

# The Composition Effect: Conjunctive or Compensatory? An Analysis of Multi-Skill Math Questions in ITS

Zachary A. Pardos – Joseph E. Beck – Carolina Ruiz – Neil T. Heffernan  
Worcester Polytechnic Institute

zpardos@wpi.edu – joseph.beck@educationaldatamining.org – ruiz@cs.wpi.edu – nth@wpi.edu

Abstract. Multi skill scenarios are common place in real world problems and Intelligent Tutoring System questions alike, however, system designers have often relied on *ad-hoc* methods for modeling the composition of multiple skills. There are two common approaches to determining the probability of correct for a multi skill question: a conjunctive approach, which assumes that all skills must be known or a compensatory approach which assumes that the strength of one skill can compensate for the weakness of another skill. We compare the conjunctive model to a learned compositional function and find that the learned function quite nearly converges to the conjunctive function. We can confidently report that system designers can implement the AND gate to represent the composition function quite accurately. Cognitive modelers may be interested in the small compensatory effect that is present. We use a static Bayesian network to model the two hypotheses and use the expectation-maximization algorithm to learn the parameters of the models.

## 1 Introduction

Real world problems often involve multiple skills. It is common to have a question in an Intelligent Tutoring System (ITS) that involves more than one skill as well. It is assumed that multi skill questions are harder but how much harder and is the composition of skills accurately modeled by an AND gate or is a more complex function required?

There are two general approaches that have been used in ITS to model how skills compose; one is a conjunctive (AND) approach which multiplies the probabilities that the student knows the skills together to deduce the probability that the student has the required knowledge to answer the question correctly. One example is the Andes physics tutor [4] which uses a leaky-AND assumption to model composition. The other approach is compensatory (OR/additive). This approach assumes that a skill with high probability of knowledge can compensate for a lower probability of knowledge of the other skill(s) involved and thus constitute the required knowledge to answer the question correct. Many varieties of the compensatory approach have been applied in practice including a mixed-effect analysis of tutor data [5] where skills with the lowest probability were used to represent the problem's difficulty. Other investigators have used a type of additive approach with learning factors analysis [3] and with a linear logistic test model [1] that found the additive approach to be inadequate for modeling difficulty using a similar dataset to ours. This paper will investigate how difficulty changes with the number of required skills and address which function is most appropriate for modeling the composition of skills at least in the context of our dataset. Previous work [6], in which the phrase "composition effect" was first

coined, focused on a different aspect of problem difficulty and found that answering a multi skill question correctly was harder than answering two single skill questions correctly.

### 1.1 About the ASSISTment tutoring system

ASSISTment is a web based tutoring and assessment system for 6<sup>th</sup>-10<sup>th</sup> grade math. The tutor started in 2004 in only a few 8<sup>th</sup> grade math classrooms and, in 2008, is now used by 300-500 students per day. The items in the tutor are all based upon publically released state test problems from the Massachusetts Comprehensive Assessment System (MCAS).

### 1.2 About the Dataset

Our dataset was from logged student use of the ASSISTment system during the 2005-2006 school year in which 8<sup>th</sup> grade math students ages 13-14 answered questions on the tutor two or three times per month at their school's computer lab. The students were given randomly chosen problems from a pool of around 300. The random selection of problems gave us a random sampling of skill data throughout the year.

Our dataset contained 441 students with 46 average responses per student. Only a student's first response to a question was considered. The number of single skill questions was 184 with 62 double skill questions and 16 triple skill questions. The majority of questions are text entry while the others are multiple-choice. We would like to note that skills were tagged to questions by subject matter experts from a skill set of 106 [10]. However, the number of skills that have data points is 84. Table 1 shows the attributes we have for each of the question responses and an example of what the response data looks like.

