Predicting Future Learning Better Using Quantitative Analysis of Moment-by-Moment Learning

Arnon Hershkovitz
Teachers College, Columbia University
ah3096@columbia.edu

Ryan S.J.d. Baker
Teachers College, Columbia University
baker2@exchange.tc.columbia.edu

Sujith M. Gowda
Worcester Polytechnic Institute
sujithmg@wpi.edu

Albert T. Corbett
Carnegie Mellon University
corbett@cmu.edu

ABSTRACT

In recent years, student modeling has been extended from predicting future student performance on the skills being learned in a tutor to predicting a student's preparation for future learning (PFL). These methods have predicted PFL from a combination of features of students' behaviors related to meta-cognition. However, these models have achieved only moderately better performance at predicting PFL than traditional methods for latent knowledge estimation, such as Bayesian Knowledge Tracing. We propose an alternate paradigm for predicting PFL, using quantitative aspects of the moment-by-moment learning graph. This graph represents individual students' learning over time and is developed using a knowledge-estimation model which infers the degree of learning that occurs at specific moments rather than the student's knowledge state at those moments. As such, we analyze learning trajectories in a fine-grained fashion. This new paradigm achieves substantially better student-level cross-validated prediction of student's PFL than previous approaches. Particularly, we find that learning which is spread out over time, with multiple instances of significant improvement occurring with substantial gaps between them, is associated with more robust learning than either very steady learning or learning characterized by a single "eureka" moment or a single period of rapid improvement.

Keywords

Moment-by-moment learning graph, preparation for future learning, student modeling.

1. INTRODUCTION

In recent years, there has been increasing emphasis in learning sciences research on helping students develop robust understanding that supports a student in achieving preparation for future learning (PFL) (cf. [9,15,17,26]), with evidence suggesting that differences in the design of educational experiences can substantially impact PFL [11,28]. Multiple approaches have now been found to be successful at supporting PFL. For example, learning-by-teaching when implemented with the use of "teachable agents", computer characters that the student have to teach during the learning process, has been shown to support PFL [11,26,28]. Another approach shown to support PFL is the use of invention activities, during which students are asked to "invent" representation of a given problem (e.g., variance of a data set) [9,25-26].

Given the existence of methods that can support PFL, there is increasing potential to enhance individualization within computer-based learning environments to optimize not just learning of the material being taught (cf. [10,12,24]), but PFL as well. However, individualization of this nature depends on student models that can effectively infer PFL.

In the last two years, approaches that can infer PFL and other forms of robust learning have begun to emerge, but these approaches are still in their early stages, and are only modestly better than simply assessing student knowledge. In specific, models that leverage data on metacognitive and motivational aspects of student behavior (e.g., off-task, help-avoidance) have achieved cross-validated correlations about 0.05-0.1 higher than classical knowledge models (e.g. Bayesian Knowledge Tracing) to both PFL and transfer tests [4-5]. In addition, retention (another aspect of robust learning) has been effectively predicted using inferences of memory decay during periods of non-practice (e.g. forgetting; [16,29]).

In this paper, we propose an alternate method for predicting PFL more precisely than the meta-cognitive/motivational behavior approach proposed in [4]: using quantitative aspects of the Moment-by-Moment Learning Graph. Similar to the classic learning curve (cf. [19,22]), a Moment-by-Moment Learning Graph (MBMLG) represents the probability that learning has occurred at a specific moment [3], for a given student and a given Knowledge Component (KC)/skill, at each step of the learning process. These probabilities are calculated based on a machine learned model that smoothes probabilities calculated using the probability that the student has learned the skill up to the point of a specific step, and the probability of their future actions given the probability that they learned the skill at that problem step.

Earlier work (discussed in greater detail in Section 3.2) suggests that visual interpretations of the patterns of the MBMLG correlate to PFL [6]. This earlier work used human coders to interpret the visual characteristics of the MBMLG. In this work, we study whether an automated approach – based on quantitative analysis of features of the MBMLG – inspired by this earlier work can improve the prediction of PFL.

2. DATASET

We use attributes of the form of individual student's MBMLG to predict student preparation for future learning. We do so in a combined data set from three studies, in total comprising 181 undergraduate and high-school students who used an intelligent tutoring system to learn Genetics. The students enrolled in Genetics courses at Carnegie Mellon University, or in high school biology courses in Southwestern Pennsylvania.

Study 1 (College Undergraduates, Three-Factor Cross). 72 undergraduates enrolled in a Genetics course or in an Introductory Biology course at Carnegie Mellon University were recruited to participate in the study for pay, at a point in the semester where the tutor software was relevant to their classroom learning. The 72 students completed a total of 22,885 problem solving attempts across a total of 10,966 problem steps in the tutor.

