Considering the influence of prerequisite performance on wheel spinning

Hao Wan
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA, USA
hale@wpi.edu

Joseph Barbosa Beck
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA, USA
josephbeck@wpi.edu

ABSTRACT
The phenomenon of wheel spinning refers to students attempting to solve problems on a particular skill, but becoming stuck due to an inability to learn the skill. Past research has found that students who do not master a skill quickly tend not to master it at all. One question is why do students wheel spin? A plausible hypothesis is that students become stuck on a skill because they do not understand the necessary prerequisite knowledge, and so are unable to learn the current skill. We analyzed data from the ASSISTments system, and determined the impact of how student performance on prerequisite skills influenced ability to learn postrequisite skills. We found a strong gradient with respect to knowledge of prerequisites: students in the bottom 20% of pre-required knowledge exhibited wheel spinning behavior 50% of the time, while those in the top 20% of pre-required knowledge exhibited wheel spinning behavior only 10% of the time. This information is a statistically reliable predictor, and considering it results in a modest improvement in our ability to detect wheel spinning behaviors: R2 improves from 0.264 to 0.268, and AUC improves from 0.884 to 0.888.

Keywords
Wheel Spinning; Prerequisite; Student Model.

1. INTRODUCTION
Many Intelligence Tutoring Systems (ITS) make use of a mastery learning framework where students continue practicing a skill until they master it. However, some students are unable to achieve mastery despite having numerous opportunities to practice the skill. As a result, these students are stuck in the mastery learning cycle of the ITS and are given additional problems on a topic they are unable to master. We refer to these students as "wheel spinning" on the skill. The term wheel spinning comes from a car that is stuck in snow or mud, and despite rapid movement of the wheels, the car is going nowhere. As defined in [1], a student who takes 10 practice opportunities without mastering a skill is considered to be wheel spinning on this skill. Based on this definition, they also point out that about 31% student-skill pairs in CAT and 38% in ASSISTments are wheel spinning. This earlier work identified the students, but did not provide an explanation for why certain students become stuck. Thus, the next question to address is to understand why students wheel spin in order to provide effective remediation to those students.

Beck and Gong [1] developed a model, consisting of 8 features, to predict which students will wheel spin on a skill. They found that there is a relationship between wheel spinning and gaming the system [12]. Beck and Rodrigo [2] constructed a causal model (using non-Western students) that situated wheel spinning in the face of affective factors. They found that wheel spinning and gaming were strongly related. This work also presented a path model that found gaming was not causal of wheel spinning, but rather, wheel spinning was related to a lack of prior knowledge, which in turn led to gaming. A more concrete wheel spinning model is developed in [3], in which three aspects of features are considered: student in-tutor performance, the seriousness of the learner, and general factors. However, these models do not provide actionable results for how to make a student less likely to wheel spin on a skill, or how to get an already wheel spinning student unstuck.

A natural question is why are some students able to learn a skill and achieve mastery, while other students fail to do so? One plausible hypothesis of what makes wheel-spinning students different from their peers is a difference in ability to learn the skill. Students certainly differ in cognitive abilities, but addressing such would be beyond the scope of most interventions ITS developers can develop. Another plausible difference in ability to learn the skill is due to differences in student preparation. For example, if students do not understand the concept of equivalent fractions, they will have great difficulty mastering the later skill of addition of fractions, which requires them to solve problems such as 1/3 + 1/4.

We define a skill S's prerequisite skills as those skills necessary to be mastered before studying skill S. This prerequisite structure has been used to improve different student models in many research works. For example, Carmona et al. [4] add a new prerequisite layer into student model based on Bayesian Networks. Their experiments suggest that the prerequisite relationships can improve the model's efficiency in diagnosing students. Botelho et al. use prerequisite structure to estimate students' initial knowledge for subsequent skills [5].

Therefore, in this paper, we incorporate the prerequisite structure into wheel spinning model, in order to check if prerequisite performance has impact in wheel spinning of post-skills. Although prior research has proposed automatic algorithms of adapting prerequisite structures [6] [7] [8], we instead use a prerequisite structure developed by a domain expert.

As an overview, we abstract students' prerequisite performance as a feature, and then add this feature into the wheel-spinning model [1]. Our main points include: 1) determine if there is connection between the prerequisite performance and the wheel spinning of post-skill; 2) explore how prerequisite factor would affect wheel spinning model; 3) compare the prerequisite factor with another possible effect that could cause wheel spinning – students' general learning ability. The rest paper is organized as following: Section 2 describes the wheel-spinning model; Section 3 introduces our method of how to represent prerequisite performance; results are shown in Section 4, and further discussion is in Section 5; conclusion and future works are made in Section 6.

