Predicting Student Aptitude Using Performance History

Anthony F. Botelho
Worcester Polytechnic Institute
100 Institute Rd.
Worcester, MA 01609-2280
abotelho@wpi.edu

Seth A. Adjei
Worcester Polytechnic Institute
100 Institute Rd.
Worcester, MA 01609-2280
saadjei@wpi.edu

Hao Wan
Worcester Polytechnic Institute
100 Institute Rd.
Worcester, MA 01609-2280
hale@wpi.edu

Neil T. Heffernan
Worcester Polytechnic Institute
100 Institute Rd.
Worcester, MA 01609-2280
nth@wpi.edu

ABSTRACT

The use of intelligent tutoring systems in classrooms provides instructors and teachers with the opportunity to observe and monitor student performance as new skills and concepts are introduced. Many such tutoring systems currently in use provide a wealth of information pertaining to student learning over long periods of time. It is important to emphasize the need to focus on utilizing this information for the benefit of both instructors and students. Providing meaningful representations of student performance can indicate levels of knowledge and understanding that can alert instructors to potential struggling students in order to provide aid where it is needed; it is the goal of many researchers to even provide such indication preemptively in order to intervene before students become frustrated when attempting new skills. The goal of this paper is to utilize student performance history to provide a means of quantizing student aptitude, defined here as the speed at which a student learns, and then using this measurement to predict the speed at which each student will learn the next skill before beginning. Using 21 observable skills (those skills containing at least 10 unique students of which at least half reach mastery status within that skill), we compare three methods of predicting aptitude to majority class predictions at both an overall and individual skill level. Our results illustrate how our three proposed methods exhibit different strengths in predicting student aptitude when compared to majority class, and may be used to direct attention to a struggling student before attempting a new skill.

Keywords

Aptitude, Student Knowledge, Intelligent Tutoring Systems, Mastery Speed

1. INTRODUCTION

Many instructors rely on intelligent tutoring systems (ITS) as a means of extending student learning outside the classroom. Many such systems, such as the ASSISTments system used in this work, provide a wealth of student performance data that is often underutilized. Much of the field of education data analytics has focused its attention to using only some of this data to predict aspects of student knowledge in the form of next problem correctness. While many systems have shown success in this area, such information is only useful to instructors in a short time-span as students are completing assignments. Furthermore, many of the methods commonly used, such as Corbett and Anderson's Standard Knowledge Tracing (KT) model [5] use estimates of student attributes in order to make these predictions but stem from calculations of training data rather from actual student performance history. Developing models that in addition to constructing reliably accurate predictions or interpretations of student performance, also include definitive observable estimates of student attributes could provide meaningful information to instructors over a much longer period of time.

The use of student performance in terms of the correctness of responses is commonly used to formulate predictive models. It is important, however, to make distinctions as to what attribute of student learning is truly being modeled. Many terms may be used to describe the learning process, but may not all be truly synonymous in describing different aspects of learning. Particularly, the use of student knowledge and aptitude, while perhaps both pertaining to student performance, describe vastly different concepts. The term knowledge in most research pertains to domain knowledge, that is, how much information a student knows about a particular topic or in a particular skill. It is for this reason, that knowledge is often represented in terms of response correctness within a particular skill or set of skills. Aptitude, however, refers to the speed, often measured in number of problems or opportunities, at which a student learns. Essentially, in the case of tutoring systems, this concept can also be described as mastery speed, or the number of opportunities to reach mastery. Many tutoring systems, such as the ASSISTments system, indicate such mastery based on a predetermined number of consecutive correct answers; a student reaches mastery status in ASSISTments, for example, after three consecutive correct responses. Both knowledge and aptitude are common representations of student performance, and we believe that these two concepts can be observed independently. Furthermore, it can also be stated that knowledge and aptitude are potential traits pertaining to both students and skills; for the purpose of our research, we observe these concepts as student attributes and have developed methodologies to represent them as such.

