## Learning Curve Exercise for DataShop Tutorial ITS2010 June 14, 2010

Ken Koedinger

This exercise has the following goals:

1. Familiarize you with the data set used in Cen, Koedinger, & Junker (2006)[[1]](#footnote-1).
2. Give you an experience in using the Pittsburgh Science of Learning Center’s DataShop, a repository of learning data open for data mining.
3. Give you a brief experience in doing some digging into this data set.

In this exercise, you first need to log into DataShop and review some help pages (part A below), then get to the “Geometry Area (1996-97)” data set (part B below), and then answer some questions about that data set (part C below).

#### Part A. Logging into Datashop and reviewing help pages

1. Go to http://www.pslcdatashop.org
2. Login – you need to register if this is your first time
3. Click on “Help” in the upper left corner. Click on “Documentation Home” at the bottom of the page that appears.
4. Click on “4. Glossary”
5. Scroll down, click on and watch “Video: Solving ‘Making Cans’, Question 1”. This video shows an example of (a more recent version of) the tutor used in the Cen paper and in the data set you will look at in Parts B and C. Note that this problem was called “Pogs” in the 1996-97 dataset, not “Making Cans”.
6. Read the rest of the Glossary pages and be sure to look at the Tables of data. Which rows in Table 1 correspond with the actions shown in the video?  
   What questions do you have? Please ask!

#### Part B. Getting to the Geometry Area (1996-97) data set

1. Click “Back to DataShop” (or get yourself to where you were after step 2 in part A).
2. Under Public Datasets, click on “Geometry Area (1996-97)”, which is under the heading “Geometry Area 1996” about halfway down the page. Click on Learning Curve at the top.
3. On the top left in the Samples section, click on All Data. You may have to wait a moment, but you should see a learning curve appear.
4. Find the “KC Models” section on the left side, third panel down. Pick a different knowledge component model (e.g., “Original”) from the “Primary” menu and inspect the learning curve that appears. The red solid line shows the data and the blue dotted line shows predictions based on the KC model (more later). Clicking on a specific red point will reveal the updated point info below. Also notice the “observation table” below the graph. Try some other KC models (e.g., Geometry, DecomposeArith, etc.) and try to make sense of the differences in the number of observations per opportunity, in the “smoothness” of the curves, and the closeness of fit of model (blue) and data (red). These different models are just some of the states in the huge space of many possible knowledge component models that the LFA algorithm described in the Cen paper you can search through.
5. Scroll down and note that you can click on individual knowledge component learning curves and they are brought up into the large display.
6. In the “KC Models” panel set the Primary model to “Original”.
7. Click on “LFA values” subtab (below “Error Report”) to see the best fitting parameters for the current Knowledge Component model. The Student and KC intercept values are just like the values you would get from an Item Response Theory model (though KC replaces item). The key difference is the addition of the slope parameter on the KC’s, which determines how fast the learning curve goes down.
8. Click back on Line Graph to see the learning curves again.

**Part C. Do some data mining!**

Write short answers to as many of the following questions as you can.

### Learning Curve tool

1. Why does the number of observations per opportunity change when you switch from one KC model to another? Hint: There is only one KC in the Geometry KC model and fifteen in the Original model.
2. For which of the knowledge component models does the learning curve appear to go down?
3. Why do the KC models for some curves not go down? Why is there no apparent reduction in error rate for some of these KC models? Hint: It is not because there isn’t learning – students learned from this unit, but only some KC models reveal it.

For the following questions, look at individual knowledge component learning curves for the Original KC model.

1. What is one KC that shows a relatively smooth dropping curve?
2. What is one KC that increases in error rate? Why might this curve be going up?
3. Do you see any opportunities for tutor redesign by looking at these individual KC learning curves?
4. Is there a KC with a curve that had a low error rate from the start and yet students received lots of practice on that KC?
5. Is there a KC with a curve that had a high error rate at the start, but students received too few practice opportunities such that the error rate was still high at the end of practice?
6. What is an example of a KC where the LFA model (the blue line) does not seem to fit the learning curve data (the red points) very well? What might be going on?

*Questions about Performance Profiler tool (click on Performance Profiler at top of page)*

1. Which problem did students have the most difficulty with? How did you figure it out? Only count problems that a reasonable number of students attempted. You can put the mouse over bars to get relevant information. Better yet you can set the minimum number of students, say to 5, in the the navigation area.
2. Which problem is easiest?
3. Which student made the fewest errors? (Do not count students who only solved a few problems.) To see data on students, click on Problem (which appears sideways on the left axis of the graph) and select Student in the pop-up menu that appears.
4. Which student made the most errors? (Do not count students who only solved a few problems.)
5. Make sure you have Original selected in the KC Models panel on the left. For what knowledge component was the overall error rate the highest? (As in question 3, you can click on the left axis Problem or Student and change it to Knowledge Component.)
6. Change the left axis back to Problem and change the top axis from Error Rate (%) to “Residual Error Rate % (Predicted – Actual)”. This variable shows the difference between the LFA model’s prediction and the actual data. Go to the Performance Profiler panel on left and change Sort By from Error Rate to Residuals. For which problem does the Original KC model predict a much lower error rate than the observed? (Hint: That is, which problem has the biggest negative residual? As before, remember to ignore problems with too few data points.)
7. For which problem does the Original KC model predict a much higher rate than observed? (Hint: That is, which problem has the biggest positive residual?)
8. For the problem identified in question 6, what are the knowledge components that are accounting for the poor fit on this problem? To find out what knowledge components go with a problem, click on Error Report at the top and then select the Problem in the panel on the lower left. You will see a display of the error rate on each step in that problem and the knowledge component(s) associated with each step.
9. For the problem identified in question 7, what are the knowledge components that are accounting for the poor fit on this problem?

*Questions to address using Export Step Rollup Table.*

1. Why does the learning curve for the Geometry KC model have big upward blips in the error rate at some opportunities like 28, 38, and 42? For instance, between opportunity 27 and 28, the error rate jumps up from 8% to 50%. To answer this question, click on Student-Step Rollup underneath Learning Curve, check Knowledge Components in the Step Rollup panel on the left and change the Knowledge Component model to Geometry. Then click on Export Step Rollup. Load the file you get into Excel. Inspect the steps, for instance, with opportunity numbers 27 and 28 using Excel’s AutoFilter or, better yet, a pivot table.  
   1. To get a lesson on pivot tables, go back to the learnlab.org main page, click Enabling Technologies, click Meetings, and view the Documents and Video of Presentation.

1. This paper can be found by following steps A1, A2, and B2 and then clicking on the “Dataset Info” tab (top-left) and then “Papers and Files” (second from left). Finally, click on the link in the “Paper” column. [↑](#footnote-ref-1)