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

Brett van de Sande
Advanced Computing and Data Science Lab,
Pearson Education

How much of of a student's performance on an exercise is explained by the associated KC and how much can be explained by a problem-specific difficulty parameter?

To answer this question, I introduce a model that is a combination of the Rasch model and the learning curves from the KC picture.

I use a corpus of tagged exercises from the Mastering platform. (62,000 exercises from 10 college-level textbooks were KC tagged by subject-matter experts.)

Knowledge Components (KCs) for Chapter 32 of "University Physics" by Young, Freedman, and Lewis.

Label	Description
yf13:32.01	Relationship between speed of electromagnetic (EM) waves, wavelength and frequency
yf13:32.02	Writing Maxwell's equations for free space. Using Faraday's Law.
yf13:32.03	Direction of propagation of an electromagnetic wave
yf13:32.04	Relationship between the amplitude of the electric and magnetic field of an EM wave
yf13:32.05	Determine the properties of EM waves given the wave function.
yf13:32.06	Speed of EM waves in dielectric materials
yf13:32.07	Energy and energy density of EM waves
yf13:32.08	Intensity of EM wave in terms of amplitudes of electric and magnetic fields
yf13:32.09	Finding the Poynting vector for EM waves
yf13:32.10	Calculate the pressure due to EM waves and the momentum density
yf13:32.11	Find the nodes and antinodes of standing EM waves

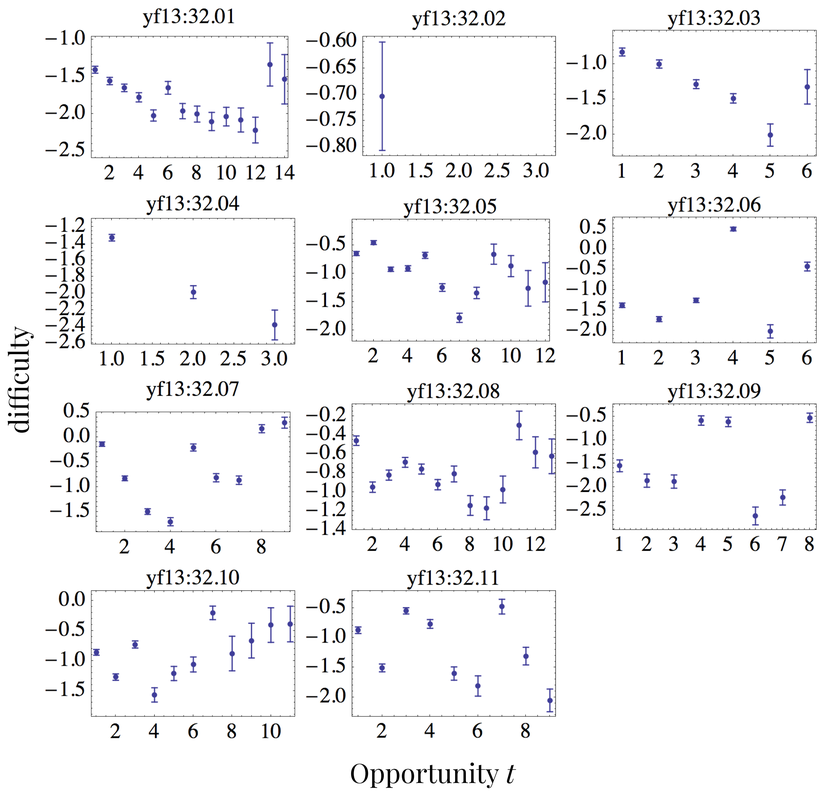
Rasch/KC model:

$P_{s,p}$ is the probability that student *s* gets problem *p* correct

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

θ_s is skill of student s
β_p is difficulty of problem p
$\zeta_{k,t}$ is the difficulty of applying KC k on opportunity t.
(The PFA model restricts $\zeta_{k,t}$ to be linear in t.)

Learning Curves for the same KCs.
Here, difficulty is -logit($P_{\text{correct first try}}$).
This is equivalent to fitting to just $\zeta_{k,t}$.



Conclusions:

- KC labels focus on content knowledge, rather than problem-solving skills. Students may be learning the latter in their homework.
- Learning curves are not like middle-school math:
  - many KCs have few opportunities
  - most curves erratic: *not* monotonically decreasing
- Careful error analysis shows KC learning can be untangled from problem difficulty.
- For the Rasch/KC model, large variance in problem difficulties β_p, with KC difficulties $\zeta_{k,t}$ nearly constant.
  - Student learning not well-explained by these KCs
  - When learning curves work (yf13:32.01, yf13:32.03), IRT-style difficulties may still effectively model student behavior.

Rasch/KC model for these KCs, showing KC difficulty $\zeta_{k,t}$ vs. *t*.
The gaussian on the right represents β_p, the distribution of exercise difficulties for the associated exercises.