

# Estimating Individual Differences for Student Modeling in Intelligent Tutors from Reading and Pretest Data

Michael Eagle<sup>1</sup>, Albert Corbett<sup>1</sup>, John Stamper<sup>1</sup>, Bruce M. McLaren<sup>1</sup>,  
Angela Wagner<sup>1</sup>, Benjamin MacLaren<sup>1</sup>, and Aaron Mitchell<sup>2</sup>

Human-Computer Interaction Institute<sup>1</sup>  
Department of Biological Sciences<sup>2</sup>  
Carnegie Mellon University

{meagle, corbett, jstamper, bmclaren, awagner,  
maclaren, apml}@andrew.cmu.edu

**Abstract.** Past studies have shown that Bayesian Knowledge Tracing (BKT) can predict student performance and implement Cognitive Mastery successfully. Standard BKT individualizes parameter estimates for skills, also referred to as knowledge components (KCs), but not for students. Studies deriving individual student parameters from the data logs of student tutor performance have shown improvements to the standard BKT model fits, and result in different practice recommendations for students. This study investigates whether individual student parameters, specifically individual difference weights (IDWs) [1], can be derived from student activities prior to tutor use. We find that student performance measures in reading instructional text and in a conceptual knowledge pretest can be employed to predict IDWs. Further, we find that a model incorporating these predicted IDWs performs well, in terms of model fit and learning efficiency, when compared to a standard BKT model and a model with best-fitting IDWs derived from tutor performance.

**Keywords:** BKT, Genetics, Machine Learning, Student Modeling

## 1 Introduction

Models of student learning have been successfully employed by intelligent tutoring systems to improve learning outcomes for more than two decades. Student modeling has been used both to individualize curriculum sequencing [1, 2, 3] and/or to individualize hint messages [4, 5]. Each of the modeling frameworks cited here employs a Bayesian method to infer student knowledge from student performance accuracy, and Bayesian modeling systems have been shown to accurately predict students' tutor and/or posttest performance [1], [3], [6,7].

These models generally individualize modeling parameters for individual knowledge components (KCs, also referred to as skills) [8], but not for individual students. Several studies have shown that individualizing parameters for students, as well as for KCs, improves the quality of the models [1], [9, 10, 11, 12].

These approaches to modeling individual differences among students have monitored student performance after the fact, in tutor logs that have been previously collected to derive individualized student parameters for the tutor module(s). While these efforts have proven successful, they complicate the actual use of student modeling within an ITS module, since the concurrent estimation and use of individualized parameters in a tutor lesson is generally quite challenging, at best. In this paper we examine whether parameter estimates can be individualized for students prior to embarking on a tutor module, based on student performance in earlier activities. In particular, we examine whether parameter estimates can be individualized based on performance in two activities that naturally precede tutor modules: reading on-line instructional text and taking a conceptual knowledge pretest.

We explore this issue in the Bayesian Knowledge Tracing modeling framework [1] and in a unit of the Genetics Cognitive Tutor [6]. In the following sections we describe Knowledge Tracing, the on-line student activities, the predictors derived from students' reading and pretest activities, and our success in using these predictors to model individual differences in student learning and performance in the tutor.

## 1.1 Bayesian Knowledge Tracing

Bayesian Knowledge Tracing (BKT) [1] employs a two-state Bayesian learning model for each knowledge component (KC) in a tutor curriculum: at any time a student either has learned or not learned a given KC. BKT employs four parameters to estimate the probability that a student has learned each KC:

- $pL_0$  the probability a student has already learned how to apply a KC
- $pT$  the probability a student learns a KC at each opportunity to apply it
- $pG$  the probability a student will guess correctly if the KC is not learned
- $pS$  the probability a student will make an error when the KC has been learned

Cognitive Tutors employ BKT to implement Cognitive Mastery, in which the curriculum is individualized to afford each student just the number of practice opportunities needed to enable the student to "master" each KC.