Study 2 (College Undergraduates, Gene Interaction). 53 undergraduates enrolled in a Genetics course or in an Introductory Biology course at Carnegie Mellon University were recruited to participate in the study for pay, at a point in the semester where the tutor software was relevant to their classroom learning. The 53 students completed a total of 33,643 problem solving attempts across a total of 22,126 problem steps in the tutor.

Study 3 (High school students, Three-Factor Cross). 56 high school students who were enrolled in high school biology courses used the tutor. The students were recruited to participate in the study for pay through several methods, including advertisements in a regional newspaper and recruitment handouts distributed at two area high schools. The 56 students completed a total of 21,498 problem solving attempts across a total of 9,204 problem steps in the tutor.

2.1 Learning System and Learning Activity

The data used in this paper was drawn from student use of the Genetics Cognitive Tutor [14]. This tutor consists of 19 modules that support problem solving across a wide range of topics in genetics (Mendelian transmission, pedigree analysis, gene mapping, gene regulation and population genetics). Various subsets of the 19 modules have been piloted at 15 universities in North America.

This study focuses on two of these tutor modules. One employs a gene mapping technique called a Three-Factor Cross. The tutor interface for this reasoning task is displayed in Figure 1. In this technique two organisms are crossed (two fruit flies in the example) and the resulting distribution of offspring phenotypes is analyzed to infer the order of three genes on the chromosome and the relative distances between the three pairs of genes.

The other module, Gene Interaction and Epistasis, engages students in extending basic Mendelian transmission to two genes. In this task, displayed in Figure 2, students cross three true-breeding strains, perform intercrosses, and based on offspring phenotype frequencies, infer the genotypes of the true-breeding strains and each of the offspring phenotypes.

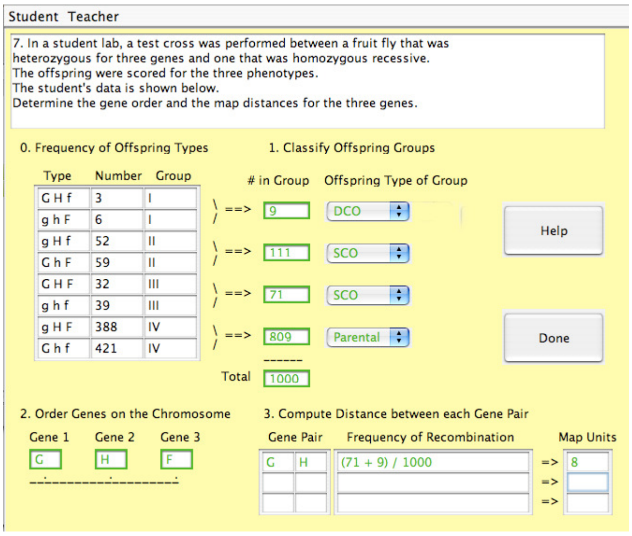


Figure 1. Screenshot from the Three-Factor Cross lesson of the Genetics Cognitive Tutor

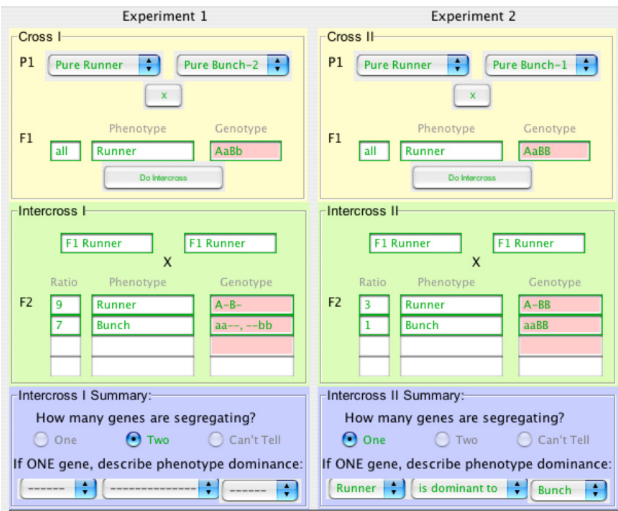


Figure 2. Screenshot from the Gene Interaction lesson of the Genetics Cognitive Tutor

2.2 Design

The studies were conducted in computer clusters at Carnegie Mellon University. All students attended study sessions on two consecutive days; in studies 1 and 2, each of these lasted 2 hours, while in study 3, each lasted 2.5 hours. All students engaged in Cognitive Tutor-supported activities for about one hour in each of the two sessions. In studies 1 and 3 all students completed standard Three-Factor Cross problems, as depicted in Figure 1, in both sessions, while in study 2 all students completed standard Gene Interaction problems, as depicted in Figure 2, in both sessions.