**Table 1.** Sample of the question response dataset

UserID	QuestionID	Skill Tags	Correct on first attempt	Timestamp
1234	5293	Equation-solving, Congruence, Perimeter	0	13-OCT-05 12:20.53
1234	5693	Congruence	1	13-OCT-05 12:27.20
1234	6755	Pythagorean Theorem	0	13-OCT-05 12:33.57

### 1.3 Bayesian Networks

We used a static Bayesian model to represent skill tagging to question nodes and to make inferences on the probability of answering a given question correct or incorrect. Bayesian networks [9] is a powerful machine learning method that we used for making posterior inferences based on binary response data; 1 for correct

0 for incorrect. The EM algorithm [2] can be used with Bayesian networks to learn the priors of latent variables and the conditional probability tables (CPT) of random variable child nodes. Exact inference with Kevin Murphy’s Bayes Net Toolkit for MATLAB was used with EM to learn the parameters of the networks.

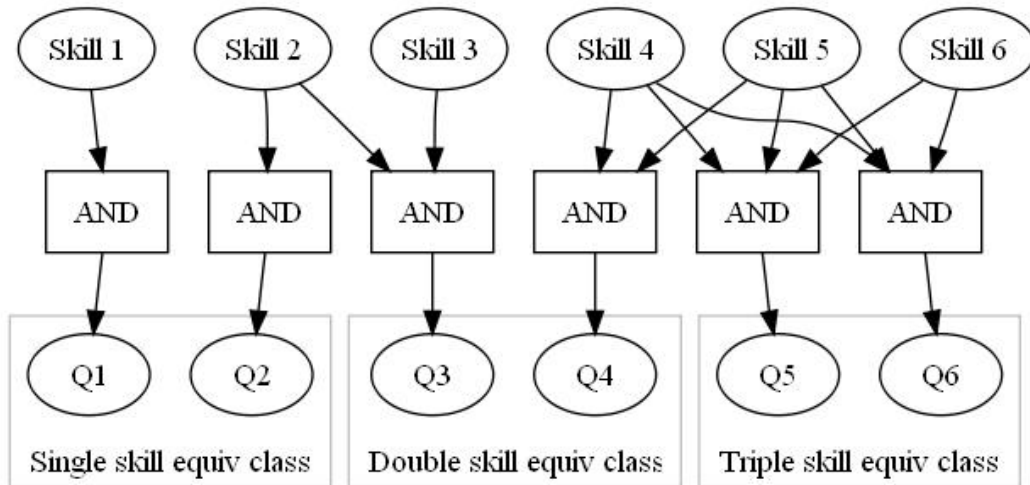
Inferring the probability a student will answer a question correctly ( $Q=correct$ ) is a function of the believed prior on the skill(s) associates with the item ( $S=known$ ) together with the guess and slip parameters of the question. The equation is shown below:

$$P(Q=correct) = (1-P(S=know)) \cdot guess + P(S=know) \cdot (1-slip) \quad (1)$$

A guess parameter dictates the probability that the student will get an item correct even if she does not have the required knowledge. A slip parameter dictates the probability that the student will get an item incorrect even if she has the required knowledge. Learning the general parameters of single and multi skill questions will tell us if questions with more skills are harder to guess.

## 2 The Conjunctive Model

The AND Gate Conjunctive model is the most common approach to skill composition in ITS. The principle behind it is that all skills involved must be known in order to answer the question correctly. The topology of this model is similar to a deterministic input noisy “AND” (DINA) model [7] except that our AND has no noise (no p-fail). The Bayesian belief network is represented by the directed acyclic graph in Figure 2.