**Individual Differences.** Knowledge Tracing and Cognitive Mastery generally employ best-fitting estimates of each of the four parameters for each individual KC but *not* for individual students. In this work, we incorporate individual differences among students into the model in the form of individual difference weights. Following Corbett and Anderson [1], four best-fitting weights are estimated for each student, one weight for each of the four parameter types,  $wL_0$ ,  $wT$ ,  $wG$ ,  $wS$ . In estimating and employing these *individual difference weights* (IDWs), we convert each of the four probability estimates to odds form ( $p/(1-p)$ ), multiply the odds by the corresponding student-specific weight and convert the resulting odds back to a probability (equation 1.) Let  $i$  represent the parameter type, ( $pL_0$ ,  $pT$ ,  $pG$ ,  $pS$ ),  $k$  represent the KC and  $s$  the student. The individually weighted parameter for each KC and student,  $p_{iks}$ , is thus:

$$p_{iks} = p_{ik} * w_{is} / (p_{ik} * w_{is} + (1 - p_{ik})) \quad (1)$$

where  $p_{ik}$  is a best fitting parameter estimate for the KC across all students and  $w_{is}$  is the corresponding individual difference weight for the student.

## 2 Student Activities in This Study

The genetics topic in this study is gene interaction, which examines how two genes can interact in controlling a single phenotypic trait. When two genes, each with a dominant and recessive allele, control a single trait, e.g., coat color in cattle, there can be up to four different resulting phenotypes (four colors). But, there are many ways the two genes can interact that result in only two or three different phenotypes. The goal of the Genetics Cognitive Tutor “process-modeling” lesson in this study is to help students understand and model the different ways two genes can interact to yield two, or three, or four phenotype values. This study focuses on three activities on this topic that students completed in succession: reading gene interaction instructional text online, taking a gene interaction pretest, and finally using the Genetics Cognitive Tutor module on Gene Interaction Process Modeling.

**On-Line Instructional Text.** The online instructional text consisted of 23 screens, structured like pages in a book. Students could move forward and backward through the screens, one screen at a time. After a student touched each page once a “done” button appeared and the student could then continue reading, or exit at any time.

**Conceptual Knowledge Pretest.** Students completed a pretest with nine conceptual questions divided into three topics. The first three questions focused on general knowledge of basic Mendelian transmission with 2 genes, the second three questions focused on process modeling — reasoning about phenotypes that could or could not result from given gene interactions, and the last three questions focused on abductive (backward) reasoning, reasoning about gene interactions that could or could not have given rise to observed phenotypes. This is not a problem-solving pretest; the last six questions are not similar to the Cognitive Tutor problems. Instead, they required students to reason about genetics processes and abductive reasoning more abstractly.

**Genetics Cognitive Tutor: Gene Interaction Process Modeling.** The Genetics Cognitive Tutor (GCT) lesson consisted of 5 process-modeling problems. In each problem, students are given a description of how two genes interact to determine a phenotype, e.g., coat color in cattle. Students (a) map the description onto one of seven gene interaction templates with 3 menus, (b) identify the phenotypes of the four true-breeding genotypes, (c) model the offspring genotypes and phenotypes resulting from two different parental crosses, and finally (e) summarize the phenotypes associated with all possible individual genotypes and how the phenotypes arise.

**The Cognitive Model for GCT Process Modeling.** There are an average of 45 steps in each of these tutor problems. Some of the KCs governing these steps are unique to a problem, while others are applicable in multiple problems. In this analysis we excluded KCs that occurred only one or two times across the 5 problems. Of the remaining 31 KCs, 28 appeared 5 times across the curriculum and 3 appeared 4 times.

### 3 Predictors

Our goal is to examine the feasibility of setting individual difference weights for students before students begin work in a tutor lesson. Several studies have focused on incorporating individual differences into BKT models from the tutor data itself. Corbett and Anderson [1] showed that a BKT model with the four IDWs described in Section 2 was a better predictor of individual differences in posttest performance than a standard, non-individualized BKT model. Pardos and Heffernan [10] individualized just the  $pL_0$  parameter, the initial probability the student knows a KC before its first application, based on either the student's first attempt at each KC within the lesson or on all attempts at each KC — and found that either individualized method yielded reliably better fits to multiple data sets than the non-individualized BKT model. Yudelson, Koedinger and Gordon [11] individualized both learning parameters,  $pL_0$  and  $pT$ , based on student accuracy in a tutor lesson, and found that individualizing  $pT$  yielded reliably better fits than the non-individualized BKT model, while individualizing  $pL_0$  did not reliably improve the goodness of fit. Lee and Brunskill [9] derived 4 individual difference parameters based on performance in a tutor module and found that the resulting model recommended substantially more practice for some students and substantially less practice for others than the non-individualized model.

