Predicting Individualized Learner Models Across Tutor Lessons

Michael Eagle, Albert Corbett, John Stamper, Bruce Mclaren
Human-Computer Interaction Institute
Carnegie Mellon University
{meagle,ac21,jstamper,bmclaren}@andrew.cmu.edu

ABSTRACT

In this work we use prior to tutor-session data to generate an individualized student knowledge model. Intelligent learning environments use student models to individualize curriculum sequencing and help messages. Researchers decompose the learning tasks into sets of Knowledge Components (KCs) that represent individual units of knowledge; the student model estimates a parameters for each KC, but not for each student. Using existing performance data to adjust parameters for each individual student improves model fit, and leads to different practice recommendations. However, in order to be implemented in a live system we need to have a method to estimate the student parameters using only the student's prior activities. In this work, we use data collected from student reading, prior tutor lessons, to predict individualized difference weights for parameters of a Bayesian Knowledge Tracing (BKT) variant. We find that best-fitting student parameters trained on previous lessons do not directly transfer to new lessons; however, we can effectively predict the student parameters for the new lesson by using features derived from prior lessons, and prior to tutor text-reading transaction data.

KEYWORDS

Individualization, Student Modeling, BKT, Genetics

1 INTRODUCTION

Learner models of domain knowledge have been successfully employed for decades in intelligent tutoring systems (ITS), to individualize both curriculum sequencing [8, 19, 23, 24] and help messages [6, 13]. Bayesian methods are frequently employed in ITSs to infer student knowledge from performance accuracy, as in the citations above, as well as in other types of learning environments [21], and Bayesian modeling systems have been shown to accurately predict students' tutor and/or posttest performance [7, 8, 14, 24]. These models generally individualize modeling parameters for individual knowledge components (KCs, also referred to as skills) [16], 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 [7, 18, 22, 27]. 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 don't achieve the goal of dynamic student modeling within an ITS, since estimating and using individualized parameters concurrently within a tutor lesson is quite difficult. In this paper we examine how well individual differences in student learning in a lesson of the Genetics Cognitive Tutor [7] can be predicted ahead of time from two types of prior online activities: reading instructional text and solving problems in prior tutor lessons. In the following sections we describe Knowledge Tracing, the on-line student activities, the predictors derived from students' reading and prior tutor activities, and our success in using these predictors to model individual differences in the tutor.

EDM'18, July 2018, El Buffalo, New York USA
.

1.1 Modeling Framework

Bayesian Knowledge Tracing (BKT) estimates the probability that a student knows each of the knowledge components (KC) in a tutor lesson. It employs a two-state Bayesian learning model — at any time a student either knows or does not know a given KC — and employs four parameters, which are estimated separately for each KC: $p(L_0)$ — initial knowledge the probability a student has learned how to apply a KC prior to the first opportunity to apply it in a lesson. $p(T)$ — learning rate the probability a student learns a KC at each chance to apply it. $p(G)$ — guessing the probability a student will guess correctly if the KC is not learned. $p(S)$ — slips the probability a student will make an error when the KC has been learned. BKT is employed in Cognitive Tutors to implement Cognitive Mastery, in which the curriculum is individualized to assign only the number of practice opportunities needed to enable the student to "master" each of the KCs, which is generally operationalized as a 0.95 probability that the student has learned the KC.

1.1.1 *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 [8], four best-fitting weights are estimated for each student, one weight for each of the four parameter types, wL0, 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. (See [8] for computational details.)

In this paper we focus on four types of BKT models for the third lesson in a Genetics Cognitive Tutor curriculum on genetic pathways analysis to examine how well IDWs in a tutor lesson can be predicted from prior online activities. The four models are: (1) a standard BKT model (SBKT) with no individualization, (2) a model with best-fitting IDWs for lesson 3 (BFIDW-L3), (3) models with best-fitting IDWs from prior lessons, and (4) a model with predicted individual difference weights derived from earlier activities. We compare how much each of the three types of individualized models improves upon the non-individualized SBKT fit (1).

