

# Infrastructure for Innovation via Data- *and* Model-Driven Design

Ken Koedinger

Professor of Human-Computer Interaction & Psychology  
Carnegie Mellon University

Director of  **LearnLab**  
Pittsburgh Science of Learning Center

Co-leader of

**Carnegie  
Mellon  
University**

The  
Simon  
Initiative

PI of



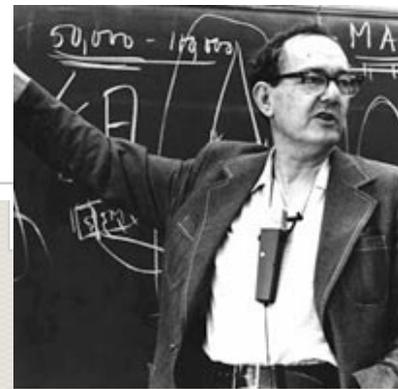
Computer-Aided Personalized Education Workshop  
Nov 12, 2015

- Do you know whether or not your students are learning (e.g., while listening to your lecture)?
- Do you know when *you* are learning (e.g., while listening to a lecture)?

# Infrastructure should support *an engineering process*

- Should we "Build into tools the latest results of how people learn"?
- No!
- Build tools for an *iterative engineering process*
  - Support using theory, *new* data, & modeling to design better learning

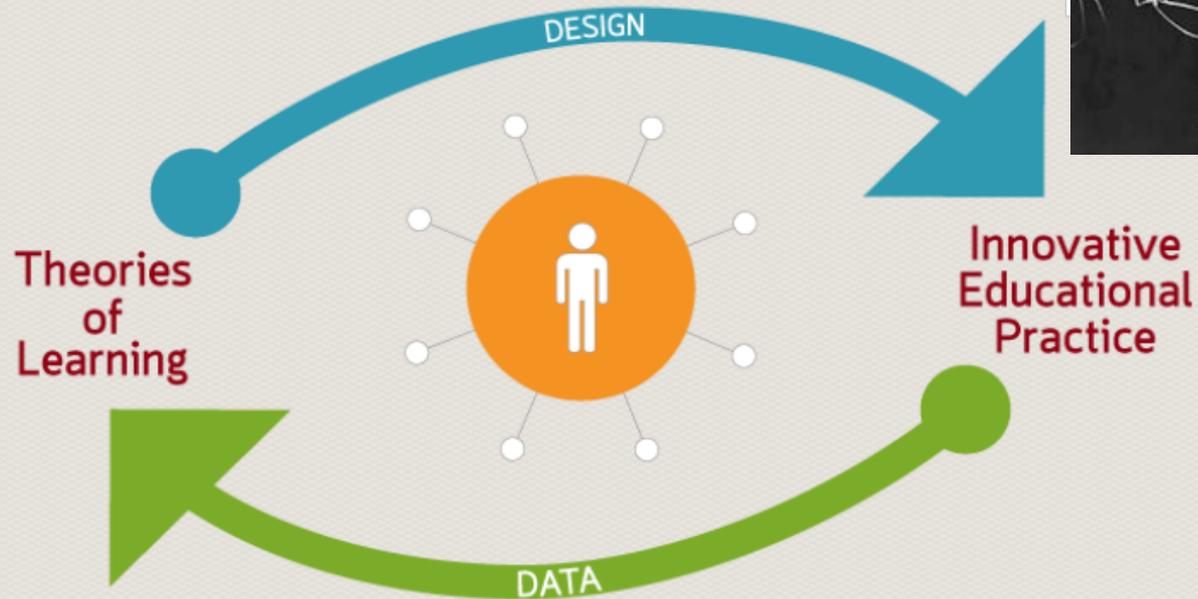
# Simon Initiative: [www.cmu.edu/simon](http://www.cmu.edu/simon)



**Carnegie  
Mellon  
University**

The  
Simon  
Initiative

About  
News  
Events  
Simon DataLab  
Global Learning  
Council



| [THE SIMON INITIATIVE](#) |

## The Simon Initiative Vision

A data-driven virtuous cycle of learning science research and innovative educational practice causes demonstrably better learning outcomes for students from any background or place.

### Recent News

[GLC Releases Draft White Paper: Addresses Key Challenges in Technology-Enhanced Learning. Data Sharing](#)

April 20, 2015

# Why is a data infrastructure & engineering process so critical?

- We don't know what we know
- We can't see learning
- Principles of learning are not general

# Adaptive Technology: Not new, but can do better

Not new:

e.g., Cognitive Tutors have provided adaptive instruction since the mid 80s  
... iStart, Andes,

Not done: Plenty room for improvement

Quality of personalization *depends* on quality of model & data used to produce it

# Cognitive Tutors

## Adaptive learning in schools since 90s

My current cell phone company charges me \$14.95 per month for service and \$.13 per minute. PPS Cellular Phone Company has offered me \$15.00 worth of free calls a month if I switch, but the charge is \$.39 per minute.