There is often a large amount of ambiguity when describing knowledge and aptitude. We believe it is important to emphasize what it is we are attempting to observe, predict, and model with this work. It is our intention to look at student learning rates, referenced throughout this paper as aptitude, and attempt to predict changes in student learning in order to better identify students who are likely to exhibit poor performance on a new skill.

Such a measure of aptitude in prerequisite skills has shown to be successful in predicting initial knowledge on a subsequent skill [3], illustrating that the two concepts are related. From that work, however, it is unclear as to whether student aptitude is transitive across skills. In this work, therefore, we strive to answer the following research questions:

1. Do students exhibit similar degrees of aptitude across skills?

2. Are changes in student aptitude across skills predictable?

3. Can a student's aptitude in previous skills be used to construct a reliable prediction of mastery speed in a new skill before it is begun?

The next section will describe a background for our work, followed by a description of the dataset used in our experiments. Section 4 describes our methodology and each of our three experiments. Section 5 discusses our results followed by conclusions in Section 6. Lastly, our contributions are described in Section 7 followed by intended future work in Section 8.

2. BACKGROUND

The use of predictive models for use in intelligent tutoring systems have the potential to supply teachers with information pertaining to various student attributes. Some systems, however, provide means of prediction, but can lead to problems if further student representations are derived from them.

An example such a case is in the case of knowledge tracing. This method is widely used and studied due to its predictive accuracy. It utilizes four parameters to represent student attributes of prior domain knowledge, learning rate (an adaption of our definition of aptitude), guess rate (the probability of a student answering correctly when in an unlearned state), and slip rate (the probability of a student answering incorrectly when in a learned state). While these learned parameters are observable for a group of students, attempting to use them leads to problems of identifiability [2]. Such a problem would perhaps be avoidable if the parameters were instead learned from actual student performance history.

Other models [7][6][8] attempt to observe and improve upon the KT model in terms of student representation. Another problem with such a model, however, is that it only views short-term performance. Modeling student knowledge is important for teachers, but observing long-term performance over multiple skills can undoubtedly provide more meaningful information. Students and teachers would benefit more from accurate predictions of future performance based on individual history.

It is this type of model that we aim to work toward; we wish to develop a model that can provide reliable predictive accuracy while also modeling student attributes that can provide teachers with more meaningful information. With such information available, more precise aid can be administered to each student earlier, before problems of frustration and wheel spinning [1] occur.

We are not the first to observe and incorporate student performance history into a predictive model. Others in the field of educational data analytics have introduced works pertaining the study and utilization of prerequisite structures defined in intelligent tutoring systems. Zapata-Rivera and Greer [10] have illustrated interesting ways for using bayesian networks to model skill relationships. Similarly, others [4] have dedicated research to utilizing skill relationships for the development of student models

3. DATASET

The dataset[1] used in this work is comprised of real-world data from PLACEments test data reported from the ASSISTments tutoring system. PLACEments tests are administered by teachers to students, in which students are given questions within their grade level for the purpose of assessing knowledge. Skills are presented in a linear fashion, assigning further skills if students experience lower performance in a particular skill. The resulting dataset provides data to construct a student performance history through skills that were either learned or relearned over the course of the testing. Such information is useful in this research to view the speed at which each student is able to learn each skill.

From the testing results, data pertaining to 21 unique observable skills was extracted. Here, we define a skill as observable if it contains data from more than 10 unique students, and no less than half of the students must have mastered the skill. As ASSISTments defines mastery in terms of 3 consecutive correct answers, answering fewer than 3 or quitting before reaching this threshold leaves the student in an unmastered state. We omit skills that contain large percentages (>50%) of such students, as it may indicate a problem with either the system or the skills assigned by the instructor. Only students within these 21 skills were observed, but the entire original dataset was used to build each observed student's performance history.

The resulting dataset contains 644 unique students across 1437 data rows. A performance history is created for each student comprised of all skills previously attempted.