**Fig. 2** Directed acyclic graph representing the AND gate Bayesian topology

The network consists of three layers of nodes with equivalence classes used to share conditional probability tables among nodes. This lets us learn a generalized

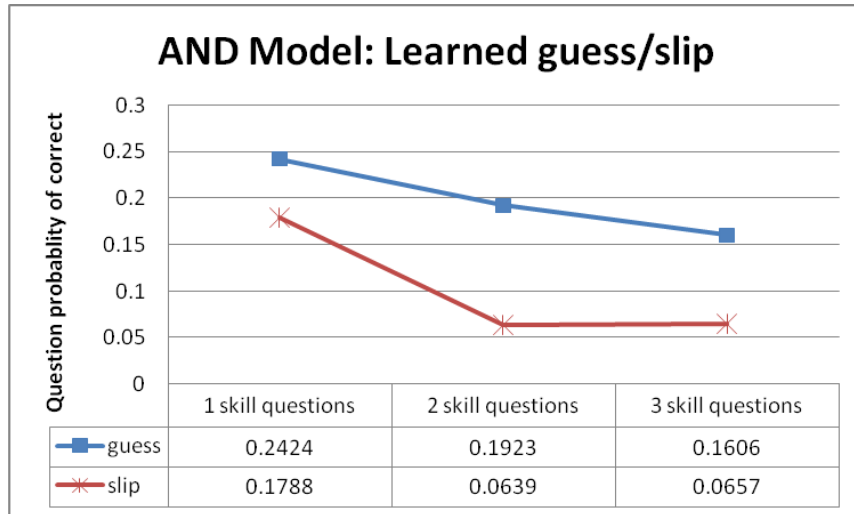
guess/slip for all nodes of a given equivalence class instead of a parameter per node, which could not be accurately learned given the size of our dataset and would also not answer the research question of how multi skill questions differ from single skill questions. The first layer consists of latent skill nodes. All the skill nodes share a single equivalence class, this was done to simplify the EM procedure. The equivalence class learns a single prior for all skills but does not constrain the individual skill performance estimations from differing. The second layer consists of the AND gates which assert that all parent skills must be known in order for the child question to be answered correctly. The last layer consists of the question nodes. All single skill questions are grouped by an equivalence class. All double skill and triple skill questions have their own equivalence class as well. We will eventually be learning a total of three sets of guess/slip values and a prior.

## 2.1 Methodology

The way in which we will approach the research question “how much harder are multi skill questions” is by learning the conditional probability tables of the Bayesian network. By learning the parameters of our model we can observe how skill knowledge determines the probability of correct for a given question. How guess and slip vary with the number of skills involved is one way of investigating the composition effect. To learn parameters in the AND model we chose to learn the single skill guess and slip first. By learning these parameters first we can establish a baseline assumption for question guess and slip that is gathered from single skill question data that does not introduce the complex issue of credit-blame that comes in to effect with multi skill questions. The skills’ prior equiv class was set at 0.50 and the starting value of the single skill equiv class was set at *ad-hoc* values; 0.15 guess and 0.10 slip that have been used in previous conjunctive model work [8]. After the guess and slip parameters have been learned for the single skill equivalence class we move on to learning the double skill and triple skill equivalence classes at once. The single skill parameter values are locked in place and the prior is reset again to 0.50 before the second EM learning begins. After this second step completes we now have three sets of guess and slip values as well as a prior for the skills.