During the first session of each study, some students were assigned to complete other cognitive-tutor activities designed to support deeper understanding; however, no significant differences were found between conditions for PFL or any other robust learning measure (this is reported for study 1 in [14]), so in this analysis we collapse across the conditions and focus solely on student behavior and learning within the standard problem-solving Cognitive Tutor activities.

All students completed a problem-solving pre-test at the beginning of the first session, and a problem-solving post-test immediately following the Cognitive Tutor activities in the second session. Following the problem-solving post-test in the second session, students also completed a transfer test and preparation-for-future-learning (PFL) test. Finally, students in studies 1 and 2 returned a week later to complete a problem-solving retention test. Within this paper, we focus all analysis on the PFL test, as a particularly strong indicator that student learning is robust (cf. [9]). A PFL prediction model was built on a dataset combining the three studies.

2.3 PFL Test

This study examines student performance on preparation-for-future-learning problem-solving tests. By definition, the reasoning in each of these two tests is related to solving Three-Factor Cross or Gene Interaction problems, respectively, but is sufficiently more complicated that a student could not be expected invent a solution method by direct transfer, and certainly not in a short period of time. Consequently, each of the PFL tests incorporated instructional text on the required reasoning, which students read prior to problem solving. The PFL tests were designed in collaboration between biology experts and a cognitive scientist (the fourth author).

In the Three-Factor Cross studies, students were asked to solve parts of a four-factor cross problem; the PFL test presented a 2.5-page description of the reasoning in a four-factor cross experiment, then asked students to solve some elements of a four-factor cross problem: identifying the middle genes, identifying all the offspring groups with a crossover between two specific genes and finding the map distance between those two genes.

In the Gene Interaction study, students were asked to reason about gene regulation problems. In these problems, three genes, an operator, a structural gene and a regulatory gene, act together to control DNA transcription. The test presented a 1.5 page description of several gene regulatory systems, then asked student to reason about the impact of dominant and recessive alleles of the component genes on transcription.

PFL tests were completed by all students in the three studies, with an average percent correct of 0.89 (SD=0.15), 0.74 (SD=0.24), and 0.66 (SD=0.28) for Study 1, Study 2, and Study 3, respectively.

3. MOMENT-BY-MOMENT LEARNING GRAPH

3.1 Construction of the Graph

The construction of the Moment-By-Moment Learning Graph (MBMLG) is based on a three-phase process, which first infers moment-by-moment learning using data from the future, then infers the same construct without data from the future, and then integrates across inferences over time to create a graph.

The first step is to infer moment-by-moment learning using data from the future, based on an approach first proposed by [2]. To obtain this inference, a Bayesian Knowledge-Tracing (BKT) model [13] is used to calculate the probability that the student knows a specific skill at a specific time, based on the student's history of success on problems or problem steps involving that skill. The BKT model is updated every time the student responds to a problem step, based on the correctness of the response, allowing for an aggregate estimate of student knowledge over time.

Then, the estimation of student knowledge and the parameters of the BKT model are combined using Bayesian formulas (discussed in mathematical detail in [3]), to infer the probability that a student learned a skill or a concept at a specific step during the problem-solving process, by looking at the probability of future actions if the student had learned the skill at this point. This probability is referred to as P(J) (J stands for "just learned"). That is to say, instead of assessing the probability that a skill is known by the time the student reaches the Nth step that involves that skill, the model assesses the probability that the skill was learned between time N-1 and time N. At an intuitive level, high values of P(J) are seen when a student's performance shifts from being mostly incorrect to mostly correct, but precise values are obtained using current estimates of the probability the student knows the skill, along with model estimates of the probability of correct answers being due to guessing, and incorrect answers being due to slipping or carelessness. This model uses information on past, current, and future performance, to predict the probability that learning occurred during each step of the student's work within the computer-based learning environment.