Finally, in an alternative approach to BKT, a variety of student modeling frameworks grounded in Item-Response Theory employ a single individual difference parameter as a basic component of the model [12,13,14].

To date, these approaches estimate individual student differences after the fact on tutor data that has already been collected. We examine whether individual differences can be modeled based on prior activities that are natural components of an on-line learning system so that they can be used when a student first begins an ITS module.

#### 3.1 Predictors Derived from Instructional Text Reading Performance.

We examine two measures of student reading performance: reading time, and revisiting pages in the text.

**Reading Time.** No prior ITS research employs reading rates to individualize parameters in a learning environment, but there is substantial evidence that reading time varies measurably with comprehension difficulty, and it follows that reading time may prove sensitive to individual differences in comprehension difficulty. Harvey and Anderson [15] showed that reading times for on-line declarative instruction in the ACT Programming Tutor are sensitive to differences in processing time necessary to encode familiar vs. novel material. More generally, an extensive research literature demonstrates that reading time is sensitive to relative comprehension difficulty [16].

**Text Pages Revisited.** Students can read through the instructional text as they would pages in a book. Some students may choose to strictly read forward through the text, while others may choose to revisit earlier pages in the text. Students who re-read text may be demonstrating a meta-cognitive self-monitoring skill, which, if it transfers to problem-solving in the tutor may be correlated with  $p(T)$ , learning rate in the tutor.

### 3.2 Predictors Derived from a Conceptual Knowledge Pretest.

Some prior projects have employed pretest accuracy to initialize ITS student models [3, 17]. We examine several measures of students' pretest performance.

**Pretest Accuracy.** We examine whether students' pretest accuracy on each of the three types of pretest questions, general knowledge, process modeling and abductive reasoning, predicts individual difference in learning or performance in the GCT gene interaction process modeling lesson.

**Pretest Answer Changes.** We examine whether changing answers in the pretest from a correct initial answer to an incorrect final answer, or vice versa, is a predictor of individual differences in learning or performance in the tutor module. Checking and changing answers may be evidence of a meta-cognitive self-monitoring skill that may translate into higher learning rates in the tutor module. Alternatively, it may be correlated with the slip rate in the tutor,  $p(S)$ , if the students slipped in making the initial error they are correcting.

**Time on Task.** Finally we examine whether time to complete the pretest is a predictor of individual differences in the tutor module.

## 4 Methods

The data analyzed in this study come from 83 undergraduates enrolled in either a genetics or introductory biology course. All students were recruited to participate in the study for pay. Students participated in two 2.5-hour sessions on consecutive days in a campus computer lab. In this study, the first session focused on gene interaction and students read the on-line gene interaction instructional text, took the on-line pretest, and used the gene interaction process modeling tutor module as the first three activities in this session. The study focuses on modeling the 83 students' first actions on 12,287 problem steps in the tutor module.

### 4.1 Fitting Procedures

First, we found best-fitting group parameter estimates for each of the 4 parameters ( $pL_0$ ,  $pT$ ,  $pG$ ,  $pS$ ) in the standard BKT ("SBKT") model for each of the 31 different knowledge components in the tutor lesson, with nonlinear optimization. The objective function takes the observed opportunities for a single skill and a set of group parameters as input and returns the negative log-likelihood (-LogLik). Optimization ultimately returns the set of group parameters that best fit the skill. Both  $pG$  and  $pS$  were bounded to be less than 0.5, as in [18] to avoid paradoxical results that arise when these performance parameters exceed 0.5 (e.g., a student with a higher probability of knowing a KC is less likely to apply it correctly.)