## 2.2 Results

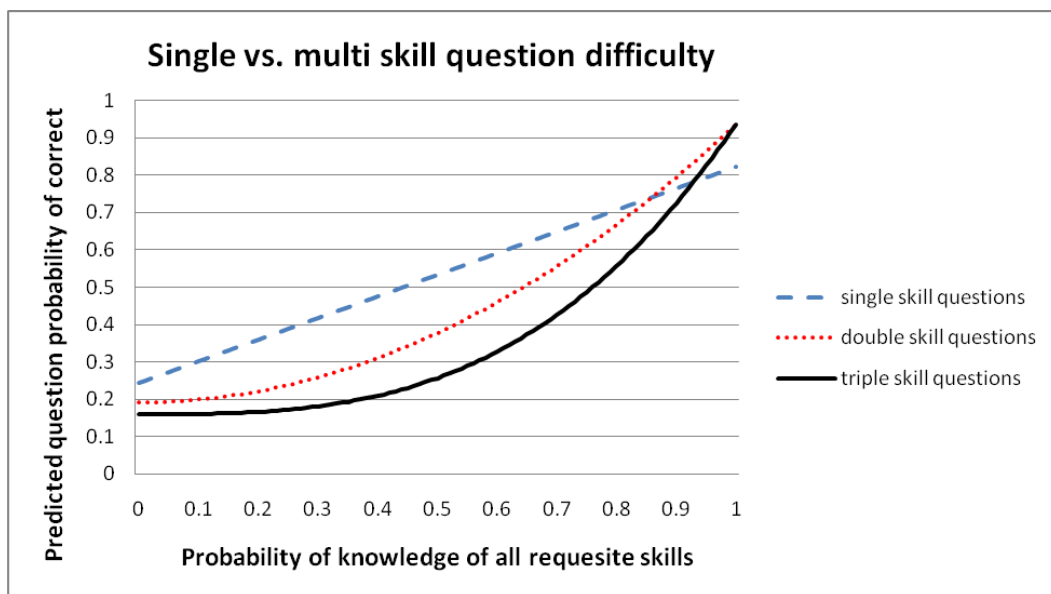
The results from the AND model parameter learning shows that the probability of guess decreases linearly as the number of skills increases; from a 24.24% guess with a single skill questions down to a 16.06% guess with a 3 skill question as shown in Figure 3. Surprisingly the slip rate, or probability of making a mistake, also goes down as the number of skills increase. This suggests that while multi skill problems are more difficult to guess, they are also more difficult to slip on.



**Fig. 3** Results of AND model parameter learning

The difficulty of a problem can be described by the probability of answering the question correctly. However, the probability of answering a question correctly is dependent on the probability of knowing the skill or skills involved. Figure 4 shows how problem difficulty differs with skill knowledge for single, double and triple skill questions. Also note that the guess values from Figure 3 are the intercepts of the left most vertical axis in Figure 4 and the slip values are the right most vertical intercepts.

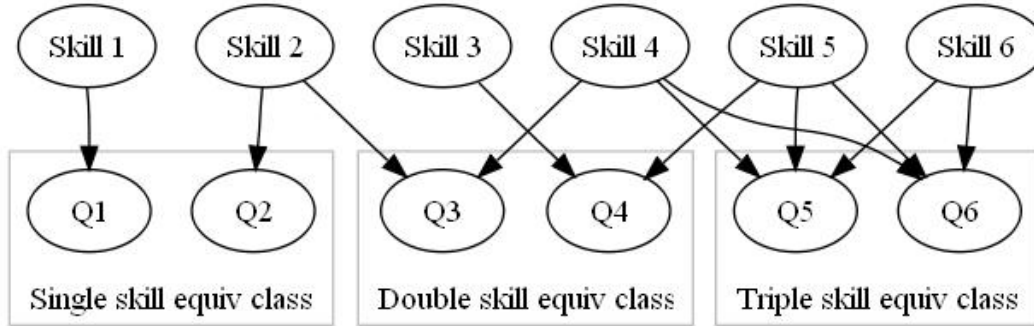
Data for the Figure 4 graph was generated by setting the probability of all skill nodes to zero and then asking the Bayes net to infer the posterior probability of correct for a single, double and triple skill question and recording the results. All skill nodes were then incremented by 0.01 and the steps were repeated up to a probability of 1.



**Fig. 4** Comparison of the difficulty of single, double and triple skill questions

### 3 The Learned Compositional Model

The learned compositional model topology looks similar to the AND model except that there is no layer of gate nodes and the skills are connected directly to the question nodes as seen in Figure 5.