2. WHEEL SPINNING MODEL

The wheel spinning model used in this work is mainly derived from the one in [1], but there are two differences between them, we will explain later. This model is fitted using logistic regression algorithm in SPSS on the following features:

a) The number of prior correct responses by the student on this skill. This feature is proved useful in the Performance Factors Analysis model (PFA) [9].
b) The number of problems in a row correctly responded by the on the skill prior to the current problem. Since for this paper we are operationalizing mastery as 3 correct responses in a row[1], the number of consecutive correct responses is an important factor. The value of this feature is from 0 to 2.
c) The exponential mean Z-score of response times on this skill. The response time for each item is transferred into a Z-score, and then exponential mean is calculated for each student by: $\gamma * \text{prior\_average} + (1-\gamma) * \text{new\_observation}$, with $\gamma = 0.7$ found to work well in practice in prior research, and so we have retained it here.
d) The exponential mean count of rapid guessing. This measures how often the student was rapidly guessing.
e) The exponential mean count of rapid response. This measures how often the student took a rapid response. This feature as well as the feature (d) reflects how serious the student is learning the skill through the tutoring system. Similar features related with "gaming" the system were used in gaming detectors as in [10] [11] [12].
f) Count of bottom-out hint. The number of times the student reached a bottom-out hint on this skill prior to the current problem.
g) The exponential mean count of 3 consecutive bottom-out-hints. This measures how often the student reached bottom out hints on 3 consecutive problems.
h) Skill identification.
i) Prior response count.

As aforementioned, the model in our experiments is different from the Beck and Gong's model [1] in two places: one is that we use one more feature in the model, the feature b) above; the other is that in some experiments, we treat the last feature – prior response count – as a covariate, not a factor like in their model. We found this parameter's affect was approximately linear, and thus treating it as a covariate made more sense. We call the model based on these 9 features the baseline model, and compare it with a model that includes the prerequisite performance.

3. METHOD

3.1 Computing Students' Performance on Skills

In this paper, our goal is to find the influence of students' prerequisite performance on wheel spinning. So the first step is to choose which measure to represent students' performance on each skill. In this work, we regard a student's percentage of correct responses to questions involving a skill to be his performance on that skill.

However, a student could answer correctly, by chance, even though this student does not understand the skill at all. Similarly, a student could give the wrong answer through a careless mistake, as in the guess and slip parameters in the Knowledge Tracing model [13]. These two cases will deviate the student's performance from his/her "true understanding" on the skill, especially if the student has very few practices. To deal with these cases, we balance the "accidental performance" with student's overall performance on all skill. The formula for calculating a student's performance on a skill i is:

$$P_i = \frac{1}{2^x} * \bar{R} * S_i + \left(1 - \frac{1}{2^x}\right) * C_i$$

- x: The number of practices on this skill;
- S_i: The percent correctness of skill i, $S_i = \frac{\#\text{correct practices}}{\#\text{overall practices}}$ (over all students). This also reflects the hardness of skill S_i.
- C_i : The student's percent correctness on skill i , $C_i = \frac{\#\text{correct practices}}{\#\text{overall practices}}$ (over the student st_1).
- $R_i = \frac{C_i}{S_i}$: This represents how well the student st_1 does on skill i comparing with the other students.
- $\bar{R} = \frac{\sum_{i=1}^{m} R_i}{m}$: m is the number of the student's started skills.

Table 1. A small sample of students' practices.

Student	Skill	Problem	Correct?
st1	s1	p1	1
st1	s1	p2	0
st1	s2	p3	1
st1	s3	p4	0
st2	s1	p1	1
st2	s1	p2	1
st2	s3	p5	1

Table 2. Calculated skills' hardness and students' performance according to the data in Table 1.

Skill	Correctness	Student performance		Normalized performance	
		st1	st2	st1	st2
s1	0.75	0.48	1.06	0.45	1
s2	1.0	0.78	1.67	0.47	1
s3	0.5	0.28	0.92	0.3	1

[1] We use this definition for consistency with prior work, and for ease of application across systems. This mastery criterion is fairly weak, and presumably underestimates the amount of wheel spinning.