Once these predictions have been obtained, a machine-learned model is built, using a set of features of student data (such as the recent history of help and errors on this skill, and time taken on the current and recent attempts) to predict P(J) values based on past and current information only. Within the work presented here, the same feature set as was used for the Cognitive Tutor in [3] was used. The list of features inputted into the machine learning algorithm is:

- Assessments of correctness:
  - Percent of all past problems that were wrong on this KC.
  - Total number of past problems that were wrong on this KC.
  - Number of last 5 problems that were wrong.
  - Number of last 8 problems that were wrong.
- Measurements of time:
  - Time taken (SD faster/slower than average across all students).
  - Time taken in last 3 actions (SD off average) Time taken in last 5 actions (SD off average)
  - Total time spent on this KC across all problems
  - Time since the current KC was last seen.
- Data on hint usage:
  - First response is a help request.
  - Bottom-out hint is used.
  - Number of last 8 problems that used the bottom-out hint.
  - Number of last 5 problems that included a help request.
  - Number of last 8 problems that included a help request.
- Other measurements:
  - Total problems attempted in the tutor so far.
  - Total practice opportunities on this KC so far.
  - Response is chosen from a list of answers (Multiple choice, etc).
  - Response is filled in (No list of answers available).

This model serves the joint purposes of enabling the model to be used for eventual intervention, and smoothing the sometimes extreme values of P(J) that can be obtained when the BKT model's parameters for guessing and slipping are low. The model used here is built using linear regression with M5' feature selection [30], in RapidMiner Version 4.6 [20]. To validate the generalizability of our models, 6-fold cross-validation at the student level was used (i.e., detectors are trained on five groups of students and tested on a sixth group of students). By cross-validating at this level, we increase confidence that detectors will be accurate for new students.

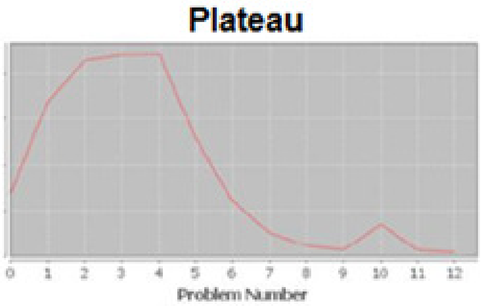
The goodness of the models was assessed using the Pearson correlation coefficient between the training labels of P(J) for each opportunity to learn each KC, and the values predicted for P(J) for the same opportunity by the machine-learned models. As both set of values are quantitative, and there is a one-to-one mapping between training labels and predicted values, linear correlation is a reasonable metric.

The P(J) model achieved solid correlations to the training labels under 6-fold student-level cross-validation, with values of 0.68 for Study 1 (college genetics 3-factor cross lesson; reported in [6]), 0.65 for Study 2 (college genetics gene-interaction lesson), and 0.48 for Study 3 (high school genetics 3-factor cross lesson). These values are moderately higher than those seen for P(J) models built for the Middle School Cognitive Tutor or ASSISTments, probably due to the more diverse collection of lessons used in these earlier studies (e.g. [3]). The difference in correlation between the college studies and the high-school study might suggest between-population differences; perhaps the high school students differed more from each other than the college students, all of whom had been accepted and chose to attend the same university.

3.2 Previous Studies: Association with PFL

In prior work, Moment-by-Moment Learning Graphs were created for the Genetics Tutor and then visually analyzed by human coders; the coders examined the graphs and chose for each instance the visual patterns that can be observed in it (either a single pattern or multiple patterns). In specific, seven specific visual patterns of the MBMLG were identified, coded by human coders (achieving high inter-rater reliability), and then those human labels were correlated with scores on a PFL test [7]. In that work, it was found that two patterns of the MBMLG are statistically significantly associated with PFL, specifically (see Figure 3): 1) *Plateau* - three or more sequential problem steps that have significantly higher values for P(J) than the rest of the student's behavior. This form represents students who have steady learning (e.g., steady improvement in performance) during only part of the learning activity. The plateau visual form was found to be negatively associated with PFL (r=-0.27, statistically significant when controlling for multiple comparisons). 2) *Immediate drop* – the first problem step for the skill has a high value for P(J), which then immediately falls to low values for the rest of the learning. Immediate drop most likely represents a student who already knows the relevant skill and simply must transfer it into the learning system, or a student immediately mastering a very easy skill. Immediate drop is positively associated with PFL (r=0.29, statistically significant when controlling for multiple comparisons), suggesting that students who already know a skill are more likely to be prepared for future learning when they start the tutor, or that the over-practice that the tutor represents for these students may be enhancing their preparation for future learning. This suggests the hypothesis that over-practice can lead students to not only develop greater speed of performance [23] and lower probability of forgetting [24], but also to deeper conceptual knowledge required to prepare them for future learning.