[1]The original dataset can be found at the following link: http://bit.ly/1DVbHdB.

4. METHODOLOGY

In order to address the research questions described in an earlier section, our methodology was divided into three experiments. The purpose of these experiments was to utilize only a small number of student attributes in order to illustrate the effectiveness of using previous observed mastery speed in order to predict aptitude in a new skill; we acknowledge that using other attributes including time between skills, a measurement of skill difficulty, and perhaps the number of previously mastered skills may provide more accurate predictions, but are not considered in this work in an attempt to isolate mastery speed as the contributing factor of our results. Before describing our experiments, our method of measuring, representing, and quantizing aptitude will be described.

We used a simple binning method implemented in similar research [9] to place students into one of five categories based on mastery speed in order to represent different levels of aptitude. As aptitude is an independent concept of domain knowledge, a student's entire recorded performance history, regardless of the prerequisite structure, was used to categorize each student. For means of clarification, the terms of "binning" and "categorizing" will be used interchangeably to describe the methodology. Observing each student's performance over several skills, we used an exponential moving average over the mastery speed of each skill ordered from oldest to most recent.

$$A_t = ((1 - \alpha) * A_{t-1}) + (\alpha * V_t) \quad (1)$$

Equation 1 displays the formula for this method, in which the variable "A" corresponds to the moving average, while "V" is the new observed value of mastery speed. This method averages values by weighing new information higher, in an attempt to better utilize the most recent information for student binning while still considering the entire performance history. For our implementation, we used a value of 0.3 for alpha.

Table 1: The ranges of mastery speed represented by each bin with corresponding the quantized aptitude value.

Bin Number	Mastery Speed	Quantized Value
1	$3 \leq \text{ms} \leq 4$	1
2	$4 < \text{ms} < 8$	0.75
3	$8 \leq \text{ms}$	0.5
4	DNF, pcor $\geq .667$	0.25
5	DNF, pcor $< .667$	0

Once an average mastery speed value is calculated for each student based on performance history, each student is then placed into a bin corresponding to different ranges of mastery speed. These bins, described in Table 1, illustrate estimated high levels of aptitude toward bin 1 and low levels of aptitude toward bin 5. Bins 4 and 5 contain students that did not finish (DNF) at least one previous skill, and are instead split based on the average percent correctness (pcor) across all previous skills. As the dataset was taken from real data collected from the ASSISTments tutoring system, and such skills are assigned by the instructor, it is difficult to determine whether the reason for not finishing a skill is due to lack of knowledge, boredom, or any other exterior factors. As such, we attempt to represent these students as having a lower level of aptitude than students who finish all previous skills.

With each student placed into one of the five categories, each bin is then given a value between 0 and 1 representing that bin's quantized level of aptitude. The reasoning behind this quantization is for more recognizable comparisons to be made, as common representations of predictive accuracy, such as RMSE, rely on this value range. For each of our experiments we use a five-fold cross validation using 80% of our dataset as training data to predict the remaining 20%.

The quantized values are chosen arbitrarily to represent the learning rate that is intended to be represented by each bin. We would like to acknowledge that this measurement itself can be greatly improved, but forms an adequate basis for which to derive the results described in a later section. Further research utilizing other student aspects such as average knowledge in terms of percent correctness may lead to more precise values of representation.

As learning rates refer to a continuous measurement, this method of binning is used to convert those values into discrete categories representing a small range of aptitude speeds.

4.1 Experiment 1

The first experiment attempts to use a very simple prediction methodology to address our first research question. In this method, we simply make a prediction that each student will exhibit a similar level of aptitude in a new skill as in previous skills. If such a claim holds, this method should provide a more accurate prediction than the majority class.

For our experiments, we compare our proposed methodology to the majority class (MC), defined here as the overall average aptitude value. For example, if it is found that every student is categorized under bin 2, the majority class prediction for every student would be 0.75. The error of this majority class is then used as a baseline.