Notice in the formula, the more practices on a skill, the more weight is assigned to the performance on this skill. Take the data in Table 1 as an example. There are in total 4 trials for skill s1, of which 3 are answered correctly, so its correctness is 0.75. The correctness of the other two skills is: s2, 1.0; s3, 0.5. The student, st1, answered two problems of s1, getting one correct and the other incorrect. So this student's correctness of s1 is 0.5, and $R_1(st1) = \frac{0.5}{0.75} = 0.67$. We can also get that $R_2(st1) = 1.0, R_3(st1) = 0$, then $\overline{R}(st1) = 0.56$. Hence, the student st1's estimated understanding on the skill s1 is: $\frac{1}{2^2} * 0.56 * 0.75 + \left(1 - \frac{1}{2^2}\right) * 0.5 = 0.48$. All the performance results are shown in Table 2. Sometimes, a student's adjusted performance is larger than 1, as the student st2's performances on skill s1 and s2. This effect can occur by a student doing very well on a very difficult skill. In this paper, we normalize the values to bring them in the range from 0 to 1.

3.2 Computing Prerequisite Performance

Once the normalized students' performances have been computed, the next step is to think about how to represent prerequisite performances, and then incorporate it into the wheel-spinning model. If a skill has only one pre-required skill, such a representation is straightforward: the student's adjusted performance on that pre-required skill. But what if a skill has multiple prerequisites? In our data set, 39 out of 128 skills have multiple prerequisites. There are a variety of approaches for handling multiple prerequisites. We chose two different methods to compute the prerequisite performance: weakest link and weighted by hardness.

3.2.1 Weakest Link

This method is based on an assumption that learning a skill requires mastery of all its prerequisites. For example, lack knowledge of square or square root might not solve the Pythagorean equation. Therefore, this method regards the prerequisite skill with the worst performance, called weakest link, as the bottom boundary of estimation of prerequisite knowledge.

In this paper, we use the lowest performance value in all prerequisite skills as the wheel-spinning model's input for prerequisite performance. For example, in Table 1, if skill s1's prerequisite skills are s2 and s3, then the prerequisite performance for student st1 on skill s1 is estimated as 0.3 (normalized).

3.2.2 Weighted by Hardness

This method assumes each prerequisite skill has different importance in affecting learning a post-skill, and this importance is determined by how hard the prerequisite skill is. Thus, we sum up a student's prerequisite performances by assigning a corresponding weight to each prerequisite skill, according to the skill hardness. Here we define a skill's hardness to be $1/correctness$. Thus, for a skill, the representation for its prerequisites is calculated as:

$$Pr_i = \frac{\sum_{j=1}^{n} w_j P_j}{\sum_{j=1}^{n} w_j}$$

- n: Number of prerequisites.
- P_j: A student's performance on the jth prerequisite.
- $w_j = \frac{1}{S_j}$: The weight assigned into the jth prerequisite. S_j is the correctness of this prerequisite.

Suppose we also have the skill s1's prerequisites are s2 and s3, then using the data from Table 1 the student st1's prerequisite performance on skill s1 is:

$$\frac{0.47 * \frac{1}{1} + 0.3 * \frac{1}{0.5}}{\frac{1}{1} + \frac{1}{0.5}} = 0.36$$

Respectively, the student st2's prerequisite representation value for s1 is 1.

3.3 Defining General Learning Ability

Our approach is to construct a variable, which we refer to as General Learning Ability (GLA), that encapsulates some of the constructs like diligence, home support, raw ability, and so on. GLA refers to a student's latent ability that affects his ability to learn new skill, similar in spirit to the unidimensional trait in Item Response Theory (IRT) [14]. In IRT, a student's trait is assumed measurable; it is measured through a series of adaptive questions given by a tutoring system.

To simplify our work, we measure student's general learning ability as following steps:

a) For each student-skill pair, randomly select the other two started skills. Here a started skill means the student has practiced at least one problem on it;
b) Compute the performance values for the two skills, as described in Section 3.1;
c) Take the average of those two performance values as the general learning ability for this student-skill pair.

Our intuition in defining GLA in this manner is that if the reason for WH's strong gradient with wheel spinning (Figure 3) is due to the knowledge of the prerequisite being important, we would expect GLA to perform poorly. However, if the power of WH comes not from estimating a particular aspect of student knowledge, but rather than providing a proxy measurement for a student's general ability and willingness to learn, we would expect estimating the student's knowledge of two random skills would work as well. We chose to use two random skills since that was the average number of prerequisites, and we wanted to avoid issues with one measure having lower variability (and hence higher reliability) simply by being an aggregate of more skills. One potential drawback of our approach is that two skills is a small number, and in some cases will certainly provide an over- or under-estimate of knowledge for a particular student. However, since our sample size is large enough, 48256 student-skill pairs in total, this approach is unlikely to produce skewed results.

