010). Evaluating an adaptive multi-user educational tool for low-resource regions. In *Proceedings of the International Conference on Information and Communication Technologies and Development (ICTD).* London, England.

149. Rafferty, A., Brunskill, E., Griffiths,T., & Shafto, P. (2011). Faster teaching by POMDP planning. In *Proceedings of 15th International Conference on Artificial Intelligence in Education*, Auckland, New Zealand.

150. Feng, M., Beck, J., Heffernan, N., & Koedinger, K. (2008). Can an intelligent tutoring system predict math proficiency as well as a standardized test? In R. Baker & J. Beck Beck (Eds.),

Proceedings of the 1st International Conference on Education Data Mining, 107-116. Montreal, Canada: Education Data Mining.

151. Beck, J. E. & Gong, Y. (2013). Wheel-Spinning: Students Who Fail to Master a Skill. *AIED 2013*: 431-440.

152. Barber, R & Sharkey, M. (2012). Course correction: Using analytics to predict course success. In S. Buckingham Shum, D. Gasevic, and R. Ferguson, Eds. LAK '12, *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 259-262. ACM, New York, NY. doi=10.1145/2330601.2330664

153. Thai-Nghe, N., Drumond, L., Krohn-Grimberghe, A., Schmidt-Thieme, L. (2010). Recommender system for predicting student performance. *Elsevier's Procedia Computer Science* 1(2) (2010) 2811-2819.

154. Paredes, W. C. and Chung, K. S. 2012. Modelling learning & performance: a social networks perspective. In S. Buckingham Shum, D. Gasevic, and R. Ferguson, Eds. LAK '12. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 34-42, ACM, New York, NY. doi= http://doi.acm.org/10.1145/2330601.2330617

155. Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258–272.

156. Beck, J. E.; and Chang, K.-m. (2007). Identifiability: A Fundamental Problem of Student Modeling. In *Proceedings of the 11th International Conference on User Modeling (UM 2007).* Corfu, Greece.

157. Baker, R. S., Corbett, A. T., & Aleven, V. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In *Intelligent Tutoring Systems* (pp. 406-415). Springer Berlin Heidelberg.

158. Huang, Y., González-Brenes, J., & Brusilovsky, P. (2014). General Features in Knowledge Tracing to Model Multiple Subskills, Temporal Item Response Theory, and Expert Knowledge. *Proceedings of the 7th International Conference on Educational Data Mining.* London, UK. pp. 84-91.

159. Khajah, M., Wing, R., Lindsey, R., Mozer, M. (2014). Integrating latent-factor and knowledge-tracing models to predict individual differences in learning. *Proceedings of the 7th International Conference on Educational Data Mining*. London, UK. pp. 99-106.

160. Beck, J. E., Chang, K. M., Mostow, J., & Corbett, A. (2008). Does help help? Introducing the Bayesian Evaluation and Assessment methodology. In *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*. Montreal, Canada. pp. 383-394.

161. Pardos, Z.A., Dailey, M. & Heffernan, N. (2011) Learning what works in ITS from non-traditional randomized controlled trial data. *The International Journal of Artificial Intelligence in Education*, 21(1-2):45-63.

162. Pardos, Z.A., Bergner, Y., Seaton, D., Pritchard, D.E. (2013) Adapting Bayesian Knowledge Tracing to a Massive Open Online College Course in edX. D'Mello, S. K., Calvo, R. A., and Olney, A. (eds.) *Proceedings of the 6th International Conference on Educational Data Mining (EDM)*. Memphis, TN. Pages 137-144.

163. van de Sande, B. (2013). Properties of the bayesian knowledge tracing model. *Journal of Educational Data Mining*, *5*(2), 1-10.

164. Reschly, A., & Christenson, S. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In S. Christenson, A. Reschly & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 3-19). Berlin: Springer.

165. Kumar, A., Palakodety, S., Wang, C., Wen, M., Rosé, C., Xing, E. (under review). Large Scale Structure Aware Community Discovery in Online Forums, submitted to *Empirical Methods in Natural Language Processing.*

166. Feng, D., Shaw, E., Kim, J., Hovy, E. (2006). An Intelligent Discussion-Bot for Answering Student Questions in Threaded Discussions, in *Proceedings of the International Conference on Intelligent User Interfaces IUI '06.*

167. Ravi, S. & Kim, J. (2007). Profiling Student Interactions in Threaded Discussions with Speech Act Classifiers, in *Proceedings of Artificial Intelligence in Education*.

168. Strijbos, J-W. & Stahl, G. (2007). Methodological issues in developing a multi-dimensional coding procedure for small-group chat communication, *Learning and Instruction* 17(4), pp 394-404.

169. Zemel, A., Xhafa, F., Cakir, M. (2007). What's in the mix? Combining coding and conversation analysis to investigate chat-based problem solving, *Learning and Instruction* 17(4), pp 405-415.

170. Singer, S.R. & Bonvillian, W.B. (2013). Two Revolutions in Learning. Science 22, Vol. 339 no. 6126, p. 1359.