Eagle et al. [11] estimated individual difference weights using reading performance data, pretest scores, resulting in a predictive model 40% as effective as the best-fitting model; the predictive model was improved for a second lesson reaching 60% of the best-fitting model by using previous lesson data [11]. As pretests do not necessarily appear in all online environments, in this paper, we examine how well we can predict IDWs in a third lesson with the same types of reading measures as in [11, 12] along with an expanded set of tutor performance measures.

2 STUDENT ACTIVITIES IN THIS STUDY

The students in this study worked through two successive topics in the genetic pathway analysis curriculum within the Genetics Cognitive Tutor. The first topic, gene interaction, examines the different ways two genes can interact in controlling a single trait. e.g., coat color in cattle. The second topic, gene regulation, focuses on three-gene systems in which two genes function together to control the expression of the third gene.

For each topic students completed five activities: reading instructional text, taking a conceptual-knowledge pretest, completing two Genetics Cognitive Tutor lessons and completing a problem-solving posttest. The two tutor lessons for each topic require students to think about the topic in contrasting ways. In the first, "forward reasoning" or process modeling lesson, students are given descriptions of how genes interact in a system and reason about the resulting behavior of the system. In the second, "backward reasoning," or abductive reasoning lesson, students are given descriptions of how genetic systems behave, and draw conclusions about how the underlying genes interact.

Online Instructional Text: The first text on gene interaction consists of 23 screens, and the second gene regulation text consists of 20 screens. The screens are structured like pages in a book. Students can move forward and backward through the screens, one screen at a time. After a student touches each page once a "done" button appears and the student can then continue reading, or exit at any time.

Cognitive Tutor Lessons: The first tutor lesson, Gene Interaction Process Modeling, consists of 5 problems, averaging 45 steps per problem. The second tutor lesson, Gene Interaction Abductive Reasoning, consists of 6 problems, averaging 25 steps each. Features of student performance in these two lessons (along with features of their reading performance) are employed to predict individual differences in the third tutor lesson, Gene Regulation Process Modeling, which consists of 9 problems with 27 steps each.

3 PREDICTORS

In this study, we examine three types of student performance variables as predictors of best fitting Lesson 3 IDWs: Aspects of reading the two texts, Lesson 1 and Lesson 2 IDWs, and features of student performance in completing tutor Lessons 1 and 2.

3.1 Instructional Text Reading Predictors

Two types of measures of students' reading performance were derived for both the Topic 1 (gene interaction) and Topic 2 (gene regulation) instructional texts: reading time per page and pages revisited in the text. Eagle et al [11, 12] found that both types of reading measures for the gene interaction text entered reliably into predictive models for IDWs for both of the gene interaction tutor lessons.

Reading Time: A factor analysis was performed on log reading times for the 23 Topic 1 pages and a factor analysis on log reading times for the 20 Topic 2 pages to reduce the number of predictors. Each analysis yielded (a different set of) four reading time factors.

Text Pages Revisited: Students may choose to strictly read forward through a text, or may choose to revisit earlier pages. Two measures of student behavior in revisiting text pages were calculated: the number of pages re-read and the number of intervening pages traversed in re-reading text pages.

3.2 Prior Lesson Model Predictors

We derived a total of total of 16 predictors from the lesson 1 and 2 student models.

Individual Difference Weights: Three sets of best-fitting individual-difference weights were derived (1) for the 31 KCs in Lesson 1, (2) for the 22 KCs in Lesson 2, and (3) for the combined set of 53 KCs in Lessons 1 and 2.

Probabilities students learned the Lesson 1 & 2 KCs: At the end of a lesson, BKT yields a probability that a student knows each KC in the lesson. Two measures of each student's knowledge at the end of each lesson were calculated: the number of unmastered skills and the minimum probability the student knows any single KC.

3.3 Tutor Performance Features

Finally, thirteen predictors based on student performance in each of the two tutor lessons were derived. Raw error rate for students' first action at each problem- solving step in each lesson, and average response time for students' first action at each problem-solving step in each lesson were calculated.