4. RESULTS

4.1 Data Set

The data in this work is from ASSISTments. We tracked all ASSISTments students when they used the system to practice Math problems for almost a full year from September 2010 to July 2011. This data set contains 7591 different students, and we randomly select 4976 of the students (about 2/3 of students) to form our training data set, while the other students comprise the testing data. There are 31301 student-skill pairs in the training set and 16955 in the testing set. In this work, we consider students who fail to achieve mastery within 10 practice opportunities for a skill (including indeterminate cases [1]) as wheel spinning, which results in 20.6% instances in the training set as wheel spinning and 19.2% in the testing set.

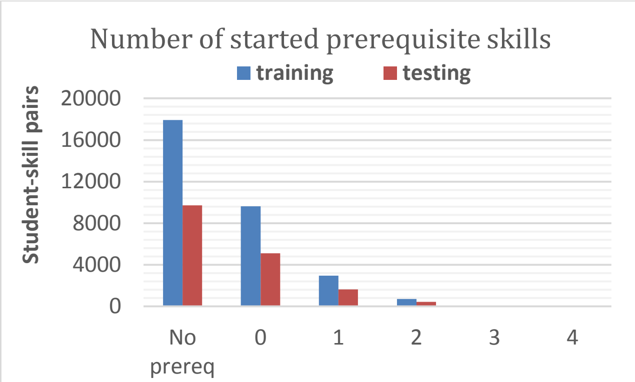

Figure 1. Distribution of number of started prerequisite skills in training set and testing set.

In the training data, there are 177713 problems solved by the students, while 97768 problems in testing data. These problems cover 128 different skills. In the training and testing set, students learn different skills. The maximum number of learned skills by a student is 61, and the average is 6.4. As aforementioned, the prerequisite-to-post skill structure is defined by domain expert as a recommended sequence of topics for instructors. Among the skills in our data set, 66 skills have at least one prerequisite. Some skills have multiple prerequisites, the max number of prerequisites is 8, and the average is 2.4.

However, it is the teacher's choice which skills and in which order to assign to students. Consequently, the majority of student-skill pairs do not have any started prerequisite skills in our data set, as shown in Figure 1. Apparently (and understandably), teachers are less likely to assign review material than to focus on new topics. The maximum number of started prerequisites is 4, and the average is only 0.37. Thus, our experiments will run over three different data sets:

- D1: the whole data set, as depicted in Figure 1, which is splitted into training and testing set.
- D2: the prerequisite data set. This data set excludes the skills that have no prerequisite skills, as identified by the domain expert, from D1. Thus, it is comprised of the points on the x-axis in Figure 1 corresponding to 0, 1, 2, 3 and 4. It is also splitted into training and testing set, and its training set is constructed from the training set in D1 by removing the non-prerequisite skills, while its testing set from testing set in D1 respectively.
- D3: the *started* prerequisite data set, and includes only student-skill pairs where the student has at least begun one of the prerequisites. This data set excludes the skills that have no started prerequisite skills from D2. Thus, it is comprised of the points on the x-axis in Figure 1 corresponding to 1, 2, 3 and 4. Similarly, its training (testing) set is generated from training (testing) set in D2 by removing non-started-prerequsite skills.

The reason for these three datasets is that they answer different research questions. D1 enables us to investigate the impact of prerequisite performance on wheel spinning in an already-existing system in a real-world deployment. That is, how much benefit would we see in the current usage context of the tutor. Unfortunately, that real-world deployment involves teachers assigning no work on most prerequisites, and thus no information about student prerequisite knowledge is available to the model. D2 enables us to examine where there is at least potential benefit. D3 enables us to answer questions about whether a system that had fuller information about prerequisite would perform better at detecting wheel spinning. D3 lets us consider possible changes to policy where teachers are more willing to assign review work, or a system is better able to access past student performance to assess prior knowledge.

4.2 Prerequisite Effect on Wheel Spinning

4.2.1 *The Gradient of the Wheel Spinning Ratio*

In order to determine how likely a student will be to wheel spin on a skill based on his corresponding prerequisite performance value, we focus on the training set of D3. We separate D3 into 5 bins according to the prerequisite performance value, calculated by the method weighted by hardness. The wheel spinning ratio in each bin is shown in Figure 2, named WS Ratio - WH.

As observed in the figure, there is a strong gradient with respect to the prerequisite performance: students in the bottom 20% of pre-required knowledge exhibited wheel spinning behavior 50% of the time, while those in the top 20% of pre-required knowledge exhibited wheel spinning behavior only 10% of the time. This expresses strong evidence supporting our hypothesis that student's wheel spinning on post-skill results from poor preparation for future learning in terms of prerequisite knowledge [15].