**Fig. 5** Directed acyclic graph representing the compositional Bayesian topology

The fundamental difference between the AND model and the learned compositional mode is that only three parameters, a guess and slip and prior, are learned in the AND model since the composition function was captured by the AND gates + guess/slip. In the learned compositional model, however, the composition function and the guess/slip parameters are captured in one complex CPT. A guess and slip value will still be learned for the single skill equivalence class since there is no composition with a single skill. However, four parameters will be learned for the double skill equivalence class and eight parameters for the triple skill class. The increase in parameters is due to the CPT needing to expand with the increased number of parent nodes. Example CPTs for single and double skill questions are shown in Figure 6.

S1	P(Q=F)	P(Q=T)	S1	S2	P(Q=F)	P(Q=T)
F	0.85	0.15	F	F	0.85	0.15
T	0.10	0.90	T	F	0.80	0.20
			F	T	0.78	0.22
			T	T	0.10	0.90

**Fig. 6** Example CPTs for single and double skill questions

In both CPTs of Figure 6 the 0.15 represents the guess parameter which is the probability the question is answered correctly ( $P(Q=T)$ ) given the skill is not known ( $S1=F$ ); 0.85 is simply the complement of the guess parameter. Observe that in the double skill CPT there is a row where  $S1=T$  and  $S2=F$  and another row where  $S1=F$  and  $S2=T$ . Why might the  $P(Q=T)$  of these two rows differ? Because, for example,  $S1$  could represent the “harder skill”. If the composition function is compensatory, knowledge of only the more difficult skill could result in a  $P(Q=T)$  of greater than 0.50 while knowledge of only the easier skill could result

in a  $P(Q=T)$  of less than 0.50. In the next section we describe how the network topology was organized to capture the notion of a “harder” skill. We found that knowing the harder skill is not much better than knowing the easier skill, suggesting against a compensatory compositional function.

### 3.1 Methodology

The methodology for learning parameters in the learned compositional model was very similar to the AND model with the exception of an initial step required to order the skill nodes. This ordering was done to capture relative difficulty among skills so that a compensatory function, which requires the notion of skill difficulty, could potentially be learned. The default order of skills in the network was alphabetical. Because this ordering has no substantive meaning we decided to order the skills by difficulty with the most difficult skill appearing first (node 1) and the least difficult appearing last (node 106). We used the metric of skill prior to represent the difficulty of a skill. In order to attain the priors on all the skills we let a separate prior be learned for each of the 106 skills during an initial parameter learning of single skill questions. This step was done solely to get a skill order. The learned guess/slip values were discarded. After the skill priors were attained, the network was reordered and the single equivalence class for skill priors was reestablished before the “real” first phase of EM parameter learning was run. This reordering gave extra power to the results and allowed us to ask composition questions such as, “does knowing only the harder skill increase the probability of answering a question correctly over knowing only the easier skill?”