In addition, for each of the two lessons the following 11 measures of students' metacognitive skills were calculated. Most of these have previously been shown [10] to correlate with measures of robust learning, including direct transfer of knowledge, which is similar students' initial knowledge, pL0, and preparation for future learning, which is similar to students; learning rate wT:

Help avoidance [1]: the proportion of problem solving steps in which the probability the student knows the relevant KC is low and the student's first action is an error instead of a hint request.

Bug Messages: the proportion of each student's actions in which a bug message (an error message generated when a student's behavior matches a known misconception) is followed by a long pause, and the proportion in which a bug message is followed by a short pause.

Hint Messages: the proportion of each student's actions in which a hint request is followed by a long pause, and the proportion in which a hint request is followed by a short pause.

Known-KCs: the proportion of each student's actions in which the student knows the relevant skill well and there is a long pause before responding, and the proportion in which the student knows the skill well and there is a short pause.

Off-Task and Gaming Variables: The proportion of actions in which an automatic detector determined the student was gaming the system [9] was calculated, (e.g., systematic guessing, or quickly drilling down through the tutor's hints to find the correct answer), as was the proportion of fast responses that were not identified as gaming by the detector. Also, we calculated for each student both the proportion of actions in which an automatic detector determined the student was off task [3] and the proportion of actions where there was a long pause not identified as off-task.

4 METHODS AND MATERIALS

The data analyzed in this study come from 80 CMU undergraduates enrolled in either genetics or introductory biology courses who were recruited to participate in this study for pay. The students participated in two 2.5-hour sessions on consecutive days in a campus computer lab. The first session focused on the first topic, gene interaction and the second session focused on the second topic, gene regulation. In each session students completed five activities: Read an on-line instructional text on the session topic; completed a pretest on the topic; completed two Genetics Cognitive Tutor modules on the session topic, a “forward“ process-modeling module and a “backward“ abductive reasoning module; and completed a problem-solving posttest. This study focuses on modeling the 22,681 problem-solving steps in the third, gene regulation process-modeling tutor lesson.

4.1 Fitting Procedures

We first found best-fitting group parameter estimates for each of the 4 parameters (pL0, pT, pG, pS) in the standard BKT (SBKT) model for each of the 47 KCs in Lesson 3, with nonlinear optimization. We optimize on negative log-likelihood and generate the best fitting set of group parameters for each of the 47 KCs. Both pG and pS were bounded to be less than 0.5, as in Baker et al., [4] 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 generate individualized BKT models by optimizing a new set of four Individual Difference Weights (IDWs,) one for each of the four standard BKT parameters, wL0, wT, wG, wS, for each of the 80 students. The optimization process takes as input the SBKT model, and the observed student opportunities, and produces the best fitting set of IDWs for each student.

Third, we derived the 6 reading features for text 2, and tutor performance measures for Lesson 1 and 2 that had not previously been derived in [11, 12]. Along with the measures from text 1, the best-fitting IDWs for Lessons 1 and 2, and the Lesson-1 measures that had been derived previously [11, 12], this yields a total set of 50 predictor variables.

We employed these 12 reading variables (6 for each topic) and the 38 tutor performance variables (19 for each lesson) to independently predict the four Lesson 3 IDWs: wL0, wT, wG, wS. Since we are predicting multiplicative weights, we fit a transformation of the weights w/(1+w). This transformation has the property that the neutral weight 0.5 (which does not modify the corresponding best-fitting group parameter), is the midpoint of the transformed scale.

4.2 Model and Feature Selection

In order to generate the predictive IDW model we first reduced the number of features with Least Angle Regression (LAR) [25] a variant of Lasso. For each of the four Lesson 3 IDWs we use LAR to select the best 12 predictors (out of 50,) Twelve predictors were selected to match with models presented in work by Eagle et al., [11, 12].

Table 1: Goodness of fit for Lesson 3 tutor performance.