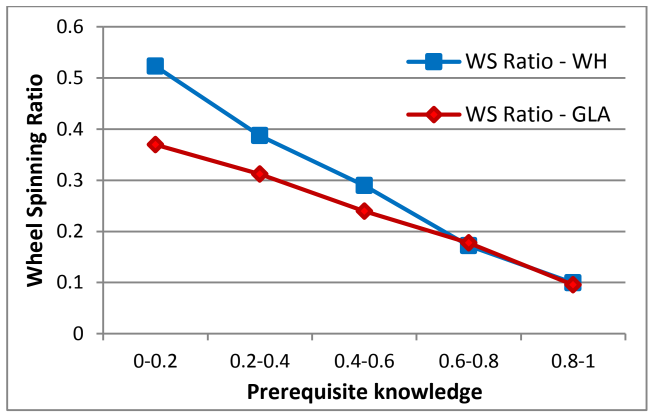

Figure 2. Wheel spinning ratio according with respect to prerequisite knowledge and general learning ability on D3.

4.2.2 *Changes in the Model*

To test the impact of prerequisite features, we integrated them into the wheel-spinning model described previously. We compare the effects of different factors in the wheel spinning model, Weakest Link (WL), Weighted by Hardness (WH), and General Learning Ability (GLA). Table 3 shows the results of training each model on the training test, and evaluating it on the test set.

In this experiment, we use the Cox and Snell R square [15] and AUC (area under curve) to measure model fit. As we can see, the model does not appreciably change in the data set D1, due to the fact that the part of the data containing started prerequisite skills is such a small component of the data. In D2 and D3, the model is improved slightly by integrating the prerequisite feature, WH or WL. This result supports that prerequisite performance is useful in determining students' wheel spinning status in postrequisite-skills. We can also notice that the model with GLA has the similar results with the ones with WH and WL.

Futhermore, to comare the difference between models, a paired t-test is applied on the results at the student's level of each pair of models, as shown in Table 4. The result shows that adding a prerequisite factor – WH or WL – into the baseline model makes it performing significantly differently in all data sets, D1, D2, and D3. On the other hand, the model "Baseline+WH" and "Baseline+WL" have the similar results in those three data sets, which also implies these two prerequisite features have similar effect in the wheel spinning model. More interesting, the p-values indicate that the model with GLA is significantly different from the model with WH (or WL respectively) in D1 and D3, but not in D2, and significantly different from the Baseline model in D2, but not in D1 and D3.

Table 3. Measurements of different models.

Model	R Square			AUC		
	D1	D2	D3	D1	D2	D3
Baseline	0.285	0.301	0.264	0.879	0.888	0.884
Baseline +WL	0.285	0.302	0.268	0.879	0.889	0.887
Baseline +WH	0.285	0.302	0.268	0.879	0.889	0.888
Baseline +GLA	0.291	0.306	0.268	0.883	0.891	0.887

Table 4. P-values of paired t-test. In each data set (D1, D2, and D3), we first compute the RMSE for each model predicting over each student. And then the t-test is applied on the RMSE results at the student's level for each pair of models. The p-values in this table are shown in the order (D1, D2, D3).

	Baseline	Baseline+WL	Baseline+WH
Baseline +WL	<0.01,<0.01, <0.01		
Baseline +WH	<0.01,<0.01, <0.01	0.62, 0.1, 0.27	
Baseline +GLA	<0.01,<0.01, 0.21	<0.01,0.29, <0.01	<0.01,0.3, <0.01

4.2.3 Impact of Prerequisite Effect on the Predictive Model

We now move to determining the impact of the prerequisite feature on the predictive model. In our intuition, the prerequisite factor might have strong effect in predicting wheel spinning when a student just starts learning a post-skill, and the effect weakens with time as the student solves problems on the postrequisite skill

In the logistic regression algorithm, researchers typically use the odds ratio, exponential the coefficient, to represent effect of the corresponding feature [15]. Then the coefficient could be also used to represent the effect on the model. Therefore, in this work, we use the coefficient of prerequisite feature to reflect its effect in predicting students' wheel spinning on post-skill.

