1. As you listen to a lecture or read a text, do you know when you are learning and when you're not?
   a. I know when I'm learning
   b. I don't know when I'm learning
2. As you give a lecture, do you know when your students are learning and when they are not?
   a. I know when my students are learning as I lecture.
   b. I don't know when my students are learning as I lecture.
3. To effectively and efficiently redesign a course that produces better student outcomes, one should:
   a. follow your intuitions
   b. apply learning science
   c. do learning science
4. Does answering an in-video question correctly mean the student has learned the associated concept or skill?
   a. Yes, a correct answer to an in-video question does mean the student has learned.
   b. No, a correct answer to an in-video question does not mean the student has learned.
Practical Learning Research at Scale

Ken Koedinger
Professor of Human-Computer Interaction & Psychology
Carnegie Mellon University

Director of LearnLab
Co-leader of The Simon Initiative

PI of LearnSphere

Learning at Scale, April 26, 2016
Overall Argument

- You cannot see learning! Beware illusions!
  - Students watching lectures
  - Instructors watching students
- Data breaks illusions
  - Students need feedback
  - Instructors need assessment *data*
- Challenges for practical learning science
  - We don’t know what we know
  - Trillions of ways to learn;
    *1 size does not fit all*
- Ed science & technology is revolutionary
  - it spreads access, personalizes, but mostly
  - it provides data to drive *iterative* improvement
Outline

• Learning Science knowns & unknowns
  – Illusions, desirable difficulties, cognitive load
  – Cannot reliably apply LS principles across courses

• What to do: Scale science through practice
  – Social-technical infrastructure to support iterative engineering & practice-relevant theory

• Scaling deep content analytics

• Scaling iterative course improvement
Do students know when they are learning? Can we see it?

Simple answer: NO!

We experience *illusions of knowing and learning*

- Students are poor at judging what they learned
  - Low correlation of students’ estimates of learning vs. actual test scores (*Eva et al. 2004, Jee et al., 2006*)

- Liking is not learning
  - Low correlation (~.15) between students’ course ratings with after-training skills (*Sitszmann et al., 2008*)
    – Observations of engagement or confusion are not strongly predictive of learning outcomes

- WHY? Students have insufficient mental resources to learn and monitor their own learning (*Moos & Azevedo, 2008*)
What do we know about what works in supporting learning?

Education wars in public dialogue
- Basics vs. understanding
- Education wars in reading, math, science...

Progress with unknowns
- Spaced practice > massed (Pashler)
  **but** massed sometimes better (Pavlik)
- “Tests” (retrieval practice) > example study (Roediger)
  **but** worked examples > pure practice (Sweller)
- Direct instruction > discovery learning (Klahr)
  **but** active learning > lecture (PNAS 2016)
- Other debates ...
  - Explaining is good v. bad (Chi v. Williams)
  - Interleaving v. blocking (Rohrer v. Carvalho)

Cognitive load vs. desirable difficulties

How big is the design space?

Focused practice
- Study examples
- Test on problems
  - Concrete
    - Block topics in chapters
    - Fade
    - Explain
  - Mix
    - Interleave topics
    - Ask for explanations
- Immediate
- Delayed
- No feedback

Gradually widen
- Study 50/50
- Test
  - Concrete
    - Block topics in chapters
    - Fade
    - Explain
    - Mix
    - Ask for explanations
  - Abstract
    - Interleave topics
    - Delayed
  - Mix
    - No feedback

Distributed practice
- Study 50/50
- Test
- Concrete
- Abstract
- Mix
- Immediate
- Delayed
- No feedback

What's best?
- More help
- Basics
- More challenge
- Understanding

Many other choices: animations vs. diagrams vs. not, audio vs. text vs. both, …

$3^{15} \times 2 = 205$ trillion options!

If cross-course experiments on the 30 dimensions produced consistent results, we could *just apply* learning science.

How to run enough experiments? Are results consistent?
Cognitive Tutors: Adaptive Support for Learning by Doing

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.

1. 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?

2. 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?

3. What is the cost from both companies for sixty minutes?

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

Authentic problems

Feedback *within* complex solutions

Progress...