Quantity Name	Time	Current cost
Unit	minutes	\$
Expression	$t$	$.13t$
Question 1		
Question 2		
Question 3		
Question 4		

The cost from my current company increases by 0.13 each minute, but remember that it starts at 14.95 dollars.

Authentic problems

Feedback *within* complex solutions

Progress...

Personalized instruction

Challenging questions

4. After how many minutes of calls will the cost for both companies be the same?

... individualization

The screenshot shows the Cognitive Tutor Algebra I interface. It includes a 'Scenario' window with a word problem, a 'Worksheet' window with a table for variables and questions, a 'Grapher' window with a coordinate plane and input fields, and a 'Hint' window with a step-by-step guide. The interface also features a toolbar with 'Solve', 'Glossary', 'Hint', 'Done', and 'Skills' buttons.

**Scenario:** My current cell phone company charges me \$14.95 per month for service and \$.13 per minute. PPS Cellular Phone Company has offered me \$15.00 worth of free calls a month if I switch, but the charge is \$.39 per minute.

- How many minutes of calls can I get from PPS Cellular Phone Company for \$50? What is the cost from my current company for that number of minutes?
- How many minutes of calls can I get from my current company for fifty dollars? What is the cost from PPS Cellular Phone Company for that number of minutes?
- What is the cost from both companies for sixty minutes?
- After how many minutes of calls will the cost for both companies be the same?

**Worksheet:**

Quantity Name	Time	Current cost
Unit	minutes	\$
Expression	$t$	$.13t$
Question 1		
Question 2		
Question 3		
Question 4		

**Grapher:** A coordinate plane with x-axis from 0 to 5 and y-axis from 0 to 10. Input fields for x and y are present.

**Hint:** If the cost from my current company and the cost from PPS Cellular Phone Company are equal, then their expressions are equal. Write an equation and solve it to find the number of minutes.

**Legend:** Enter Label, Enter Equation

**Equations:**  $y =$  Enter Equation

**Individualization Steps:**

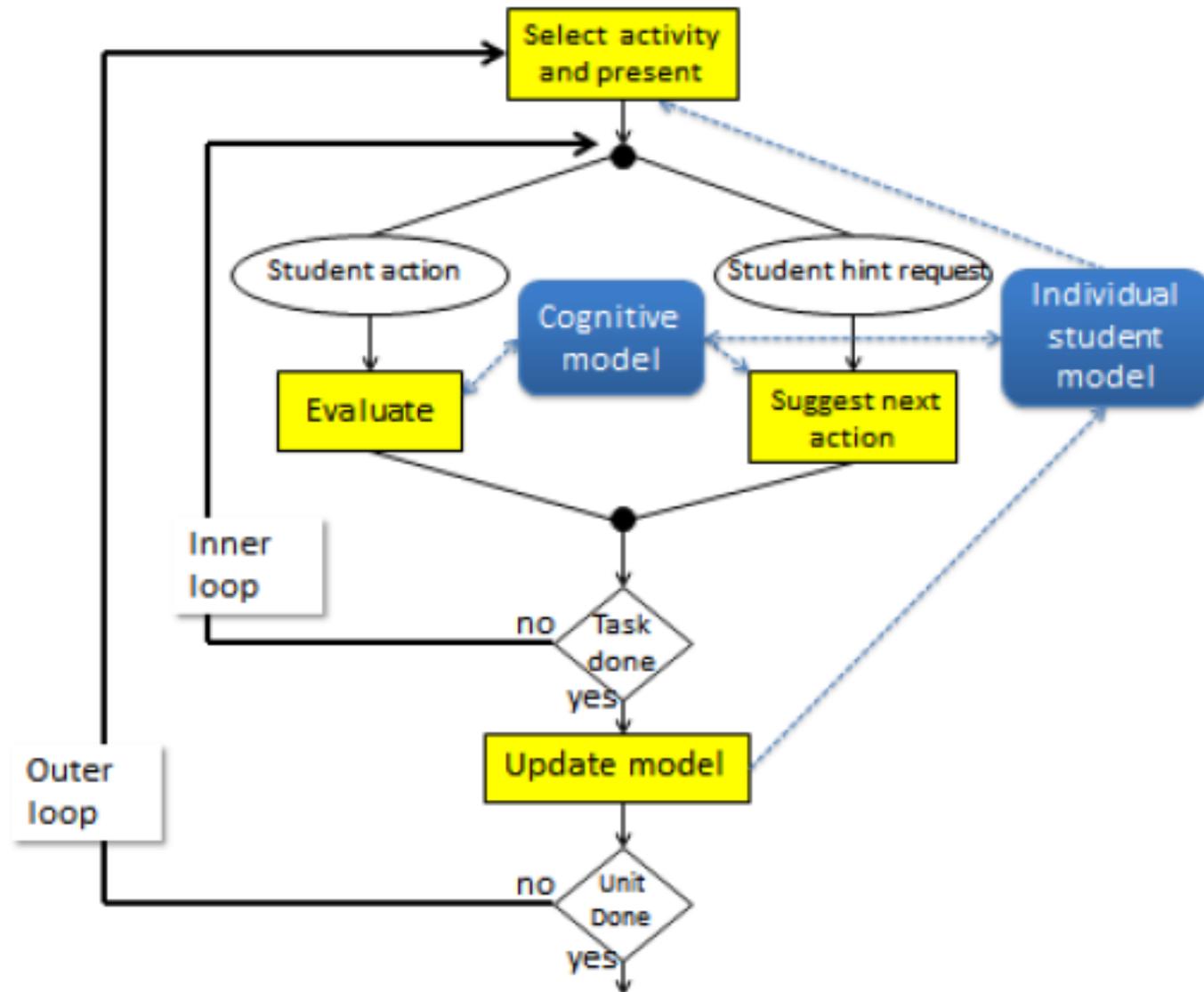
- Calculate input value.
- Writing expression, any form.
- Set axis bounds.
- Label point of intersection.
- Enter given.