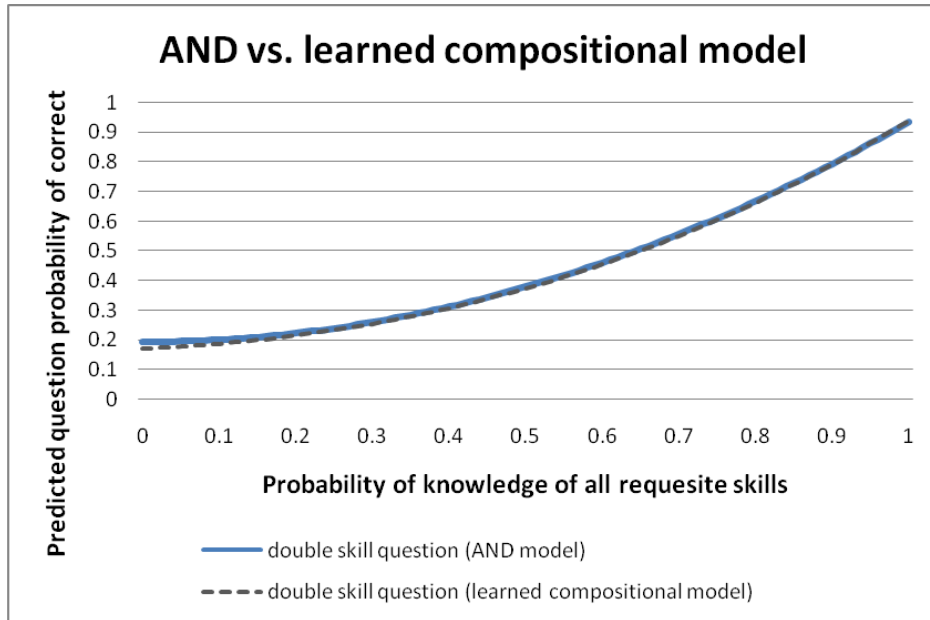
### 3.2 Results

Results of the compositional model parameter learning indicate that the learned composition function is conjunctive. Evidence against a compensatory function can be drawn from Table 2 which shows that knowing one skill only slightly increases the probability of answering a double skill question correctly over knowing no skills.

**Table 2.** Learned double skill CPT for the learned compositional model

Harder skill	Easier skill	$P(Q=F)$	$P(Q=T)$
<b>F</b>	<b>F</b>	0.83	0.17
<b>T</b>	<b>F</b>	0.77	0.23
<b>F</b>	<b>T</b>	0.78	0.22
<b>T</b>	<b>T</b>	0.06	0.94

The table also shows that knowing only the harder skill does not help significantly over only knowing the weak skill. For double skill questions the learned guess is 0.17 and 0.06 slip, nearly the same as the AND model. To further compare and verify the learned compositional model’s similarity to the AND model we generated a graph similar to Figure 4.



**Fig. 7** Comparison of the AND model and learned compositional model

Figure 7, above, shows that the lines from the AND model and learned compositional model overlap. The similarity of the behavior of these functions, arrived at through two different analytic approaches, favors the AND gate as being a very close approximation to the composition function.

### 3.3 Further analysis: deriving the composition function

The values in Table 2, estimated by our model, determine how likely a student is to respond correctly to a multi-skill question on the basis of her knowledge of the two associated skills. For example, a student who knows the harder skill but does not know the easier skill has a 23% chance of responding correctly to the question. These parameters values are affected by two components: the effect of composition and the likelihood of slipping or guessing. Unfortunately, it is impossible to simultaneously model both effects separately since the model is underdetermined. Both the guess/slip and composition effects are latent, and therefore during the parameter estimation phase of EM there are an infinite number of simultaneous solutions. Therefore, we will reuse the slip and guess values from the AND model (guess=0.1923 and slip=0.0639, from Figure 3) as estimates of the effect of slipping and guessing. We then partial those out (analogous to a partial correlation) of the table by using the following equation:  $P(\text{correct}) = (1 - P(\text{known})) * \text{guess} + P(\text{known}) * (1 - \text{slip})$ . In this formula,  $P(\text{known})$  refers to the probability the student knows how to solve the problem accounting for the composition effect;  $P(\text{correct})$  is the value in Table 2, and slip and guess are 0.1923 and 0.0639, respectively. By solving for  $P(\text{known})$ , we can compute the effect of composition: that is, when a student knows neither, one or both of the two skills, how much effective knowledge does she bring to bear on the problem? Solving for  $P(\text{known})$  for each entry in Table 2 yields Table 3.



**Table 3.** Compensatory model with guess/slip factored out

P(known)	Hard skill is not known	Hard skill is known
Easy skill not known	-0.028	0.055
Easy skill is known	0.045	1.004

Table 3 represents the composition function. Although some of the values are impossible in terms of being probabilities, they are only (close) approximations since we were forced to use slip and guess values from a related analysis. In spite of these shortcomings, Table 3 is quite interpretable: it is extremely similar to an AND gate. When a student knows neither skill she has, effectively, zero knowledge. When she knows both skills her knowledge is approximately 1. When she only knows 1 of the skills, she has a slight amount of knowledge, but still fairly close to 0. Replacing these numbers with an AND gate would produce largely similar results, as can be seen in Figures 4, 5 and 6. Therefore, composition is well-modeled as an AND gate. Furthermore, we see no evidence that it is necessary to use a leaky-AND gate [4] to model composition.