Model	RMSE	Accuracy
SBKT	0.399	0.765
BFIDW-L3	0.368	0.806
BFIDW-L1	0.4	0.766
BFIDW-L2	0.394	0.774
BFIDW-L12	0.389	0.778
PrIDW-L12	0.38	0.791

We then 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 [2]. In order to control for the false discovery rate, we adjusted for multiple comparisons in the coefficient significance tests [5].

Finally, we employed the standard BKT model for lesson 3, the best fitting IDWs from each of the three lessons, and the various sets of predictor variables to generate 5 new IDW BKT models for Lesson 3, yielding a total of six BKT model variants displayed below. Analysis work was performed using R [15], Optimx [20], rlm [26], and lars [25].

Six BKT models calculated in this analysis for Lesson 3:

SBKT: Standard BKT non-individualized model with best-fitting group parameter estimates

BFIDW-L3: Individualized BKT model with best-fitting IDWs for Lesson 3

BFIDW-L1: Individualized BKT model with best-fitting IDWs for KCs in Lesson 1

BFIDW-L2: Individualized BKT model with best-fitting IDWs for KCs in Lesson 2

BFIDW-L12: Individualized BKT model with best-fitting IDWs for KCs in both Lessons 1 & 2

PrIDW-L12: Individualized BKT with predicted IDWs from reading and from Lesson 1 and Lesson 2 tutor performance features.

5 RESULTS AND DISCUSSION

Table 2 displays the overall fit to students‘ Lesson 3 tutor performance of the six models. 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).

Best-fitting IDWs for Lesson 3. The RMSE for the SBKT model with best fitting Lesson 3 parameter estimates, but no individualization is 0.399, as displayed in row 1. The remaining five rows display the five individualized models. BFIDW-L3 in row 2 employs best-fitting IDWs derived from the lesson 3 data. This model necessarily yields the best fit; it improves the goodness of fit by 7.8% over the SBKT model, reducing RMSE from 0.399 to 0.368.

Direct transfer of IDWs from Lessons 1 and 2. The next 3 rows display goodness of fit when the best fitting IDWs from Lesson 1, from Lesson 2, and from Lessons 1 & 2 combined, are employed directly in modeling Lesson 3 performance. As can be seen, BFIDW-L1, with IDWs from Lesson 1, and BFIDW-L2 with IDWs from Lesson 2 have little impact on the overall goodness of fit compared to SBKT, changing RMSE -0.03% and 1.6% respectively. BFIDW-L12 with refitted IDWs for the 53 KCs in both lessons has a slightly larger effect, improving on the SBKT fit by 3.2% reducing it to 0.394.

Predicted IDWs based on reading and Lessons 1 and 2 performance. The last row in the table displays RMSE for the PrIDW-L12 model in which reading measures from both texts and tutor performance measures from lessons 1 and 2 are employed to predict Lesson 3 IDWs. This model reduces RMSE to 0.380; it is about 60% as successful as the best-fitting BFIDW-L3 in reducing RMSE (and twice as successful as BFIDW-12).

Individualization and Mastery. Small differences in model fits can have large effects on the amount of practice assigned to students [11, 12, 17]. Following [11, 12], we calculated the approximate amount of practice that would be necessary for students to reach mastery under each of the six models in Table 2, and found general agreement among the five IDW models compared to the standard SBKT model. On average 51 students would have needed less practice under any of the 5 IDW models than under the SBKT model (range 46-57) and on average they would have required 54 fewer practice opportunities across all the lesson-3 KCs (range 42-64). On average 29 students would have needed more practice (range 22-30) and they would have needed an average of 23 more opportunities across all KCs (range 18-23). We take BFIDW-L3 (with best fitting Lesson-3 IDWs) as the gold standard in this comparison, and while the PrIDW-L3 model fits the lesson 3 data better than BFIDW-L12, the latter model agrees slightly better with BFIDW-L3 than does PrIDW-L3 (94% vs 91%). More work is needed to understand the relationship between model fit and mastery recommendations, but the general agreement between the IDW models suggests that a variety of evidenced-based IDW sets can improve efficiency in guiding students to mastery, compared to the SBKT model.