Second, we re-fit the tutor data with an individualized BKT model: We obtained four best-Fitting Individual Difference Weights (IDWs) for each of the 83 students, one weight for each of the four parameter types,  $wL_0$ ,  $wT$ ,  $wG$ ,  $wS$  to construct this "FIDW" model. As described in Section 1 equation 1, each student's four weights are mapped across the best-fitting group learning and performance parameter estimates

for each of the 31 KCs to individualize these parameter estimates. The objective function takes the fixed group parameters, the observed opportunities for a student, and a set of IDWs ( $wL_0$ ,  $wT$ ,  $wG$ ,  $wS$ ) and returns the -LogLik. Optimization ultimately returns the set of IDWs that maximize the fit for the student.

Third, we derived 12 predictive features, 6 from the on-line reading data and 6 from the pretest data to predict these four individual difference weights for the 83 students, as displayed in Table 1. We performed a factor analysis on log reading times for the 23 individual pages to reduce the number of predictors. The factor analysis yielded a total of four factors (RTF1, RTF2, RTF3, RTF4), which each account for at least 10% of the variance and align with subtopics in the text, as summarized in the table. These four factors accounted for 54% of the total variance and additional factors each accounted for less than 5% of the variance.

**Table 1.** 12 Predictor variables extracted from the on-line reading and pretest data

<b>RTF1</b>	Reading: Time for a 5-page intro with familiar content on basic Mendelian genetics
<b>RTF2</b>	Reading: Time for 6 pages with charts of various ways 2 genes can interact
<b>RTF3</b>	Reading: Time for 3 pages on parental crosses with offspring genotypes & traits
<b>RTF4</b>	Reading: Time for 2 pages with full-page diagrams of dominant & recessive alleles
<b>RRNP</b>	Reading: Total number of previous pages re-read
<b>RRTD</b>	Reading: Total distance traversed (intervening pages) in re-reading text pages
<b>PACC1</b>	Pretest: % Correct for 3 general knowledge questions
<b>PACC2</b>	Pretest: % Correct for 3 process modeling questions
<b>PACC3</b>	Pretest: % Correct for 3 abductive reasoning questions
<b>PCIC</b>	Pretest: Number of answers initially incorrect changed to correct
<b>PCCI</b>	Pretest: Number of answers initially correct changed to incorrect
<b>PTime</b>	Pretest: Total time to complete the pretest

Fourth, we employed each of these 12 variables to independently predict the four sets of IDWs:  $wL_0$ ,  $wT$ ,  $wG$ ,  $wS$ . Since these are multiplicative weights, we fit a transformation of the weights  $w/(1+w)$ . This transformation has the property that the neutral weight 1.0 (which does not modify the corresponding best-fitting group parameter) is the midpoint of the transformed scale. We built a robust regression model with the 12 predictors for each of the IDWs. Robust regression is less sensitive to outliers, variable normality, and other violations of standard linear regression assumptions.

Finally, after deriving the 4 predicted IDWs for each of the 83 students, we recalculated the earlier FIDW BKT model with the predicted IDWs, in place of the best-fitting IDWs to construct the “PIDW” model. In summary, we have three BKT model variants:

1. SBKT: Standard BKT model with best-fitting group parameter estimates,
2. FIDW: Standard BKT model with Fitted Individualized Difference Weights
3. PIDW: Standard BKT model with Predicted Individualized Difference Weights

## 5 Results and Discussion

Table 2 summarizes our results. Columns 2 and 3 summarize the overall fit of the standard BKT and the two IDW models to the tutor data. Column 2 displays root mean squared error (RMSE) for the fits and column 3 displays Accuracy (the probability a model correctly predicts students' correct or incorrect responses, with a 0.5 threshold on predicted accuracy). As can be seen, the FIDW model with best-fitting IDWs fits the tutor data best; it reduces RMSE by 8.7% compared to the standard, non-individualized SBKT model (0.2794 vs. 0.3059). The new PIDW model with predicted IDWs is about 40% as successful as the best-fitting FIDW model: The new model reduces RMSE by 3.6% compared to the standard SBKT model (0.2950 vs. 0.3509). The FIDW model is also about 2.4% more accurate than the SBKT model (0.8948 vs. 0.8742) while the PIDW model is about 0.8% more accurate than the SBKT model (0.8812 vs. 0.8742).

