Learning curves versus problem difficulty: an analysis of the Knowledge Component picture for a given context

Brett van de Sande
Pearson Education
brett.vandesande@pearson.com

ABSTRACT

The Knowledge Component (KC) picture of learning has proven useful for constructing models of student learning in a number of subject areas. However, it is still unclear how well this picture generalizes to other contexts and subject areas. A corpus of 62,000 exercises for 10 textbooks on the Mastering platform has been tagged by content experts. In this report, I introduce a strategy for investigating the importance of a given set of KCs in describing student performance as the students solve problems. The strategy is to see how much of the student's performance on an exercise is explained by the associated KC and how much it is predicted by a problem-specific difficulty parameter. To do this, I introduce a model that is a combination of the Rasch model and the learning curves from the KC picture. For this corpus and set of KC tags, a rather striking picture emerges: problem difficulty accounts for most of the student behavior while KC learning accounts for only a small portion of the student behavior. I hypothesize that these KC tags do not accurately capture the skills students are using while doing their homework.

Author Keywords

Learning Curves, Knowledge Components

ACM Classification Keywords

I.2.6 Learning: Knowledge acquisition

Knowledge components (KCs) are bits of information needed to solve a problem [5, 2]. KCs generally have some sort of pre-requisite relations. However, aside from prerequisites, a KC can, by definition, be mastered independently from other KCs. This definition assumes that KCs are *context independent*. That is, the student's ability to apply that KC correctly or quickly does not depend on the particular problem the student is solving or the other KCs needed to solve that problem.

Since KCs are *defined* to have these properties, then it remains to be seen whether a given set of KC labels for a particular curriculum provides a useful description of skill acquisition. Much of the pioneering work on KCs focused on middle school math [4]. It is unclear whether this picture extends to the corpus examined here.

Table 1. Some Knowledge Components for Chapter 32 of "University Physics" by Young, Freedman, and Lewis [6].

1	Relationship between speed of electromagnetic (EM) waves, wavelength and frequency
2	Writing Maxwell's equations for free space. Using Faraday's Law.
3	Direction of propagation of an electromagnetic wave

One way to determine how well the KC picture is working is to examine the associated learning curves. If the curves increase/decrease more-or-less monotonically (depending on the measure of competence) then the KC picture is working. A smooth learning curve implies that the associated KCs account for most of the student performance on a problem while other aspects of the problem are less important.

A corpus of over 62,000 exercises on the Mastering platform has been tagged by content experts. This corpus covers homework exercises for 10 college-level textbooks in anatomy and physiology, biology, organic chemistry, general chemistry, and physics. An typical set of KCs is shown in Table 1. On average, there are about a dozen KCs per chapter.

We examined log data from problems solved on the Mastering platform during the Spring of 2014. We selected students whose coursework spanned more than 25 days and who were enrolled in a course containing more than 50 students.

Before we address the main question of the validity of the KC picture for this corpus, we mention some general properties of the log data. The learning curves (see Fig. 1) are expressed in terms of "difficulty" which is defined to be minus the logistic of the probability of "correct on first try."

The mean number of opportunities to practice a given KC is 3.84, averaged over students and KCs. So, students have very few opportunities to practice a given KC.

Also, the number of students practicing a KC usually decreases rapidly with increasing opportunity number t. This can result in a selection bias, since the population is changing with t. Thus, to produce a learning curve for a given KC, we rank the students by the total number of opportunities for that KC and take the uppermost portion as our student population. An example learning curve is shown in Fig. 1. In general, we

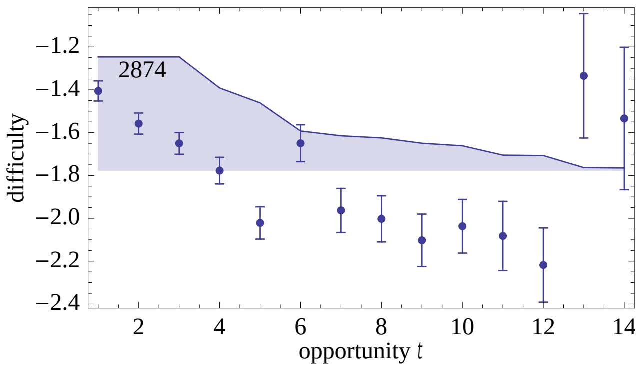

Figure 1. Learning curves for the first KC listed in Table 1. Difficulty should decrease as students learn. The shaded region represents the relative number of students who completed that opportunity and the number in the upper left corner is the initial number of students.