In this experiment, we group the D3 of training set by amount of practice on the skill, and construct a wheel spinning model for each group. The coefficients of prerequisite feature (for the WH model) in the corresponding models are shown in Figure 3. As we can see, the coefficient representing the impact of prerequisite knowledge has the highest value at the beginning, and it decreases in influence as students obtain more practice on the skill. This result support our intuition that the prerequisite factor is a good predictor for wheel spinning only at the beginning stage of learning post-skill. Thus, prerequisite knowledge is useful for overcoming the cold start problem in student modeling. When a student first starts working on a skill, his performance on that skill provides little basis with whether to classify him as likely to wheel spin or not. In this situation, knowing how he performed on the prerequisite skills provides some information in his ability to master the current material. As the system observes more and more performances on the skill, those performance provide a much more pertinent source of information about the student's likely trajectory, and the relative importance of prerequisite skills diminishes.

The decrease in in predictive performance for the WH coefficient is monotonic and roughly linear. From a standpoint of statistical significance, the WH coefficient is reliably different than 0 for practice opportunities 1 through 7 (p=0.026 at the 7th opportunity). At the 8th opportunity, the impact of the WH coefficient has p=0.51.

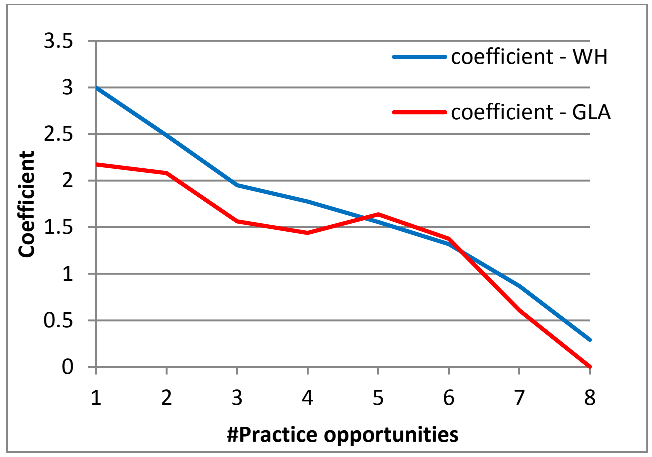

Figure 3. The changes of coefficient with respect to number of practice opportunities on D3.

4.3 Understanding What Prerequisite Performance Really Represents

The performance of the WH feature raises an interesting question: to what does it owe its predictive power. Although we refer to this feature as representing student's prerequisite knowledge, it captures much more than just knowledge. For example, if one student demonstrates strong performance on prerequisite skills and the other does not, those students probably differ in many dimensions beyond knowledge of the skill: diligence in doing math homework, support at home, raw ability at learning new concepts, and perseverance when stuck. Wrapping this bundle of constructs together and calling it "prerequisite knowledge" certainly simplifies discussion, but does a disservice to accuracy. Therefore, we perform a baseline experiment to investigate what prerequisite knowledge represents.

4.3.1 Compare GLA with WH

Since the effects of two prerequisite features, WL and WH, are pretty much the same in the wheel spinning model. Therefore, we will compare only the WH with the GLA. These two features are compared though three different experiments.

The first experiment is to construct wheel spinning ratio gradient for GLA. As we can see in Figure 2, there is the same broad trend for both GLA and WH. For both measures, students with lower general learning ability are more likely to be wheel spinning, which is in accord with our common sense. By comparing the two wheel spinning ratio gradients, we notice that the ratio is the same when the WH and GLA values are high; that is, if a student's performance is relative high (> 0.6) for WH and GLA, then there is a similar chance the student will wheel spin. However, in the lower range of 0 to 0.6, students are more likely to be wheel spinning according to WH value than the students having the same GLA value. This result suggests that prerequisite factor has stronger correlation with wheel spinning than general learning ability, although general learning ability has strong overlap.

The second experiment is to add the GLA into wheel spinning model and compare the model measurements. According to the results in Table 3, adding the GLA into the baseline model makes more improvement than adding the WH on the data set D1 and D2. This is because the student-skill pairs with pre-required knowledge are very rare in those data sets, while every student-skill pair is assigned with a computed GLA value based on that student's performance on a pair of random skills. The model with GLA and the model with WH on the data set D3 have nearly identical performance.

The third experiment is to compare the effect in the learning procedure. As seen in Figure 3, the GLA coefficient also decreases with respect to the number of practice. But in the first 5 practices, the slope of GLA coefficient is more moderate than the slope of WH coefficient, which defends the statement that the prerequisite factor is useful in predicting wheel spinning at early learning stage. By examine the GLA coefficient Wald statistic p-value, it is also statistically reliable ($p<0.05$) before the 7^{th} practice.

5. DISCUSSION AND FUTURE WORK

It should be noticed that even though we found that prerequisite knowledge is related to wheel spinning on post-skills, the general learning ability also has the similar relation. Therefore, it is hard to identify which factor has a stronger connection with wheel spinning in this data set. This is because of two possible reasons: improper prerequisite structure and indirect prerequisite-post relation.