# Loops of data-responsive adaptation

Step loop:  
Support  
student paths

Task loop:  
Select what's  
needed

Design loop:  
Design for  
needs of all



# Adapt to “hidden skills” that *all students* find difficult

Which is harder for algebra students?

*Story Problem*

As a waiter, Ted gets \$6 per hour. One night he made \$66 in tips and earned a total of \$81.90. How many hours did Ted work?

*Word Problem*

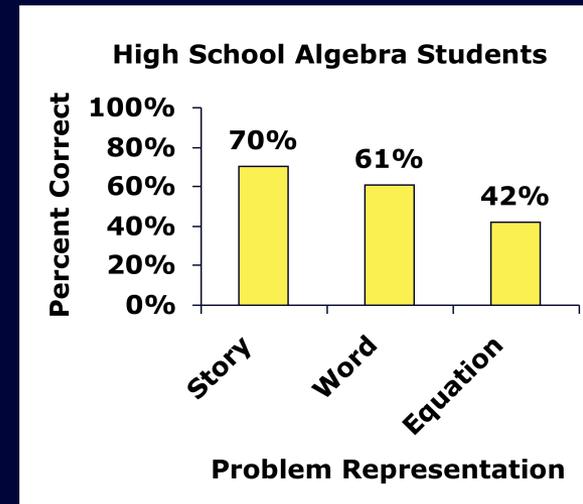
Starting with some number, if I multiply it by 6 and then add 66, I get 81.90. What number did I start with?

*Equation*

$$x * 6 + 66 = 81.90$$

Math educators say:  
story or word is hardest

Students:  
equations  
are hardest



*Expert blind spot!*

Algebra teachers, especially,  
incorrectly think equations are easy

# Collect data to check your assumptions!

## Problem

Ann is in a rowboat on a lake. She is 800 yards from the dock. She then rows for  $m$  minutes back towards the dock. Ann rows at a speed of 40 yards per minute. Write an expression for Ann's distance from the dock.

## Solution   %Correct\*

800-40m   40%

Original  
symbolization

# Elements of infrastructure for sharing innovations

What should we share?

- Data
- Analytic methods
- Explanatory models of learners
  - Models of cognition, metacognition, motivation, self-regulated learning, dialogue, social interaction
- Ed tech
  - generates data, uses models

Li, Matsuda, Cohen, & Koedinger (2015). Integrating representation learning and skill learning in a human-like intelligent agent. *Artificial Intelligence*.

# Formal models of learning can make non-obvious predictions about hidden skills

- SimStudent combines
  - *Representation* learning (probabilistic grammars)
  - *Skill* learning mechanisms
- Novel prediction

Step
$2x=12$
$6=3x$
$-x=5$
$-24=-4x$



# A community data infrastructure to support online learning improvement.



## Existing Resources



Send us data, models, data-generating tools (or links to such)!

# Some existing review articles

- Data-driven ITS development
  - Koedinger, *Brunskill*, Baker, McLaughlin, Stamper (2013). New potentials for data-driven intelligent tutoring system development and optimization. *AI Magazine*.
- Educational Data Mining
  - Koedinger, *D'Mello*, McLaughlin, Pardos, & Rosé (2015). Data mining and education. *WIREs Cognitive Science*
- Learning science results
  - Pashler et al. (2007). Organizing Instruction and Study to Improve Student Learning (NCER 2007-2004).
  - Ambrose et al. (2010). *How learning works: Seven research-based principles for smart teaching*.
  - Clark & Mayer (). E-learning
- Toward a learning engineering process
  - Koedinger, Corbett, & Perfetti (2012). The Knowledge-Learning-Instruction (KLI) framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*.
  - Koedinger, Booth, & Klahr (2013). Instructional complexity and the science to constrain it. *Science*.

# Thank you!



<http://learnlab.org>



<http://cmu.edu/simon>

**Carnegie  
Mellon  
University**

The  
Simon  
Initiative

<http://learnsphere.org>



Ken Koedinger  
koedinger@cmu.edu