Our first prediction method, referenced as Same Bin Prediction (SBP) in our results section, simply uses the average mastery speed of each student's performance history to determine in which bin to place each student. The method then simply uses that bin's quantized value as a prediction for the new skill. Both the SBP and majority class are then compared to each student's actual mastery speed, expressed as a quantized bin value, to determine both error rates.

This prediction method was used at two levels of observation to view results at different levels of granularity. The first level made predictions over all skills in the dataset. As each prediction is made at a student level, this mainly effected the calculation of the majority class prediction. The second level was an analysis performed at the skill level, described in our results as SBP by skill (SBPbySkill). Again, as our method of prediction is always at the student level, this alteration effects only the majority class predictions, which we hypothesize to be more accurate when viewed at the skill level.

4.2 Experiment 2

Our second experiment attempts to make predictions again using each student's performance history, but by also taking into account changes in aptitude. Our first experiment assumes that most students will exhibit the same level of aptitude in a new skill as in previous skills. This experiment takes into account the realization that differences in skill difficulty may cause fluctuations in our aptitude measurements. Aside from skill difficulty, each individual student may find particular skills harder than others, again effecting that student's mastery speed. Considering these points, our second method, referenced as Transitioning Bin Prediction (TBP) in our results section, attempts to calculate the degree in which each skill effects the aptitude estimation of each bin.

The TBP method of predictions builds off of the previous SBP prediction using a calculated offset. Again, the idea is that some skills may have different effects on each level of aptitude; one skill may be difficult, causing students from each bin to need more opportunities to master than normal, while other skills may only have such an effect on particular bins.

An offset value is calculated for each bin using the training set. For each bin, an average difference value is calculated based upon to which bin each student transitioned. For example, if half the students in bin 1 (value = 1) remained in that bin for the new skill, while half transitioned to bin 2 (value = 0.75), an offset value of -0.125 would be applied to all predictions of bin 1. A negative offset indicates that many students more opportunities to master than normal, while a positive offset indicates that many students required fewer opportunities to master than normal.

Using the original SBP prediction as a basis, the offset value corresponding to each student's bin is applied to the prediction. Similar to experiment 1, this method is applied at the same 2 levels of granularity. As this method relies more on the use of the training set, each level of observation should effect both the TBP and majority class predictive accuracy in a similar manner. The first level again compares TBP and majority class predictions across all skills, using 80% of the students from each skill in the training set to predict the remaining 20%. The second level of observation, referenced as TBP by Skill (TBPbySkill) in our results section, compares the two methods at the skill level.

4.3 Experiment 3

The third and last experiment utilizes the method developed in the second experiment and attempts to expand upon it. While the previous method attempts to make use of common changes in aptitude from skill to skill, offsetting the prediction values will inherently bias predictions toward the center of our aptitude spectrum. In order to ensure that our prediction values are utilizing the entire 0 to 1 prediction range, a third method, referenced as Normalized TBP (NTBP) in our results section, is also compared to the majority class.

The NTBP is a simple modification to the previous method. Using the TBP prediction as a basis, as well as the minimum and maximum TBP value of the training set, the value is normalized to values between 0 and 1. As the value ranges are simply scaled, the prediction is still meant to represent changes in aptitude across skills.

$$NTBP_k = \frac{(TBP_k - min(TBP_{tr}))}{(max(TBP_{tr}) - min(TBP_{tr}))} \tag{2}$$

Again, this method of prediction is compared to majority class predictions at the same two levels of granularity. Similar to TBP, the prediction value of NTBP relies on the training set to calculate its offset value and for normalization. Following the same naming convention as the previous two experiments, the NTBP by Skill (NTBPbySkill), compares predictive accuracy at the skill level.