5.1 Prerequisite Structure

As aforementioned, the prerequisite structure used in this work is defined by domain experts. Through this structure, the experts suggest a general curriculum over all grades, not specified in a single year or a single class. It is certainly possible that our structure is in error either by missing some links and incorrectly creating others. Such errors would impact the results.

Moreover, in the method of computing prerequisite performance for a post-skill, we assume that the prerequisite skill with the worst performance (or the hardest prerequisite skill) has the strongest influence in learning post-skill. However, this assumption might be inappropriate here. Botelho [5] et al. also illustrate in their experiments that the prerequisite relation in some post-skills are not as stable as expected by domain experts.

Therefore, there are two possible ways of improving our experiments. The first one is to construct a prerequisite structure specifically for the data. Previous works have been focused on this area. For example, Vuong et al. [8] introduce a method for finding prerequisite structure within a curriculum. Their method calculates the overall graduation rate for each unit, and regards Unit A as prerequisite knowledge for Unit B if the experience in Unit A promotes graduation rate in Unit B.

The other possible way is to measure the correlation between each prerequisite skill and a post-skill, and then we can obtain which prerequisite skill is most effective in affecting learning post-skill. Vuong et al. also distinguish the prerequisite relationship between significant and non-significant in their work [8].

5.2 Prerequisite-post Relation

Obviously, students' general learning ability influences their performance in both prerequisites and post-skills. Therefore, one might argue that there is no direct causal prerequisite-post relationship. The student who is wheel spun on learning post-skill as well as lack of pre-required knowledge is mainly because he/she has weak learning ability, as shown in Figure 4. In this view, GLA is the primary driver of both prerequisite and postrequisite performance.

According to this argument, a consequent case would be: a student who is wheel spun on a skill, he/she will be wheel spun on every skill, due to the weak learning ability. However, in our data set, the wheel spinning ratio of the students who have at least one wheel spinning case is about 23%. Thus, the GLA is an effective factor in wheel spinning, but not a unique or crucial one. Another drawback of this model is that, for low levels of performance, prerequisite knowledge is more strongly related to wheel spinning than GLA. Therefore, even if GLA is the primary driver, there is apparently some impact of prerequisite knowledge on postrequisite performance, represented by the dotted line in Figure 4.

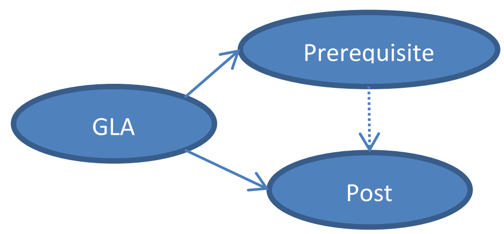

Figure 4. A structure to explain indirect prerequisite-post relationship.

In order to validate the structure in Figure 4, a subtler model should be constructed, in which students' GLA is finely measured. A proper way is to utilize the IRT model to estimate a student's trait; this trait is regarded as the GLA value. And then it is used in predicting if the student will be wheel spinning or not. Meanwhile this trait is updated for each item practiced or for each skill learned. The similar work is in [16], the authors integrate temporal IRT into Knowledge Tracing model, in order to track students' knowledge stage and predict next problem correctness.

6. CONTRIBUTIONS AND CONCLUSION

This work makes two contributions. First, it examines the relationship between prerequisite performance and wheel spinning. One plausible hypothesis for why some students are stuck in the mastery learning cycle is due to inadequate preparation in the building block skills. We found such an association, with students who performed less well on the prerequisite skills being more likely to wheel spin. This work represents an advance over what is known about wheel spinning [1][2].

The second contribution of this work is unpacking what is meant by knowledge of prerequisite skills, and discovering that it is not always related to relevant knowledge. Specifically, by showing that two random skills work approximately as well as prerequisite performance, we show that, for this study, the impact is largely due to general properties of the student than the student's knowledge about particular skills. This reasoning is more than a semantic game, as it directly impacts the conclusions we can draw from our data.

Given just the WH line in Figure 2, a reasonable interpretation is that we can reduce wheel spinning by increasing student prerequisite knowledge, and we could imagine interventions designed to target such. Given the additional context of the results for GLA, we realize that most of the effect attributed to prior knowledge is really just how well the student learns math in general. Unfortunately, interventions to target diligence, grit, math ability, and home support are outside the scope of plausible interventions to deliver with an ITS. However, the difference in the gradients of the two lines suggests there is some benefit from improving student knowledge to at least a moderate level to reduce wheel spinning. This analysis also raises the question of how much work reporting effects related to student prior knowledge is really talking about some other construct than knowledge. Unless the difference in knowledge is caused by a randomized manipulation, differences in knowledge are a proxy for a collection of variables. Hopefully this work will spur EDM researchers to more carefully investigate the meaning of the constructs they are reporting.