#### **4 Contributions**

We have shown that the more skills involved in a question the harder it is to guess the correct answer. We have also shown that the probability of slipping goes down as well with more skills. We speculate that the reason for decreased slip is that students who are believed to know multiple skills are less likely to make a mistake. Another possibility is that multi skill questions demand more concentration and thus a student is less likely to make a careless mistake due to not paying attention. We also found that knowing the more difficult skill in a multi skill question does not help much over knowing only the less difficult skill or over knowing neither skill.

We have investigated the composition function and found that it is approximated very well by the AND gate + guess/slip. While cognitive modelers may be interested in the slight compensatory effect seen in Table 3, ITS developers can have confidence in using an AND assumption for accurate assessment when dealing with multi skill items.

#### **Acknowledgements**

We would like to thank the Worcester Public Schools and all of the people associated with creating the ASSISTment system listed at [www.ASSISTment.org](http://www.ASSISTment.org) including investigators Kenneth Koedinger and Brian Junker at Carnegie Mellon. We would also like to acknowledge funding from the U.S. Department of Education, the National Science Foundation, the Office of Naval Research and the Spencer Foundation.

## References

- [1] Ayers E., & Junker B. W. (2006). Do skills combine additively to predict task difficulty in eighth grade mathematics? In Beck, J., Aimeur, E., & Barnes, T. (Eds). *Educational Data Mining: Papers from the AAAI Workshop*. Menlo Park, CA: AAAI Press. pp. 14-20. Technical Report WS-06-05.
- [2] Baum, L.E., Petrie, T., Soules, G., & Weiss, N. (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The annals of Mathematical Statistics*, 41, 164-171
- [3] Cen, H., Koedinger, K., & Junker, B. (2006). *Learning Factors Analysis - A General Method for Cognitive Model Evaluation and Improvement*. Paper presented at the 8th International Conference on Intelligent Tutoring Systems.
- [4] Conati, C., Gertner, A., & VanLehn, K. (2002). Using Bayesian networks to manage uncertainty in student modeling. *User Modeling & User-Adapted Interaction*, 12(4), 371-417.
- [5] Feng, M., Heffernan, N. T., Mani, M., & Heffernan, C. (2006). Using Mixed-Effects Modeling to Compare Different Grain-Sized Skill Models. In Beck, J., Aimeur, E., & Barnes, T. (Eds). *Educational Data Mining: Papers from the AAAI Workshop*. Menlo Park, CA: AAAI Press. pp. 57-66. Technical Report WS-06-05. ISBN 978-1-57735-287-7.
- [6] Heffernan, N. T. & Koedinger, K.R. (1997). The composition effect in symbolizing: The role of symbol production vs. text comprehension. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum. pp. 307-312.
- [7] Junker, B.W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258–272.
- [8] Pardos, Z.A., Heffernan, N.T., Anderson, B., Heffernan, C.L. The Effect of Model Granularity of Student Performance Prediction Using Bayesian Networks. In the Complete On-Line Proceedings of the Workshop on Data Mining for User Modeling, at the 11th International Conference on User Modeling (UM 2007) Pages 91-100
- [9] Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA
- [10] Razzaq, L., Heffernan, N., Feng, M., Pardos Z. (2007). Developing Fine-Grained Transfer Models in the ASSISTment System. *Journal of Technology, Instruction, Cognition, and Learning*, Vol. 5. Number 3. Old City Publishing, Philadelphia, PA. 2007. pp. 289-304.