5. RESULTS

The results of each experiment are illustrated in the tables below. The predictive accuracy of each method are expressed in two metrics to exemplify the strengths and weaknesses of each comparative method. The first metric, RMSE, is used to illustrate the overall error when comparing the actual quantized mastery speed value of each student to each prediction value. A lower value of RMSE indicates a more accurate model. The second metric utilizes a simple binary "right or wrong" measurement of correctness. This metric, referenced simply as percent correct, represents the number of times each method correctly predicted into which bin each student would be categorized for the new skill. More specifically, this metric is calculated as the number of correct predictions of each bin divided by the number of total predictions made for that bin. For example, if a method predicts that four students will be placed into bin 1, while only 2 are actually placed there, the percent correct would be 0.5. A higher value of this metric indicates a more accurate and dependable prediction model. Percent correctness gives an indication of method dependability, but it is important to consider both metrics when comparing such models, as a model with superior RMSE may not be as accurate in terms of percent correctness as well as the reverse.

The results of each method are depicted by bin, that is, in terms of how well each is able to predict the level of aptitude represented by each bin. As it is the intention of this work to provide more meaningful information to instructors, such a distribution indicates where methodologies are most successful. Observing the results in Table 2, for example, shows that Majority Class was successful in predicting students that were placed in the second bin, but was less successful in predicting students of lower aptitude in bins 4 and 5.

5.1 Results Across All Skills

Table 2 displays the RMSE values of each method when predictions are made across all skills. As a majority of the dataset contained students displaying higher aptitude, the majority class was successful in predicting these students' level of aptitude before beginning a new skill. Ignoring the first two bins, however, the NTBP method outperforms all other methods in terms of predictive accuracy. Identifying students who may be prone to problems of wheel spinning has been a challenge in previous works [1], and the methods described here provide a higher degree of reliability in doing

Table 2: The RMSE of each method across all skills divided by bin.

Bin of New Skill	Majority Class	SBP	TBP	NTBP
1	0.230	0.498	0.262	0.461
2	0.022	0.356	0.023	0.241
3	0.271	0.362	0.241	0.206
4	0.518	0.526	0.489	0.359
5	0.767	0.659	0.740	0.618

Table 3: Percent correctness of each method across all skills.

Bin of New Skill	Majority Class	SBP	TBP	NTBP
1	0.000	0.479	0.000	0.439
2	0.249	0.245	0.249	0.270
3	0.000	0.251	0.000	0.221
4	0.000	0.029	0.000	0.063
5	0.000	0.041	0.000	0.043

so; using estimates of aptitude can direct the attention of teachers to such students before problems occur.

The results in Table 3 illustrate the percent correctness of each method. The distribution here illustrates how dependable each method is as a predictive model. As shown, the majority class and TBP have difficulty expressing larger changes in aptitude and therefore fall into a biased averaging across all skills. This, of course is opposed to the NTBP method that takes the same distribution of predictions from TBP, but normalizes it to span the entire 0 to 1 prediction range.

Another interesting observation from this table is the results of the SBP method. This method simply uses the average mastery speed across all previous skills and predicts that the student will exhibit the same level of aptitude. The distribution indicates that students in the first bin are the most likely students to remain in that bin for a new skill. This observation is not necessarily surprising, but is a positive indicator that aptitude is transitive across skills to some degree.

5.2 Results of the Skill Level Analysis

Observing skills at different levels of granularity provides an indication of a prediction method's scalability. At a skill level, it is expected that certain methods will perform with a higher degree of accuracy. This is, of course, especially the case with the majority class, as it has a much narrower focus of each skill's baseline as opposed to the baseline across all skills. It is unlikely that each skill has differing levels of difficulty associated with it, and viewing results at this level inherently include some of that information; this is particularly the case whenever a training set is utilized as it builds off of information pertaining to previous student performance in the observed skill.

Table 4 contains the RMSE results, again divided by each bin of the new skill, of each of the prediction methods. Similar to the results across all skills, the success of the majority class prediction extends primarily to students of higher aptitude. This alludes to a property of our dataset, in which a majority of students exhibit fast mastery speeds. Focussing on the lower three bins, NTBP again provides the most accurate predictions over the bins representing lower aptitude.