In conclusion, this paper investigates the effect of prerequisite performance on wheel spinning and finds that they are related. The addition of prerequisite or GLA features provides a small enhancement in predictive accuracy to our wheel spinning model, improving R2, on skills for which we have prerequisite data, from 0.264 to 0.268, and AUC from 0.884 to 0.888. The baseline model results are quite strong for ITS research, so third-decimal improvement in both metrics is fairly good.

This work also found that prerequisite performance and GLA are both effective for overcoming the cold start problem in student modeling. When students begin working on a skill, the tutor has little knowledge of the student's capabilities on that skill. We found that the new factors in our model had the greatest impact when students were first starting to work with a skill, and diminish in importance as we acquire additional data about his knowledge of the skill.

7. ACKNOWLEDGEMENTS

We acknowledge funding from NSF (# 1440753, 1316736, 1252297, 1109483, 1031398, 0742503), ONR's 'STEM Grand Challenges' and IES (# R305A120125 & R305C100024) grant for ASSISTments and support of the author.

REFERENCES

[1] J. E. Beck and Y. Gong, "Wheel-Spinning:Students Who Fail to Master a Skill," in *Proceedings of 16th International Conference, AIED 2013*, Memphis, TN, USA, 2013.

[2] J. E. Beck and M. M. T. Rodrigo, "Understanding Wheel Spinning in the Context of Affective Factors," in *Proceedings of 12th International Conference, ITS 2014*, Honolulu, HI, USA, 2014.

[3] Y. Gong and J. Beck, "Towards Detecting Wheel-Spinning: Future Failure in Mastery Learning," in *Learning at Scale 2015*, 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 *Proceedings of 10th International Conference, UM 2005*, Edinburgh, Scotland, UK, 2005.

[5] A. Botelho, H. Wan and N. Heffernan, "The Prediction of Student First Response Using Prerequisite Skills," in *Learning at Scale 2015*, 2015.

[6] E. Brunskill, "Estimating Prerequisite Structure From Noisy Data," in *EDM*, 2011.

[7] P. I. Pavlik Jr, H. Cen, L. Wu and K. R. Koedinger, "Using Item-Type Performance Covariance to Improve the Skill Model of an Existing Tutor," in *Proceedings of the 1st International Conference on Educational Data Mining*, Montreal, Canada, 2008.

[8] A. Vuong, T. Nixon and B. Towle, "A Method for Finding Prerequisites Within a Curriculum," in *EDM*, 2011.

[9] P. I. Pavlik Jr, H. Cen, L. Wu and K. R. Koedinger, "Performance Factors Analysis - A New Alternative to Knowledge Tracing," in *the 14th International Conference on Artificial Intelligence in Education*, Brighton, UK, 2009.

[10] I. Arroyol and B. P. Woolf, "Inferring learning and attitudes from a Bayesian Network of log file data," in *Artificial Intelligence in Education*, 2005.

[11] Y. Gong, J. E. Beck, N. T. Heffernan and E. Forbes-Summers, "The impact of gaming (?) on learning at the fine-grained level," in *Proceedings of the 10th International Conference on Intelligent Tutoring Systems (ITS2010)* , 2010.

[12] R. S. J. d. Baker, A. T. Corbett, I. Roll and K. R. Koedinger, "Developing a generalizable detector of when students game the system," *User Modeling and User-Adapted Interaction*, vol. 18, no. 3, pp. 287-314, 2008.

[13] A. T. Corbett and J. R. Anderson, "Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge," *User Modeling and User-Adapted Interaction*, pp. 253-278, 1995.

[14] S. E. Embretson and S. P. Reise, Item Response Theory for Psychologists, Psychology Press, 2013.

[15] R. S. J. D. Baker, S. M. Gowda, A. T. Corbett and J. Ocumpaugh, "Towards automatically detecting whether student learning is shallow," in *Intelligent Tutoring Systems*, 2012.

[16] D. W. Hosmer Jr and S. Lemeshow, Applied Logistic Regression, John Wiley & Sons, 2004.

[17] Y. Huang, J. González-Brenes and P. Brusilovsky, "General Features in Knowledge Tracing to Model Multiple Subskills, Temporal Item Response Theory, and Expert Knowledge," in *Educational Data Mining 2014*, 2014.