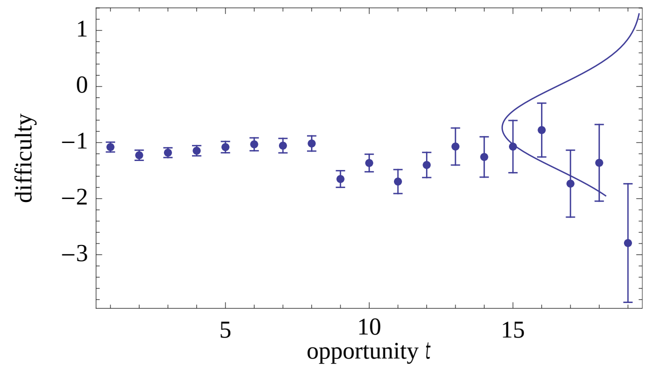

Figure 2. KC difficulties $\zeta_{k,t}$ versus opportunity number t from the Rasch/KC model applied to student log data for the first KC in Table 1. The curve on the right is a gaussian that represents the distribution of problem difficulties for the exercises labeled with the associated KC.

find that learning curves are not monotonically decreasing. In fact most do not even show a decreasing trend.

There must be important aspects of the exercises that are not captured by these KCs. Thus, we introduce problem difficulty β_p to capture the aspects of a problem not explained by the KCs. This leads us to introduce the Rasch/KC model: a hybrid of the Rasch model [3], and the learning curve picture.

If $P_{s,p}$ is the probability that student s gets problem p correct, then we define $P_{s,p}$ by the logistic equation:

$$\operatorname{logit}\left(P_{s,p}\right) = \theta_s - \beta_p - \sum_{(k,t)\in\mathcal{T}_{s,p}} \zeta_{k,t} \qquad (1)$$

where θ_s is the skill of student s, β_p is the difficulty of exercise p, and $\zeta_{k,t}$ is the difficulty of applying KC k on opportunity t. $\mathcal{T}_{s,p}$ is the set of KC, opportunity pairs where $(k,t) \in \mathcal{T}_{s,p}$ means that problem p is opportunity t for student s to apply KC k. The log-likelihood for a set of students and problems to obtain a particular set of outcomes is

$$\log\left(\mathcal{L}\right) = \sum_{s,p\in\mathcal{C}_s} \log\left(P_{s,p}\right) + \sum_{s,p\in\mathcal{I}_s} \log\left(1 - P_{s,p}\right) + \qquad (2)$$

where $\mathcal{C}_s$/$\mathcal{I}_s$ is the set of problems s got correct/incorrect.

If we drop $\zeta_{k,t}$, then we obtain the usual Rasch model. Likewise, if we drop θ_s and β_p and fit the resulting model to student data, a plot of $\zeta_{k,t}$ versus opportunity t will yield the conventional learning curve for KC k; this is precisely what we have plotted in Fig. 1. This model is similar to the Additive Factors Models (AFM) [1] except that AFM restricts $\zeta_{k,t}$ to be linear in t.

We can apply this model to student log data associated with the KCs listed in Table 1. We find that both student skills $\{\theta_s\}$ and problem difficulties $\{\beta_p\}$ are Gaussian distributed with standard deviations of 1.02 and 1.15, respectively.

Looking at the KC difficulties $\zeta_{k,t}$ in Fig. 2 we see that the difficulties vary little with opportunity number. We also, see that the associated problem difficulties, represented by the Gaussian distribution on the right, vary significantly more than the KC difficulties. The same qualitative behavior is seen for all KCs we have analyzed. We conclude that, for this corpus and KC labeling, problem difficulty is much more important than KC mastery when predicting student performance on an exercise.

If we look at the KCs, see Table 1, we see that they represent content knowledge rather than more abstract problem solving skills. It may be that the students have already learned the content knowledge in lecture or reading and, during their homework, they are really learning how to apply that content knowledge to various physical situations. If this is the case, it may be more appropriate to label problems with labels that are more oriented towards problem-solving skills, like "given description of situation, determine that one should relate velocity, frequency, and wavelength." Also, it may mean that one can explain student performance with just a few KCs like "solve physics word problem" or "solve problem with kinematics graphs."

REFERENCES

1. Chi, M., Koedinger, K., Gordon, G., Jordan, P., and VanLehn, K. Instructional Factors Analysis: A Cognitive Model For Multiple Instructional Interventions. In *Proceedings of the 4th International Conference on Educational Data Mining* (Eindhoven, the Netherlands, June 2011).
2. Koedinger, K. R., Corbett, A. T., and Perfetti, C. The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive Sci. 36*, 5 (2012), 757–798.
3. Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests.* MESA Press, Chicago, 1993.
4. Ritter, S., Anderson, J. R., Koedinger, K. R., and Corbett, A. Cognitive Tutor: Applied research in mathematics education. *Psychon. B. Rev. 14*, 2 (Apr. 2007), 249–255.
5. VanLehn, K. The Behavior of Tutoring Systems. *Int. J. Artif. Intell. Ed. 16*, 3 (Jan. 2006), 227–265.
6. Young, H. D., Freedman, R. A., and Ford, A. L. *University Physics with Modern Physics*, 13 edition ed. Addison-Wesley, Boston, Jan. 2011.