Table 4: Average RMSE of the skill level analysis divided by bin.

Bin of New Skill	Majority Class	SBP	TBP	NTBP
1	0.230	0.498	0.274	0.358
2	0.120	0.356	0.165	0.170
3	0.284	0.362	0.302	0.205
4	0.307	0.526	0.330	0.251
5	0.571	0.659	0.577	0.497

Table 5: Percent correctness at the skill level divided by bin.

Bin of New Skill	Majority Class	SBP	TBP	NTBP
1	0.709	0.479	0.590	0.500
2	0.280	0.245	0.262	0.268
3	0.102	0.251	0.171	0.200
4	0	0.029	0.091	0.129
5	0	0.041	0.333	0.333

The dependability of these methods is again further illustrated by the results in Table 5. The majority class was unsuccessful in predicting lower aptitude students when compared with the other methods. This truly illustrates the importance of representing the results using multiple metrics. Looking only at RMSE, majority class provides reliably accurate predictions, but the second metric of percent correctness reveals that it completely omitted students in bins 4 and 5 from its predictions. This is precisely the scenario that is important to avoid when providing teachers with estimates of any student attribute. While the NTBP method does not outperform the majority class in every bin, it models a greater range of students with comparable accuracy. The level of precision, whether focusing on higher or lower aptitude, can be gained from using different methods.

6. CONCLUSIONS

As it is the goal of the paper to provide instructors with information to better aid struggling students, larger focus should be given to predicting students in bins 3, 4, and 5. Identifying these students with greater accuracy can allow teachers to provide aid before a problem occurs.

It is important to note, however, that each method described in this work exhibited different strengths, including the simple majority class predictions. It is often for the benefit of both teachers and students that a model represent meaningful information beyond the provision of predictive accuracy. The SBP method, for example, while not excelling in any one category, illustrates tendencies of aptitude mobility. It is the hope of the teacher that once students reach higher levels of aptitude that they remain at that level for future skills. The SBP method provides a means of observing such information and is useful as such while perhaps being less successful in terms of making predictions.

In regard to the research questions proposed in an earlier section, each can be addressed from our methods and results. The first question asks if students exhibit similar degrees of aptitude across skills. The SBP method, as described earlier, addresses this question directly. Observing the results of Table 3 and Table 5, it can be seen that less than half of the students in each bin remain in that bin for the new skill. Several factors including skill difficulty can effect this, but it appears that only certain levels of aptitude, particularly high ones, display similar trends across skills.

The TBP and NTBP methods address our second research question pertaining to the predictability of changes in aptitude across skills. As both of these methods improve upon majority class accuracy in lower aptitude bins, it can be argued that such changes are predictable in many cases.

The third research question, pertaining to the reliability of predicting mastery speed in a new skill certainly seems to gain support from our results. This indication, however, does not come from a single method, but rather by observing each method for what it intends to represent. In many cases, each level of aptitude can be represented by one of our proposed methods, but no one method excels at predicting all levels of aptitude. This is an interesting observation, as it perhaps alludes to the fact that several factors impact changes in aptitude across skills, and therefore require different predictive models to emulate such factors.

7. CONTRIBUTIONS

This work introduces three methods of representing and predicting aptitude on a new skill using student performance history. We illustrate that such predictions can lead to better indications of students in danger of performing poorly or struggling on a new skill before that skill is begun. While no one method outperforms all others in predicting each level of aptitude, each also provides meaningful information to teachers.

We have shown that levels of aptitude change from skill to skill in a predictable manner. It is undoubtedly the case that providing more information regarding each skill such as difficulty or perhaps even utilizing prerequisite structure relationships could lead to more precise predictions.