**Table 2.** Goodness of fit of the 3 models and differences in practice needed to reach mastery.

Model	RMSE	Accuracy	# Students Needing Less	# Fewer Opportunities Needed	# Students Needing More	# More Opportunities Needed
SBKT	0.3059	0.8742	-	-	-	-
FIDW	0.2794	0.8948	56 (46)	17.27 (17.24)	27 (19)	27.04 (27.37)
PIDW	0.2950	0.8812	54 (46)	10.48 (10.96)	27 (19)	11.59 (13.58)

Even small differences in model fits, such as what we found in this study, can have large effects on the amount of recommended work assigned to the student [19]. In order to explore the practical impact of the individualized models, we examined the number of practice opportunities that were necessary for students to reach mastery under each of the three models — that is, the number of opportunities required for  $pL$  (the probability the student has learned a rule) to reach 0.95. This analysis is possible because students completed a fixed curriculum in this study with 4 or 5 opportunities per KC, and most students reached mastery for all of the KCS in the available number of opportunities under all three models.

On average students mastered 94% of the skills under the SBKT model, 90% under the FIDW model, and 93% under the PIDW model. If a student failed to reach mastery on a KC under one model, we conservatively estimated that the student would reach mastery on the next opportunity. On average students needed 57.22 total opportunities to reach mastery of the 31 KCs under the SBKT model, 53.65 total opportunities under the FIDW model, and 53.71 under the PIDW model.

The bottom two rows in the last four columns of Table 2 show how many students need less practice to reach mastery under each of the individualized BKT models than under the standard BKT model, and how many students need more practice. The numbers in parentheses show how many students are common to the two sets under

the two models. These columns also show how much more or less practice the students need before the model would consider them to have mastered the KCs.

Both individualized models, FIDW and PIDW, substantially modify the amount of practice needed to reach mastery compared to the standard SBKT model. Under the best-fitting FIDW model, 56 students needed less practice to master all the KCs than under the standard SBKT model and on average these students required 17.3 fewer practice opportunities to reach mastery under FIDW than under SBKT. Under the predicted PIDW model, 54 students needed an average of 10.5 fewer opportunities to master all the KCS than under the SBKT model. The two individualized model agree on a set of 46 students who need fewer practice opportunities to reach mastery, but again the FIDW model requires less practice (17.2 opportunities) of these students than the PIDW model (11.0 opportunities).

Under both the FIDW and PIDW models, 27 students need more practice opportunities to reach mastery than under the SBKT model, but students need 27 more practice opportunities under the FIDW model and only 11.6 more opportunities under PIDW model. The two models agree on a set of 19 students who need more practice, but again the FIDW model requires more practice than the PIDW model.

Overall, the FIDW and PIDW models were in 78% agreement on which students needed fewer or more opportunities to master all the KCs than under the standard SBKT model. The new predicted PIDW model reaches roughly 60% of the potential learning efficiency gains identified by the best-fitting FIDW model, and does so without the use of the student tutor performance data.

### 5.1 The Predictive Models for the Four Individual Difference Weights

Table 3 displays the coefficients for each of the 12 predictors in the regression model for each of the four IDWs. The predictors that entered reliably into the robust regression model are highlighted with asterisks.

The most interesting result is that student behaviors in reading the text are, in fact, reliable predictors of some individual difference weights. Three of the reading time factors, RTF1, RTF2, RTF4 each reliably predicted one of the four individual differences weights ( $wT$ ,  $wG$ , and  $wG$  respectively). The pages that load on RTF1 specifically are introductory pages on basic Mendelian transmission that should be familiar to all the students and this factor is inversely related to  $wT$  — the longer students take reading what should be familiar text, the lower their learning rate in the tutor. However, student behaviors in re-visiting pages did not reliably predict any IDWs.

Not surprisingly, more pretest variables reliably entered into the four IDW models. Differences in student accuracy on general knowledge (PACC1) and on process-modeling (PACC2) — the same type of reasoning as in this tutor unit — each reliably predict two of the four IDWs. Three other pretest measures, including student accuracy on abductive reasoning questions (PACC3) — a type of reasoning not employed in this tutor unit, total time (PTime) and number of changes from an initially incorrect answer to a correct answer (PCIC) each marginally predicted one IDW.