5.1 IDW Predictive Models

Table 3 displays the coefficients for each of the predictors in the regression models for each of the four Lesson 3 IDWs. As in [11], Lasso was used to identify the best 12 predictors for each of the four IDWs. The predictors that enter reliably into the four robust regression models are highlighted with asterisks.

The predictors that enter into the four models are rather eclectic. Reading time factors from the first text are among the top 12 predictors in three of the four IDWs models, as are reading time factors for the second text. The first text is on a different topic (gene interaction) than Lesson 3 (gene regulation). This suggests the reading time factors may be tapping learning strategy rather than the specific knowledge acquired.

Among the tutor performance measures in Table 3, slightly more came from Lesson 2 than Lesson 1, 25 vs. 15, but the difference is not significant. Whereas Lesson 1 and Lesson 3 employ related reasoning strategies — “forward“ process modeling rather than “backward“ abduction, Lessons 2 and 3 are closer in time; both of these relationships may contribute to predictive effectiveness, with perhaps a slight advantage for recency.

Table 2: Coefficient Summary Table

Pred.	wL0	wT	wG	wS
(Inter.)	1.012***	0.866***	0.242	0.306*
RT	T1F1[1]0.63	T2F3 0.043	T1F1 -0.034	T1F4 -0.034*
RT	T1F3 -0.066		T1F4 0.060	T2F1 -0.025
RT			T2F3 -0.039	
RT			T2F1 -0.017	
Pg re.				
Pg dist.				
wL0			L1 0.106	
wT		L2 0.171	L2 -0.080	
wG	L1[2]0.095		L2 0.034	L2 0.026
wS	L1 -0.235	L1 -0.433***		L2 0.214
	L2 -0.239			
Min. pLn				
Mast. KC		L2 -0.006***		
Err Rate.	L2 -0.411	L2 0.068		
Mean RT	L2 0.010	L2 0.016		
Help Av	L1 -2.996	L1 -1.773		L1 1.036
				L2 1.714*
Bug-LP			L2 15.672	
Bug-SP		L2 -5.514	L1 9.728	L2 -4.978
Hint-LP				
Hint-SP				
Kn-LP	L2 -0.726	L2 -0.275		L2 -0.616
Kn-SP	L2 -1.869***		L1 1.287	L2 0.386
				L1 0.791
Gaming	L1 -0.107	L2 -0.851	L2 0.534	
SP-NotG				
Off-Task	L1 -1.766	L1 -4.94**	L1 2.847	L1 2.378
				L2 -2.624*
LP-NotOT		L2 0.033		
RMSE	0.16	0.157	0.192	0.139

(* < 0.10, ** < 0.05, *** < 0.01)
[1] T1F1 = Topic 1 (gene interaction), Factor 1
[2] L1 = Lesson 1 wS (slip IDW)

The 19 total tutor performance variables fall into four broad types: the 4 IDWs, two BKT measures of student knowledge at the end of each lesson, two raw measures of performance, error rate and mean response time, and finally, the 11 “metacognitive“ measures, including use of help, response time in specific contexts, gaming and off-task behaviors. None of these four categories emerges as a stronger predictor than the others. Overall, each of the 19 variables enters into an average of 2.1 models, and the average number of models for the variables within any of the four categories does not depart much from this mean. Perhaps most surprisingly, the Lesson 1 and Lesson 2 IDWs are not especially strong predictors of Lesson 3 IDWs. Lesson 1 wL0 is among the top 12 predictors for just one model, Lesson 2 wT appears twice in Table 3, Lesson 1 or 2 wG appears three times, and Lesson 1 or 2 wS appears four times. The average number of models in which these variables appear, 2.5, is not much different from the overall average of 2.1.

Finally, among the 11 metacognitive features, Lesson 1 off-task behavior is perhaps the strongest predictor of Lesson 3 IDWs; it appears among the top 12 variables in all four models, and is significant in one of the models.