Additionally, "Spikiness" in the MBMLG – that is, the extent to which there is a prominent peak in the graph, which might indicate a "eureka" moment (cf. [3,4]) – was shown to be correlated with PFL and as a significant factor in a PFL machine-learned prediction model [5]. These results suggested that the visual or functional form of the Moment-By-Moment Learning Graph can be highly associated with preparation for future learning. However, the results in [7], as they rely upon human labels, are not sufficient for use to improve the automatic adaptive behavior of educational software; also, human labels are not easily available at scale for larger studies. The model developed in [5] uses measures other than the MBMLG, and only simple measures of the MBMLG; hence it does not fully demonstrate the potential of the graph to individually predict PFL. In this paper, we attempt to extend these approaches by assessing the mathematical properties of the MBMLG in an automatic fashion, and developing a model that relies solely on these properties to predict PFL.



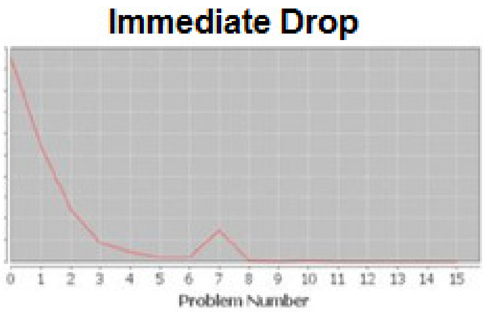


Figure 3. Examples of MBMLG patterns that were found to be significantly related to PFL: plateau (top) and immediate drop (bottom)

4. FEATURE ENGINEERING

In this paper, we attempted to distill quantitative features of the MBMLG to use in automated prediction of PFL. 15 features were computed for each MBMLG, and used as potential predictors of PFL. The full list of features is given here. Features included in the prediction model are highlighted in boldface; in square brackets, a short name is given for each variable, to be used later in the article:

- Average moment-by-moment learning [avgMBML]
- Sum of moment-by-moment learning values [sumMBML]
- Number of opportunities to learn the KC [graphLen]
- **Area under the graph [area]**
- Height of the largest peak [peak]
- Height of the 2nd-largest peak [2ndPeak]
- **Height of the 3rd-largest peak [3rdPeak]**
- First index of the largest peak (index = 1 equals the first step involving the skill, index = 2 equals the second step involving the skill, and so on) [peakIndex]
- First index of the 2nd-largest peak [2ndPeakIndex].
- Distance (in terms of number of problem steps) between the largest and the 2nd-largest peaks [2PeaksDist]
- Distance between the largest and the 2nd-largest peaks, divided by the total number of steps involving the KC [2PeakRelDist]
- Decrease [%] of magnitude from largest to 2nd-largest peak [2PeakDecr]
- Decrease [%] from largest to 2nd-largest peak, divided by the distance (# steps) between them [2PeakRelDecr]
- **Decrease [%] from largest to 3rd-largest peak [3PeakDecr]**
- **Decrease [%] from largest to 3rd-largest peak, divided by the distance (# of steps) between them [3PeakRelDecr]**

5. PFL PREDICTION MODEL

To predict the PFL test results using the 15 MBMLG features, we averaged values of these variables across the sets of genetics problem-solving skills in each of the two tutor modules. Using this data set, we built a model to predict the PFL test from the quantitative attributes of the MBMLGs, using linear regression with forward selection of model features (cf. [21]). The model was validated using student-level leave-one-out cross-validation. In addition, a first pass was conducted prior to model selection where features were eliminated if, taken individually, they had cross-validated correlation below zero. This procedure was adopted as an extra control and over-fitting. This first pass eliminated five variables ([peakIndex], [2ndPeakIndex], [graphLen], [2PeaksDist], [2PeakRelDist]).

Goodness of fit was assessed using the Pearson correlation between the predicted PFL score and the actual score. The best-fitting model has a cross-validated correlation of r=0.532 with actual PFL scores, substantially better than the cross-validated correlations previously found (e.g., [4]) for models based on meta-cognitive and behavioral features (0.360) or models assessing student skill within the software (0.285). The best model is presented in Table 1.

Interestingly, the 3rd-largest learning peak is involved in three of the four variables selected into the best predictive model. In specific, the magnitude of the 3rd-largest learning peak is positively associated with PFL, and large gaps in the size between the largest and 3rd-largest learning peak (measured by *decrease [%] from largest to 3rd-largest peaks*) are positively associated with PFL. These relationships may suggest that single "eureka" moments might indeed be meaningful for robust learning. Having the steepness of the decrease between the largest and to the 3^{rd}-largest peaks (*decrease [%] from largest to 3rd-largest peaks, divided by the distance between them*) with a negative coefficient emphasizes the importance of multiple (though more minor than the most prominent one) learning events which are spread out over time (as opposed to occurring more rapidly). As such, the best pattern appears to be a pattern where the student has multiple substantial moments of learning (at least three) of comparable magnitude, spread out over time.