**Table 3.** Coefficient Summary Table ( $\cdot < 0.10$ ,  $* < 0.05$ ,  $** < 0.01$ )

	wL0	wT	wG	wS
<b>(Intercept)</b>	0.0528	0.0785	0.0623	0.8752**
<b>RRTD</b>	-0.0314	0.0317	-0.0061	0.0215
<b>RRNP</b>	0.0221	-0.0055	-0.0056	-0.007
<b>RTF1</b>	-0.0046	-0.0834*	-0.0193	-0.0089
<b>RTF2</b>	0.0338	0.0335	-0.0627*	0.0053
<b>RTF3</b>	0.0131	0.017	0.0192	-0.0215
<b>RTF4</b>	-0.004	0.0204	-0.052*	0.0029
<b>PACC1</b>	0.3504**	0.1469	0.1109	-0.3038**
<b>PACC2</b>	0.2154	-0.0398	0.563**	-0.3021**
<b>PACC3</b>	0.0841	0.4373	0.144	-0.1699
<b>PCIC</b>	0.0096	-0.0248	-0.0327	0.005
<b>PCCI</b>	0.0143	-0.0092	0.0352	0.0189
<b>Ptime</b>	0	0.0004	0	0.0001
<b>RMSE</b>	0.1809	0.2245	0.2055	0.1443

## Conclusion

We have developed and discussed a method of inserting individual student differences into a traditional Bayesian Knowledge Tracing model that employs pre-tutor reading and test data to predict individual difference weights. This is important because integrating IDWs into an intelligent tutor is much easier if the IDWs can be assigned before the student starts working with the tutor. An advantage of our method is that it can be implemented easily; only a single adjustment needs to be made to each of the group parameters before the student starts the lesson. This initial attempt to pre-set individual difference weights is already quite successful.

The goodness of fit of this new predictive PIDW BKT model falls almost midway between the standard non-individualized SBKT model and the fitted FIDW BKT model. Further, the individualized practice recommendations for the predictive PIDW BKT model are similar to the practice recommendations for the fitted FIDW BKT model, although the new PIDW model does not identify all the opportunities to decrease the amount of practice for some students, nor the need to increase the amount of practice for other students, that are identified in the best-fitting FIDW model. However, if implemented, the PIDW model would result in some students needing an average of 18% fewer total practice opportunities to reach mastery with other students needing an average of 20% more total practice opportunities. This is a potentially meaningful difference, as it could lead to students spending just the right amount of time with the tutor to achieve mastery.

An important finding is that student data from the reading the instructional text is a useful predictor of learning and performance in an intelligent tutor. Three reading time factors entered reliably into predictive models for individual difference weights in the study. Several conceptual pretest variables also reliably predicted individual differences in learning and performance in an ITS. These results suggest that it is possible to assign IDWs to students before they begin to use the tutor. We expect, but it remains for future research to explore, that other individual difference frameworks can also benefit from using data from the prior to tutor activities as predictors for initial IDW assignment.

**Acknowledgements.** This research was supported by the National Science Foundation under the grant “Knowing What Students Know: Using Education Data Mining to Predict Robust STEM Learning”, award number DRL1420609.