6 CONCLUSION

This study examines methods for predicting individual difference weights for students in BKT learning parameters (intercept and rate) and performance (guess and slip) for the third lesson in a Cognitive Tutor curriculum. This is an important issue because integrating IDWs into an intelligent tutor lesson is easier if the IDWs can be assigned before the student starts working in the lesson. We evaluate the different estimated IDWs by examining how well they fit student performance in Lesson 3, compared to (1) standard SBKT with no IDWs, and (2) a model with best-fitting weights for Lesson 3.

We find that directly applying the best-fitting IDWs from either of two prior lessons in the curriculum, or from both lessons combined, does not appreciably improve goodness of fit for Lesson 3, compared to the SBKT model. In contrast, estimating lesson-3 IDWs from measures of students‘ prior reading performance, and performance in the two prior tutor lessons, is more successful; it is 60% as successful as the best-fitting Lesson-3 IDW model in improving the goodness of fit compared to the SBKT model.

Several secondary conclusions emerge. First, a prior study [12] obtained very similar success in predicting IDWs based on reading performance, pretest performance and a smaller set of tutor performance measures. This study demonstrates that IDWs can be successful predicted without including pretest measures. This is potentially important since pretests may not be available in online learning environments. Second, among reading time measures and a wide range of tutor performance measures, no category of measures emerged as an especially strong predictor of Lesson 3 IDWs; instead it appears that predictive success depends on a broad range of predictor variables. Finally, reading time measures prove to be useful predictors of students‘ problem-solving behaviors in a subsequent tutor lesson, including reading time measures for text on a topic unrelated to that tutor lesson. This suggests that the reading time measures may reflect knowledge-acquisition strategies, as well as any knowledge acquired.

REFERENCES

[1] Vincent Aleven, Ido Roll, Bruce M McLaren, and Kenneth R Koedinger. 2016. Help helps, but only so much: Research on help seeking with intelligent tutoring systems. *International Journal of Artificial Intelligence in Education* 26, 1 (2016), 205–223.

[2] Robert Andersen. 2008. *Modern Methods for Robust Regression*. SAGE Publications, Inc. https://doi.org/10.4135/9781412985109

[3] Ryan SJd Baker. 2007. Modeling and understanding students' off-task behavior in intelligent tutoring systems. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1059–1068.

[4] Ryan SJ Baker, Albert T Corbett, and Vincent Aleven. 2008. More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In *International Conference on Intelligent Tutoring Systems*. Springer, 406–415.

[5] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)* (1995), 289–300.

[6] Cristina Conati, Abigail Gertner, and Kurt Vanlehn. 2002. *User Modeling and User-Adapted Interaction* 12, 4 (2002), 371–417. https://doi.org/10.1023/a:1021258506583

[7] Albert Corbett, Linda Kauffman, Ben Maclaren, Angela Wagner, and Elizabeth Jones. 2010. A Cognitive Tutor for Genetics Problem Solving: Learning Gains and Student Modeling. *Journal of Educational Computing Research* 42, 2 (feb 2010), 219–239. https://doi.org/10.2190/ec.42.2.e

[8] Albert T. Corbett and John R. Anderson. 1995. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modelling and User-Adapted Interaction* 4, 4 (1995), 253–278. https://doi.org/10.1007/bf01099821

[9] Ryan SJ d Baker, Albert T Corbett, Ido Roll, and Kenneth R Koedinger. 2008. Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction* 18, 3 (2008), 287–314.

[10] Ryan S. J. d. Baker, Albert T. Corbett, and Sujith M. Gowda. 2013. Generalizing automated detection of the robustness of student learning in an intelligent tutor for genetics. *Journal of Educational Psychology* 105, 4 (2013), 946–956. https://doi.org/10.1037/a0033216

[11] Michael Eagle, Albert Corbett, John Stamper, Bruce M McLaren, Ryan Baker, Angela Wagner, Benjamin MacLaren, and Aaron Mitchell. 2016. Predicting Individual Differences for Learner Modeling in Intelligent Tutors from Previous Learner Activities. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*. ACM, 55–63.