Table 1. Best-fitting linear regression model predicting PFL

Variable	Coefficient
Area under the graph [area]	-8.459
Height of the 3rd-largest peak [3rdPeak]	+2.634
Decrease [%] from largest to 3rd-largest peaks [3rdPeakDecr]	+0.641
Decrease [%] from largest to 3rd-largest peaks, divided by the distance between them [3rdPeakRelDecr]	-0.296
(Constant)	+0.607

Another finding worth discussing is the negative coefficient of total area under the graph, which indicates that students who have relatively higher learning in the environment have generally lower PFL (when controlling for the other features). While this finding is at a surface-level non-intuitive, it is worth noting that in an effective learning environment, most students learn the skills being taught if they do not already know them. As such, many or most of the students who do not learn the skills being taught already knew the skills to begin with. Indeed, problem-solving pre-test and area are statistically significantly negatively correlated in our study, as shown earlier. These students may have therefore been able to focus on developing more robust learning while using the environment, rather than needing to focus on the exact skills directly taught by the environment.

5.1 Correlations of MBMLG Feature with PFL

Having built the full model, it may be worth examining the types of relationships between the individual MBMLG features and PFL score. In particular, some variables may reverse direction in a complex model. Therefore, in order to understand individual features' relationship to PFL, we computed correlations between each feature the PFL score. In addition, we computed correlation with pre-test scores, in order to explore the possibility that some of these relationships might be explained by prior knowledge. Results are summarized in Table 2 (full names of the variables, along with their shortened names, appear in Section 4, Feature Engineering), and are discussed in this section.

Three features that aggregate overall measures of moment-by-moment learning were found to be significantly negatively correlated with both pre-test and PFL scores; these features are: *average moment-by-moment learning [avgMBML]*, *sum of moment-by-moment learning values [sumMBML]*, and *area under the graph [area]*. That is, low values of learning as reflected in the MBMLG are indicators of high prior knowledge, which is in turn a good predictor of PFL.

The features that measure values of the largest peaks of the MBMLG are significantly negatively correlated with PFL. That is, the larger the values of the graph's highest peaks, the lower the PFL score is. Interestingly, the correlation between the 2nd-largest peak and PFL is stronger than that of the largest peak; and the correlation between the 3rd-largest peak and PFL is stronger than that of the top two largest peaks. It is important to notice that *height of the 3rd-largest [3rdPeak]* is significantly (though mildly) negatively correlated with prior knowledge, while the *height of the 2nd-largest peak [2ndPeak]* is marginally negatively correlated with prior knowledge, and *height of the largest peak [peak]* is not significantly correlated with prior knowledge. Hence, based only on a few meaningful learning events (three, to be more specific), we can conclude that the student was not properly prepared to the learning to begin with, and as such – the student is probably not prepared for future learning as well.

Lastly, the four features that measure the decrease in the graph largest peaks – both absolute and relative to the distant between the peaks – are significantly positively correlated with both pre-test and PFL scores. It is important to first fully understand the meaning of these four features. The larger the absolute decrease between the largest peak values of the graph – measured by *decrease [%] of magnitude from largest to 2nd-largest peaks [2PeakDecr]*, *decrease [%] of magnitude from largest to 3rd-largest peaks [3PeakDecr]* – the more likely it is that there was a single meaningful learning event across the learning process. The higher the value of the relative decrease between the largest peak values – measured by *decrease [%] from largest to 2nd-largest peaks, divided by the distance between them [2PeakRelDecr]*, *decrease [%] from largest to 3rd-largest peaks, divided by the distance between them [3PeakRelDecr]* – the more likely it is that the graph peaks are either different in value from each other, or that they are close to each other. So, these features' positive correlations with PFL suggest that either single learning events, or temporally close multiple learning events are associated with robust learning.

That said, there are two interesting sign-flips observed between the individual features' correlations with PFL scores and their coefficients in the prediction model: the coefficient of the *3rd-largest peak [3rdPeak]* is positive in the model while it is negative when correlation is examined individually; and the coefficient of the steepness of the decrease from the largest to the 3rd-largest peaks *[3rdPeakRelDecr]* is negative in the model while it is positive when correlation is examined individually.

Considering these results along with the full model, we might conclude that when controlling for overall learning (*area under the graph [area]*) and for the prominence of the most meaningful learning event *(measured by Decrease [%] from largest to 3rd-largest peaks [3rdPeakDecr])*, another pattern emerges as an indication to PFL, which is having multiple learning events spread over the learning process.