## References

1. Corbett, A.T., Anderson, J.R.: Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4, 253-278. (1995)
2. Mayo, M., Mitrovic, A. Optimising ITS behaviour with Bayesian networks and decision theory. *International Journal of Artificial Intelligence in Education*, 12, 124-153 (2001).
3. Shute, V.: Smart: Student Modeling Approach for Responsive Tutoring. *User Modeling and User-Adapted Interaction*, 5 (1), 1-44. (1995)
4. Ganeshan, R., Johnson, L., Shaw, E., Wood, B.: Tutoring diagnostic problem solving. In G. Gauthier, C. Frasson, K. VanLehn (eds.) *ITS2000 Intelligent Tutoring Systems*, LNCS vol. 1839, pp. 33-42. Springer, Heidelberg. (2000)
5. Conati, C., Gertner, A., VanLehn, K. Using Bayesian networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interaction*, 12, 371-417. (2002)
6. Corbett, A.T., MacLaren, B., Kauffman, L., Wagner, A., Jones, E. A.: Cognitive Tutor for Genetics Problem Solving: Learning Gains and Student Modeling. *Journal of Educational Computing Research*, 42 (2), 219-239. (2010)
7. Gong, Y., Beck, J., Heffernan, N.: Comparing knowledge tracing and performance factor analysis by using multiple model fitting. In V. Aleven, J. Kay, J. Mostow (eds.) *ITS2010 Intelligent Tutoring Systems*. LNCS vol. 6094, pp. 35-44. Springer, Heidelberg. (2010)
8. Koedinger, K., Corbett, A. Perfetti, C.: The Knowledge-Learning-Instruction (KLI) framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, 36 (5), 757-798. (2012)
9. Lee, J., Brunskill, E.: The impact of individualizing student models on necessary practice opportunities. In: Yacef, K., Zaiane, O., Hershkovitz, A., Yudelson, M., Stamper, J. (eds.) *EDM2012 Proceedings of the 5<sup>th</sup> International Conference on International Educational Data Mining Society*, 118-125. (2012)
10. Pardos, Z., Heffernan, N.: Modeling individualization in a Bayesian networks implementation of Knowledge Tracing. In De Bra, P., Kobsa, A., Chin, D. (eds.) *UMAP2010 Proceedings of the 18<sup>th</sup> International Conference on User Modeling, Adaptation, and Personalization*, LNCS vol. 6075, pp. 255-266. Springer, Heidelberg (2010)
11. Yudelson, M., Koedinger, K., Gordon, G.: Individualized Bayesian knowledge tracing models. In: Lane, C., Yacef, K., Mostow, J., Pavlik, P. (eds.) *AIED2013 Artificial Intelligence in Education*, LNCS vol. 7926, pp. 171-180, Springer, Heidelberg. (2013)

12. Pirolli, P., Wilson, M.: A theory of the measurement of knowledge content, access, and learning. *Psychological Review*, 105(1), 58-82. (1998)
13. Cen, H., Koedinger, K., Junker, B.: Comparing two IRT models for conjunctive skills. In B. Woolf, E. Atmour, R. Nkambou, S. Lajoie (eds.) *ITS2008 Intelligent Tutoring Systems LNCS vol. 5091*, pp. 796-798. Springer, Heidelberg. (2008)
14. Pavlik, P., Yudelson, M., Koedinger, K.: Using contextual factors analysis to explain transfer of least common multiple skills. In G. Biswas, S. Bull, J. Kay, A. Mitrovic (eds.) *AIED2011 Artificial Intelligence in Education, LNCS vol. 6738*, pp. 256-263. Springer, Heidelberg. (2011)
15. Harvey, L., Anderson, J.: Transfer of declarative knowledge in complex information processing domains. *Human-Computer Interaction*, 11 (1), 69-96. (1996)
16. Zwann, R., Singer, M.: Text comprehension. In A. Graesser, M. Gernsbacher, S. Goldman (eds.) *Handbook of discourse processes*, pp. 83-121. Mahwah, NJ: Erlbaum. (2003)
17. Arroyo, I., Beck, J., Woolf, B., Beal, C., Schultz, K.: Macroadapting Animalwatch to gender and cognitive differences with respect to hint interactivity and symbolism. In G. Gauthier, C. Frasson, K. VanLehn (eds.) *ITS2000 Intelligent Tutoring Systems, LNCS vol. 1839*, pp. 574-583. Springer, Heidelberg. (2000)
18. Baker, R., Corbett, A., Aleven, V.: More Accurate Student Modeling Through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing. In B. Woolf, E. Aimeur, R. Nkambou, S. Lajoie (eds.) *ITS2008 Intelligent Tutoring Systems. LNCS vol. 5091*, pp. 406-415. Springer, Heidelberg. (2008)
19. Koedinger, K., Stamper, J., McLaughlin, E., Nixon, T.: Using data-driven discovery of better student models to improve student learning. In: Lane, C., Yacef, K., Mostow, J., Pavlik, P. (eds.) *AIED2013 Artificial Intelligence in Education, LNCS vol. 7926*, pp. 421-430, Springer, Heidelberg. (2013)