[12] Michael Eagle, Albert Corbett, John Stamper, Bruce M. McLaren, Angela Wagner, Benjamin MacLaren, and Aaron Mitchell. 2016. Estimating Individual Differences for Student Modeling in Intelligent Tutors from Reading and Pretest Data. In *Intelligent Tutoring Systems*. Springer International Publishing, 133–143. https://doi.org/10.1007/978-3-319-39583-8_13

[13] Rajaram Ganeshan, W. Lewis Johnson, Erin Shaw, and Beverly P. Wood. 2000. Tutoring Diagnostic Problem Solving. In *Intelligent Tutoring Systems*. Springer Berlin Heidelberg, 33–42. https://doi.org/10.1007/3-540-45108-0_7

[14] Yue Gong, Joseph E. Beck, and Neil T. Heffernan. 2010. Comparing Knowledge Tracing and Performance Factor Analysis by Using Multiple Model Fitting Procedures. In *Intelligent Tutoring Systems*. Springer Berlin Heidelberg, 35–44. https://doi.org/10.1007/978-3-642-13388-6_8

[15] Ross Ihaka and Robert Gentleman. 1996. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* 5, 3 (sep 1996), 299–314. https://doi.org/10.1080/10618600.1996.10474713

[16] Kenneth R. Koedinger, Albert T. Corbett, and Charles Perfetti. 2012. The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive Science* 36, 5 (apr 2012), 757–798. https://doi.org/10.1111/j.1551-6709.2012.01245.x

[17] Kenneth R. Koedinger, John C. Stamper, Elizabeth A. McLaughlin, and Tristan Nixon. 2013. Using Data-Driven Discovery of Better Student Models to Improve Student Learning. In *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 421–430. https://doi.org/10.1007/978-3-642-39112-5_43

[18] Jung In Lee and Emma Brunskill. 2012. The Impact on Individualizing Student Models on Necessary Practice Opportunities. *International Educational Data Mining Society* (2012).

[19] Michael Mayo and Antonija Mitrovic. 1999. Estimating Problem Value in an Intelligent Tutoring System Using Bayesian Networks. In *Advanced Topics in Artificial Intelligence*. Springer Berlin Heidelberg, 472–473. https://doi.org/10.1007/3-540-46695-9_42

[20] John C. Nash and Ravi Varadhan. 2011. Unifying Optimization Algorithms to Aid Software System Users:optimxforR. *Journal of Statistical Software* 43, 9 (2011). https://doi.org/10.18637/jss.v043.i09

[21] Zachary Pardos, Yoav Bergner, Daniel Seaton, and David Pritchard. 2013. Adapting bayesian knowledge tracing to a massive open online course in edx. In *Educational Data Mining 2013*.

[22] Zachary A. Pardos and Neil T. Heffernan. 2010. Modeling Individualization in a Bayesian Networks Implementation of Knowledge Tracing. In *User Modeling, Adaptation, and Personalization*. Springer Berlin Heidelberg, 255–266. https://doi.org/10.1007/978-3-642-13470-8_24

[23] Steve Ritter, Michael Yudelson, Stephen E Fancsali, and Susan R Berman. 2016. How mastery learning works at scale. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. ACM, 71–79.

[24] Valerie J. Shute. 1995. SMART: Student modeling approach for responsive tutoring. *User Modelling and User-Adapted Interaction* 5, 1 (1995), 1–44. https://doi.org/10.1007/bf01101800

[25] Robert Tibshirani, Iain Johnstone, Trevor Hastie, and Bradley Efron. 2004. Least angle regression. *The Annals of Statistics* 32, 2 (apr 2004), 407–499. https://doi.org/10.1214/009053604000000067

[26] W. N. Venables and B. D. Ripley. 1999. *Modern Applied Statistics with S-PLUS*. Springer New York. https://doi.org/10.1007/978-1-4757-3121-7

[27] Michael V. Yudelson, Kenneth R. Koedinger, and Geoffrey J. Gordon. 2013. Individualized Bayesian Knowledge Tracing Models. In *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 171–180.