Another interesting finding regarding the individual feature correlations is that *Number of opportunities to learn the KC [graphLen]* is significantly positively correlated with pre-test, but is not significantly correlated with PFL. This might suggest that over-practice within the tutor is not necessarily associated with robust learning; however, practice within the tutor after gaining knowledge in another fashion might be useful (as immediate drop was found to be positively associated with PFL in Baker, Hershkovitz, et al., in press).

Table 2. Correlations between MBMLG features (N=179) and Pre-test, PFL scores; significant results (two-tailed) are boldfaced (* p<0.05, ** p<0.01), marginally significant results (p<0.1) are italicized

Feature	Pre-test	PFL
[avgMBML]	**-0.35****	**-0.48**
[sumMBML]	**-0.19**	**-0.40**
[graphLen]	**0.30**	0.00
[area]	**-0.35**	**-0.48**
[peak]	-0.09	**-0.35**
[2ndPeak]	*-0.13*	**-0.41**
[3rdPeak]	**-0.20**	**-0.44**
[peakIndex]	*-0.15*	-0.09
[2ndPeakIndex]	-0.03	0.03
[2PeaksDist]	*0.15*	0.05
[2PeakRelDist]	*0.14*	0.07
[2PeakDecr]	**0.29**	**0.45**
[2PeakRelDecr]	**0.26**	**0.40**
[3PeakDecr]	**0.35**	**0.49**
[3PeakRelDecr]	**0.21**	**0.38**

6. CONCLUSIONS AND DISCUSSION

In this paper, we present a new method of predicting preparation for future learning (PFL), based on quantitative analysis of the Moment-by-Moment Learning Graph (MBMLG; [2-3]) the patterns of which were previously shown to be associated with PFL [6]. Overall, we find that using MBMLG features in machine-learned prediction models outperforms previous attempts to predict PFL using BKT parameters and behavioral and metacognitive variables [4-5].

A crucial part of many EDM applications is the feature engineering. In this case, the features we defined and tested were derived from two main streams of literature. First, this work is a natural continuation of previous studies that showed that certain patterns of the MBMLG were strongly associated with PFL, or robust learning in general (cf. [6]). In particular, the presence of *immediate drop* and *plateau* were shown as good indicators of better/poorer PFL, respectively. Hence, the use of measures of the three largest peaks of the graph – the decrease in their values, and the distances between the first and the second. A second relevant line of work is the broader educational research of robust learning. As many studies suggest, learner characteristics have a strong influence on robust learning; of these characteristics, cognitive ability [1,18] can be easily measured via the student model, hence the use of learning measures (which are at the core of the MBMLG).

Another potentially interesting line of future work might be to present MBMLG graphs to teachers and content developers to find irregularities in the learning process (an idea that had previously inspired the creation of *learnograms*, cf. [22]). Teachers and content developers may have insights about the meaning of the MBMLG graphs; they may also find ways to incorporate these graphs in their work to understand their students better.

Overall, our findings suggest that the pattern most associated with a better PFL consists of a process where the student has at least three substantial moments of learning of comparable magnitude, spread out over time. One limitation of the current approach, as it is implemented here, is that PFL prediction is made only after practice has been completed. That is, data cannot be used in real-time like it was used in [4]. However, truncated forms of the MBMLG might be explored for that purpose.

Analyzing the MBMLG qualitatively and using its features to predict PFL is an instance of “discovery with models”, that is of using an existing model (in our case, the MBMLG) in a new analysis (predicting PFL). Discovery with models was suggested as a promising EDM approach in [7]. We build on previous studies involving the same model (e.g., [2-3,6]), and advance the potential automatization of the use of the MBMLG as a component in future studies.

7. ACKNOWLEDGMENTS

This research was supported by grants “NSF #DRL-0910188 and IES #R305A090549, and by the Pittsburgh Science of Learning Center, NSF #SBE-0836012. We would also like to thank Adam Goldstein, Neil Heffernan, Lisa Rossi, Angela Wagner, Zakkai Kauffman-Rogoff, and Aaron Mitchell for their help in this study.

8. REFERENCES

[1] Brandsford, J., & Schwartz, D. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research Education, 24*, 61-100.

[2] Alexander, P. A., & Murphy, P. K. (1999). Nurturing the seeds of transfer: A domain-specific perspective. *International journal of educational research*, *31*(7), 561-576.

[3] Baker, R.S.J.d., Goldstein, A., & Heffernan, N. (2010). Detecting the moment of learning. In *Proc. Intelligent Tutoring Systems* (pp. 25-34). Springer Berlin/Heidelberg.

[4] Baker, R. S., Goldstein, A. B., & Heffernan, N. T. (2011). Detecting learning moment-by-moment. *International Journal of Artificial Intelligence in Education*, *21*(1), 5-25.