Personalized instruction

... individualization
Scaled applications of learning by doing *with feedback*

**K12 Math Cognitive Tutors**
- Widely used
  - ~500K students per year
  - ~80 minutes per week
- **2x better learning**

**College Online Courses**
- Widely used
  - ~50K-1M students per year
  - ~30 courses at 1K colleges
- **2x faster learning**

Pane et al. (2013). Effectiveness of Cognitive Tutor Algebra I at Scale. Santa Monica, CA: RAND Corp.

Social cyberinfrastructure for doing science *in practice*

Ed tech + wide use = “Basic research *at scale*”

> 360 *in vivo* experiments

> 820 ed tech data sets in DataShop

NSF: $47M, 2004-15
Lessons from 360 cross-domain *in vivo* experiments => KLI Framework

1. Ideal instruction depends on knowledge goals
2. Because *different learning processes* are at work
3. Makes sense of competing recommendations

### Table 2. Recommendations and corresponding Level of Evidence to support each

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Level of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Space learning over time. Arrange to review key elements of course content after a delay of several weeks to several months after initial presentation.</td>
<td>Moderate</td>
</tr>
<tr>
<td>2. Interleave worked example solutions with problem solving exercises. Have students alternate between reading already worked solutions and trying to solve problems on their own.</td>
<td>Moderate</td>
</tr>
<tr>
<td>5. Use quizzing to promote learning. Use quizzing with active retrieval of information at all phases of the learning process to exploit the ability of retrieval directly to facilitate long-lasting memory traces.</td>
<td>5a. Low</td>
</tr>
<tr>
<td>5a. Use pre-questions to introduce a new topic</td>
<td>5b. Strong</td>
</tr>
<tr>
<td>5b. Use quizzes to re-expose students to key content</td>
<td></td>
</tr>
<tr>
<td>7. Ask deep explanatory questions. Use instructional prompts that encourage students to pose and answer “deep-level” questions on course material. These questions enable students to respond with explanations and supports deep understanding of taught material.</td>
<td>7. Strong</td>
</tr>
</tbody>
</table>
Learn by doing or by studying?

- **Testing effect** (e.g., Roediger & Karpicke, 06)
  - “Tests enhance later retention more than additional study of the material”

- **Worked example effect**
  - “a worked example constitutes the epitome of strongly guided instruction”

- **Theory**
  - Testing produces “desirable difficulties”
  - Worked examples reduce “cognitive load”

- **What’s right?**
  And what’s the mechanism?
Testing effect: Sample instruction

<table>
<thead>
<tr>
<th>Study</th>
<th>Assistance</th>
<th>Test</th>
<th>Practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Students</td>
</tr>
<tr>
<td>Difficulty/Load</td>
<td></td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Example</td>
<td></td>
<td></td>
<td>Instructors</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lao3shi1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-&gt; teacher</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A -&gt; B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A -&gt; ?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Testing effect: Typical result

- **Example**
  - High Assistance
  - Low Difficulty/Load

- **Test**
  - Low Instructors
  - High Practice

- **Study**
  - Low
  - (SSST)

- **Test**
  - High
  - (STTT)

Roediger & Karpicke, 06
Testing effect vs. worked examples

Instructors Assistance High Low

Students Difficulty/Load Low High

Example

Test

Study

Practice

Testing Effect

Study

lao3shi1

-> teacher

Test

lao3shi1

-> ?

Solve

(a+b)/c = d

->

(a+b)/c = d

a+b = dc

a = dc - b
Testing effect vs. worked examples: Conflicting results!

High Assistance Low Instructors
Example Test
Low Difficulty/Load High Students
Study Practice

Testing Effect
Study (SSST) 57%
Test (STTT) 64%

Roediger & Karpick, 06
Is the worked example effect a result of weak controls?