The work in this paper also supports the use of student history as definitive estimates of student attributes. By omitting the use of any latent factors in our model, it attempts to avoid problems of identifiability, providing definitive sources of our information. If such information could be implemented into prediction models like KT, it could become a better overall student model.

Our research extends into the study of student attributes by making distinctions between knowledge and learning rate. While knowledge is often domain-specific, student aptitude has been shown to exhibit similar trends that exist independent of the skills observed. As this trait is domain independent, it is scalable not only to the skills observed, but to other domain areas as well.

8. FUTURE WORK

While we have proposed several methods of prediction, these representations of student aptitude are not necessarily formulated into a model. We have illustrated that each method exhibits different strengths in predicting levels of aptitude with differing degrees of accuracy and dependability. With these in mind, a more definitive model may be constructed.

As already discussed, our methods may also benefit from further skill and student information. Currently we utilize mastery speed as a basis of binning and prediction, but other aspects may also be beneficial to consider. Aspects such as response time, estimations of knowledge in terms of correctness, and other system-dependent information such as hint usage may also be considered. The use of prerequisite structures and inter-skill relationships can be used as well to formulate more precise predictions; knowing which skills exhibit stronger relationships could be used to create a more adaptive model that also takes these skill-level attributes into account.

Furthermore, the ability to predict aptitude may benefit other models like KT. Our quantization of student aptitude may be appropriated to calculate the learn parameter of KT, defined as the probability of a student transitioning from an unlearned to a learned state at each opportunity. Removing problems of identifiability from such a recognized model could also make it easier for systems to implement and test the use of such information sooner. Incorporating this and other similar factors could provide a wealth of information to the teachers that rely on dependable systems in and outside the classroom.

9. ACKNOWLEDGMENTS

We thank multiple NSF grants (ACI-1440753, DRL-1252297,DRL-1109483,DRL-1316736 & DRL-1031398), the US Dept of Ed's (IES R305A120125 & R305C100024 and GAANN), the ONR, and the Gates Foundation.

10. REFERENCES

[1] J. Beck and M. M. T. Rodrigo. Understanding wheel spinning in the context of affective factors. In *Intelligent Tutoring Systems*, pages 162–167. Springer International Publishing, 2014.

[2] J. E. Beck and K. min Chang. Identifiability: A fundamental problem of student modeling. In *User Modeling 2007*, pages 137–146. Springer Berlin Heidelberg, 2007.

[3] A. Botelho, H. Wan, and N. Heffernan. The prediction of student first response using prerequisite skills. In *Learning at Scale*, 2015.

[4] C. Carmona, E. Millán, J.-L. Pérez-de-la Cruz, M. Trella, and R. Conejo. Introducing prerequisite relations in a multi-layered bayesian student model. In *User Modeling 2005*, pages 347–356. Springer, 2005.

[5] A. Corbett and J. Anderson. Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modeling and User-Adapted Interaction*, 4(4):253–278, 1994.

[6] Z. A. Pardos and N. T. Heffernan. Modeling individualization in a bayesian networks

implementation of knowledge tracing. In *User Modeling, Adaptation, and Personalization*, pages 255–266. Springer Berlin Heidelberg, 2010.

[7] Z. A. Pardos and N. T. Heffernan. Navigating the parameter space of bayesian knowledge tracing models: Visualizations of the convergence of the expectation maximization algorithm. *EDM*, 2010:161–170, 2010.

[8] Y. Wang and N. Heffernan. Extending knowledge tracing to allow partial credit: Using continuous versus binary nodes. In *Artificial Intelligence in Education*, pages 181–188. Springer Berlin Heidelberg, 2013.

[9] X. Xiong, S. Li, and J. E. Beck. Will you get it right next week: Predict delayed performance in enhanced its mastery cycle. In *FLAIRS Conference*, 2013.

[10] J.-D. Zapata-Rivera and J. E. Greer. Interacting with inspectable bayesian student models. *International Journal of Artificial Intelligence in Education*, 14(2):127–163, 2004.