[5] Baker, R.S.J.d., Gowda, S. M., & Corbett, A.T. (2011). Automatically detecting a student's preparation for future learning: Help use is key. In *Proc. of the 4th International Conference on Educational Data Mining* (pp. 179-188).

[6] Baker, R.S.J.d., Gowda, S., & Corbett, A.T. (2011). Towards predicting future transfer of learning. In *Proc. Artificial Intelligence in Education* (pp. 23-30).

[7] Baker, R.S.J.d., Hershkovitz, A., Rossi, L.M., Goldstein, A.B., & Gowda, S.M. (in press). Predicting robust learning with the visual form of the moment-by-moment learning curve. To appear in *Journal of the Learning Sciences*.

[8] Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, *1*(1), 3-17.

[9] Belenky, D. M., & Nokes-Malach, T. J. (2012). Motivation and transfer: The role of mastery-approach goals in preparation for future learning. *Journal of the Learning Sciences*, *21*(3), 399-432.

[10] Cen, H., Koedinger, K., & Junker, B. (2006). Learning factors analysis–a general method for cognitive model evaluation and improvement. In *Proc. Intelligent Tutoring Systems* (pp. 164-175). Springer Berlin/Heidelberg.

[11] Chin, D. B., Dohmen, I. M., Cheng, B. H., Oppezzo, M. A., Chase, C. C., & Schwartz, D. L. (2010). Preparing students for future learning with Teachable Agents. *Educational Technology Research and Development*, *58*(6), 649-669.

[12] Corbett, A.T. (2001). Cognitive computer tutors: Solving the two-sigma problem. In *Proc. User Modeling*, 137-147.

[13] Corbett, A.T., MacLaren, B., Kauffman, L., Wagner, A., & Jones, E. (2010). A Cognitive Tutor for genetics problem solving: Learning gains and student modeling. *Journal of Educational Computing Research*, 42, 219-239.

[14] Haskell, R. E. (2000). *Transfer of Learning: Cognition and instruction*. San Diego, CA: Academic Press.

[15] Jastrzembski, T. S., Gluck, K. A., & Gunzelmann, G. (2006, January). Knowledge tracing and prediction of future trainee performance. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*(Vol. 2006, No. -1). National Training Systems Association.

[16] Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36*(5), 757-798.

[17] Li, N., Cohen, W., & Koedinger, K. (2010). A computational model of accelerated future learning through feature recognition. In *Proc. Intelligent Tutoring Systems* (pp. 368-370). Springer Berlin/Heidelberg.

[18] Mazur, J. E., & Hastie, R. (1978). Learning as accumulation: A reexamination of the learning curve. *Psychological Bulletin*, *85*(6), 1256.

[19] Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006). YALE: Rapid Prototyping for Complex Data Mining Tasks. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD 2006), 935-940.

[20] Myers, R. H. (1990). *Classical and modern regression with applications* (Vol. 2). Belmont, CA: Duxbury Press.

[21] Nachimas, R., & Hershkovitz, A. (2007). A case study of using visualization for understanding the behavior of the online learner. In *Proceedings of the International Workshop on Applying Data Mining in e-Learning 2007*.

[22] Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive skills and their acquisition*, 1-55.

[23] Pavlik, P. I., & Anderson, J. R. (2010). Practice and Forgetting Effects on Vocabulary Memory: An Activation-Based Model of the Spacing Effect. *Cognitive Science*, *29*(4), 559-586.

[24] Roll, I., Aleven, V., & Koedinger, K. R. (2009). Helping students know 'further'-increasing the flexibility of students' knowledge using symbolic invention tasks. In *Proc. of the Cognitive Science Society* (pp. 1169-74).

[25] Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22*(2), 129-184.

[26] Segedy, J. R., Kinnebrew, J. S., & Biswas, G. (2011). Modeling learner's cognitive and metacognitive strategies in an open-ended learning environment. In *AAAI Fall Symposium on Advances in Cognitive Systems, Arlington, VA*.

[27] Tan, J., & Biswas, G. (2006). The role of feedback in preparation for future learning: A case study in learning by teaching environments. In *Proc. Intelligent Tutoring Systems* (pp. 370-381). Springer Berlin/Heidelberg.

[28] Wang, Y., & Heffernan, N. T. (2011). Towards Modeling Forgetting and Relearning in ITS: Preliminary Analysis of ARRS Data. In *Proc. Fourth International Conference on Educational Data Mining* (pp. 351-352).

[29] Wang, Y., & Witten, I. H. (1996). Induction of model trees for predicting continuous classes.