- Prior studies: students solve whole problems & get whole solution feedback
- Cognitive Tutors provide step-by-step support
  - Might they reduce cognitive load & make worked examples unnecessary?
Practice
Students solve problems step-by-step & explain

Given is circle A with arc BD.
If the measure of arc BD is 32°, what is the measure of arc BFD?

m Arc BFD = [Blank]
Rule = [Blank]
Worked out steps with calculation shown by Tutor

Worked examples:
Half of steps are given as examples

Student still has to self explain worked out step
Worked examples improve efficiency & understanding

Lab results
- 20% less time on instruction
- Conceptual transfer in study 2

In Vivo
- Adaptively fading examples to problems yields better long-term retention & transfer


Worked example effects generalize, limit generality of testing effect

- Geometry tutor studies (just discussed)
- Chemistry tutor studies *in vivo* (McLaren et al.)
  - Same outcomes in 20% less time
- Algebra Tutor study *in vivo* (Anthony et al.)
  - Better long term retention in less time

- Theory: SimStudent model (Matsuda et al.)
  - Examples: basis for inducing patterns
  - Problems: negative examples to prune misconceptions

KLI explains discrepancy: Target **Knowledge** => **Learning process needed** => optimal **Instruction**

<table>
<thead>
<tr>
<th>Processes (simpler on bottom)</th>
<th>Understanding &amp; sense making</th>
<th>Induction &amp; refinement</th>
<th>Memory &amp; fluency building</th>
</tr>
</thead>
<tbody>
<tr>
<td>Many examples support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing recall supports</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Knowledge Components</th>
<th>Facts (constant - constant KCs)</th>
<th>Rules (variable condition KCs)</th>
<th>Principles (verbal KCs with a rationale)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Testing effect</strong></td>
<td>Facts must be memorized</td>
<td>rules (&amp; instances) must be remembered (dual paths may help)</td>
<td>principles must be remembered, but can be reconstructed</td>
</tr>
<tr>
<td><strong>Testing effect</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Worked examples</th>
<th>Worked examples</th>
<th>+ + + principles can be inert without associated rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ - nothing to explain, learning is more efficient</td>
<td>- no generalization needed</td>
<td>+ rules must be induced</td>
</tr>
</tbody>
</table>

Outline

• Learning Science knowns & unknowns
  – Illusions, desirable difficulties, cognitive load
  – Cannot reliably apply LS principles across courses

• What to do: Scale science through practice
  – Social-technical infrastructure to support iterative engineering & practice-relevant theory

• Scaling deep content analytics
• Scaling iterative course improvement
Is this what we should be doing?

THEORIES OF LEARNING

Apply learning science

PRACTICE

INSTRUCTIONAL DESIGN
Simply Applying Learning Science *Is Not Enough!*

- Only ~10% of US Dept of Ed funded randomized controlled trials (RCT) find positive effects
  - 11/90 (or 7/77) IES RCTs succeeded
    - See 2013 review at Coalition4evidence.org
- Further “learning scientists” don’t agree
  - especially those from different disciplines Cog Neuro, Psychology, Ed Psych, Ed …
Can’t just *apply* learning science,

must *do* learning science
Data-driven iterative engineering

Contribute use-inspired basic research

Design explanatory models of learners

Use insights from data to redesign

Use models to design instruction

THEORIES OF LEARNING

MODEL DESIGN

MODEL OF EXPERTISE

PRACTICE

INSTRUCTIONAL DESIGN

DISCOVERY

… design, deploy, data, discover, design, deploy … [repeat]
Social cyberinfrastructure for doing science in practice

Ed tech + wide use = “Basic research at scale”

Researchers

Schools

Learn Lab

> 360 in vivo experiments

> 820 ed tech data sets in DataShop
Every school & university should be a LearnLab!

- LearnLab was an experiment in building a cyberinfrastructure but funding is over
- Need more of these ...
  - See 5 recommendations in
  - See Global Learning Council report


http://globallearningcouncil.org/documents/
Three avenues to scaling practical learning research

1. Make every university a LearnLab
   a. Foster a culture of iterative improvement
   b. Share assessments, data & analytics

2.

3.
Outline

• Learning Science knowns & unknowns
  – Illusions, desirable difficulties, cognitive load
  – Cannot reliably apply LS principles across courses

• What to do: Scale science through practice
  – Social-technical infrastructure to support *iterative engineering* & practice-relevant theory

• Scaling deep content analytics

• Scaling iterative course improvement
Cognitive Task Analysis
Most powerful way to improve a course!

• Goal: Identify underlying cognitive processes students need to learn
  – Creates an accurate cognitive model

• Many methods
  – Think alouds, structured interviews, computational modeling

• Courses modified by CTA produce much better learning
  – 1.5 sd effect size! (Clark et al)
  – Why? 70% of expertise is tacit!

Improves learning of catheter insertion
(Velmahos, Clark, et al., 2004)
Social cyberinfrastructure for doing science in practice

Ed tech + wide use = “Basic research at scale”

LearnLab
Pittsburgh Science of Learning Center

Researchers

Schools

NSF $47M, 2004-15

> 360 in vivo experiments

> 820 ed tech data sets in DataShop

Statistics Online
Algebra Cognitive Tutor
English Grammar Tutor
Educational Games

> 360 in vivo experiments
> 820 ed tech data sets in DataShop
Quantitative Cognitive Task Analysis

• A accurate Cognitive Model should produce a “smooth learning curve”
Learning Curves

Dataset: Geometry Area (1996-97)
Sample(s): All Data

All Selected Knowledge Components

Error Rate (%)

opportunity

All Data

<table>
<thead>
<tr>
<th>Opportunity Number</th>
<th>Number of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>59</td>
</tr>
<tr>
<td>2</td>
<td>59</td>
</tr>
<tr>
<td>3</td>
<td>59</td>
</tr>
<tr>
<td>4</td>
<td>58</td>
</tr>
<tr>
<td>5</td>
<td>56</td>
</tr>
<tr>
<td>6</td>
<td>48</td>
</tr>
<tr>
<td>7</td>
<td>47</td>
</tr>
<tr>
<td>8</td>
<td>45</td>
</tr>
<tr>
<td>9</td>
<td>42</td>
</tr>
<tr>
<td>10</td>
<td>42</td>
</tr>
<tr>
<td>11</td>
<td>41</td>
</tr>
<tr>
<td>12</td>
<td>41</td>
</tr>
<tr>
<td>13</td>
<td>40</td>
</tr>
<tr>
<td>14</td>
<td>39</td>
</tr>
<tr>
<td>15</td>
<td>39</td>
</tr>
<tr>
<td>16</td>
<td>39</td>
</tr>
<tr>
<td>17</td>
<td>39</td>
</tr>
<tr>
<td>18</td>
<td>39</td>
</tr>
<tr>
<td>19</td>
<td>39</td>
</tr>
<tr>
<td>20</td>
<td>39</td>
</tr>
<tr>
<td>21</td>
<td>39</td>
</tr>
<tr>
<td>22</td>
<td>39</td>
</tr>
<tr>
<td>23</td>
<td>39</td>
</tr>
<tr>
<td>24</td>
<td>39</td>
</tr>
<tr>
<td>25</td>
<td>39</td>
</tr>
<tr>
<td>26</td>
<td>39</td>
</tr>
<tr>
<td>27</td>
<td>39</td>
</tr>
<tr>
<td>28</td>
<td>39</td>
</tr>
<tr>
<td>29</td>
<td>39</td>
</tr>
<tr>
<td>30</td>
<td>39</td>
</tr>
<tr>
<td>31</td>
<td>39</td>
</tr>
<tr>
<td>32</td>
<td>39</td>
</tr>
<tr>
<td>33</td>
<td>39</td>
</tr>
<tr>
<td>34</td>
<td>39</td>
</tr>
<tr>
<td>35</td>
<td>39</td>
</tr>
<tr>
<td>36</td>
<td>39</td>
</tr>
<tr>
<td>37</td>
<td>39</td>
</tr>
<tr>
<td>38</td>
<td>39</td>
</tr>
<tr>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td>40</td>
<td>39</td>
</tr>
</tbody>
</table>
Cognitive Task Analysis using DataShop’s learning curve tools

Without decomposition, using just a single “Geometry” KC, no smooth learning curve.

But with decomposition, 12 KCs for area concepts, a smoother learning curve.

Upshot: Can automate analysis & produce better cognitive models
Statistical version of cognitive model: Generalization of item response theory

**GIVEN:**
- $p_{ij} =$ probability student $i$ gets step $j$ correct
- $Q_{kj} =$ each knowledge component $k$ needed for this step $j$
- $T_{ik} =$ opportunities student $i$ has had to practice $k$

**ESTIMATED:**
- $\theta_i =$ proficiency of student $i$
- $\beta_k =$ difficulty of KC $k$
- $\gamma_k =$ gain for each practice opportunity on KC $k$

*Cen, Koedinger, & Junker (2006)*
*Draney, Pirolli, & Wilson (1995)*
*Spada & McGaw (1985)*
Visualizing learning curves to find opportunities for improvement

High rough curve => concept/skill is more complex
Labeled steps vary in error rate =>
Hypothesized skill is wrong =>
Inspect problems to find new difficulty factors
Same math but different difficulty!? 

To make metal cans, the ends for the cans are stamped out of square pieces of metal. The part of the square that is left over is then recycled as scrap. The 

Hard 

Worksheet 

<table>
<thead>
<tr>
<th>Diagram Label</th>
<th>Unit</th>
<th>radius of the end of the can</th>
<th>length of the square ABCD</th>
<th>Area of the scrap metal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>inches</td>
<td>4</td>
<td>8</td>
<td>13.76</td>
</tr>
<tr>
<td>Question 2</td>
<td>inches</td>
<td>8</td>
<td>16</td>
<td>55.04</td>
</tr>
<tr>
<td>Question 3</td>
<td>inches</td>
<td>12</td>
<td>24</td>
<td>123.84</td>
</tr>
</tbody>
</table>

Easy 

Scenario 

A manufacturing plant makes the bottom of aluminum cans by stamping a circle from a square piece of aluminum. The remaining metal is scrap. The side length of each square piece of aluminum is 5.6 centimeters. The diameter of the can is equal to the side length of the square piece of aluminum.

Use 3.14 for π.

1. What is the area of the scrap metal?

Worksheet 

<table>
<thead>
<tr>
<th>Diagram Label</th>
<th>Unit</th>
<th>Side of the metal square</th>
<th>Area of the metal square</th>
<th>Radius of the bottom of the can</th>
<th>Diameter of the bottom of the can</th>
<th>Area of the bottom of the can</th>
<th>Area of Scrap Metal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>centimeter</td>
<td>5.6</td>
<td>31.36</td>
<td>2.0</td>
<td>5.6</td>
<td>24.6176</td>
<td>6.7424</td>
</tr>
</tbody>
</table>
Discovered Cognitive Model (green line) produces a better fit

Bars (shaded from the left) represent the actual error rate.

Lines show the predicted error rate for KC models.

The original KC model (square points).

A new KC model (round points) in which compose-by-addition is split into 3 KCs. The predictions of these three KCs produce better fits (with some exceptions) for the steps with high (decompose KC), medium (reduced compose-by-addition KC), and low (subtract KC) error rates.
Closing the loop
Data-driven continuous improvement

• *Deploy v1* of online course
  – Use *data* to make *discoveries*
  – Make *design* changes

• *Deploy v2 vs. v1* in randomized controlled experiment
  – Use *data* to evaluate improvement
New problem type to focus instruction on greatest need

- Practice planning step only

The given figure consists of a square and a parallelogram. The base of parallelogram QUAR is 7.5 meters and the height is 2.5 meters.

What is the area of the given figure?

- A. Multiply area of SQRE by the area of QUAR: (7.5 * 7.5)(7.5 * 2.5)
- C. Add the area of SQRE to the area of QUAR: (7.5 * 7.5) + (7.5 * 2.5)
- D. Add together all sides of the figure and multiply by the height of the parallelogram: (7.5 + 7.5 + 7.5 + 7.5) * 2.5
Students learn more in much less time

More efficient mastery: 25% less time!  
And better learning of planning skills
Practice yes, but practice what?

• You might say:
  – “students learn what they spend time on”

• BUT! both groups spent time on problems involving combining formulas
  – In fact, control spent more time on such problems

• What matters:
  – *Not* practice time on a “topic”
    but practice on *difficult cognitive skills*
  – *Not* the act of doing per se
    but *context in which a decision to act is made*

• *Well-engineered* multiple choice is much better than *intuitively designed* “active” “hands-on”
Scale CTA using theory & computational modeling

• Vision: Computational models of human learning that could
  – Automatically generate accurate cognitive models of students & misconceptions
  – Serve as a “crash test” to reliably evaluate course changes before implemented with real students
    Li, Cohen, & Koedinger (2012). Problem order implications for learning transfer. *Intelligent Tutoring Systems*
    Matsuda, Cohen, & Koedinger (2015). Teaching the Teacher: Tutoring SimStudent leads to more Effective Cognitive Tutor Authoring. *International Journal of Artificial Intelligence in Education*
Is it obvious what the cognitive components are? Which problem is hardest?

<table>
<thead>
<tr>
<th>Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>2x = 12</td>
</tr>
<tr>
<td>6 = 3x</td>
</tr>
<tr>
<td>-x = 5</td>
</tr>
<tr>
<td>-24 = -4x</td>
</tr>
</tbody>
</table>
SimStudent: A computational theory of skill induction

- Learning by *induction* from examples & doing
  - Tacit learning of representations & skills
  - No verbal instruction involved

**Diagram:**
- **Learning System**
  - Representation Learning
  - Skill Learning
    - Perceptual Learner
    - Feature Test Learner
    - Operator Function Sequence Learner

- **Performance System**
  - What?
    - Perceptual Representation Hierarchy
  - Production Rule
    - If
      - Where?
        - Generalized Information Finding Paths
    - When?
      - Feature Tests
    - Then
      - How?
        - Operator Function Sequence

**Flow Diagram:**
- Exposure to input
- Feedback on errors
- Examples of input-response
- Structures input (WM)
- Selects needed info
- Decides when to act
- Constructs behavioral output from info
SimStudent is cognitive architecture that acquires expertise from exposure to perceptual stimuli (not shown in video) and tutoring with example actions & feedback on its
SimStudent Learning
Fraction Arithmetic
Representation learning through perceptual chunking

- Use Probabilistic Context Free Grammar (pCFG) Induction
  - Underlying structure in the problem → Grammar
- Combined with if-part learning mechanisms, learns better than deep belief networks

Cognitive Models learned by SimStudent are more accurate than human-generated models

**Table 8.1: Number of KCs in SimStudent models and Human-Generated Models.**

<table>
<thead>
<tr>
<th></th>
<th>Human-Generated Model</th>
<th>SimStudent-Discovered Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algebra</td>
<td>12</td>
<td>21</td>
</tr>
<tr>
<td>Stoichiometry</td>
<td>44</td>
<td>46</td>
</tr>
<tr>
<td>Fraction Addition</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Article Selection</td>
<td>19</td>
<td>22</td>
</tr>
</tbody>
</table>

**Table 8.3: CV RMSE on SimStudent-Generated models and Human-Generated Models.**

<table>
<thead>
<tr>
<th></th>
<th>Human-Generated Model</th>
<th>SimStudent-Discovered Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algebra</td>
<td>0.4024</td>
<td><strong>0.3999</strong></td>
</tr>
<tr>
<td>Stoichiometry</td>
<td>0.3501</td>
<td><strong>0.3488</strong></td>
</tr>
<tr>
<td>Fraction Addition</td>
<td><strong>0.3232</strong></td>
<td>0.3343</td>
</tr>
<tr>
<td>Article Selection</td>
<td>0.4044</td>
<td><strong>0.4033</strong></td>
</tr>
</tbody>
</table>
But *insight* from models is more important than prediction

**Which problem is hardest?**

<table>
<thead>
<tr>
<th>Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2x = 12$</td>
</tr>
<tr>
<td>$6 = 3x$</td>
</tr>
<tr>
<td>$-x = 5$</td>
</tr>
<tr>
<td>$-24 = -4x$</td>
</tr>
</tbody>
</table>

*Expert blind spot* caused by lack of conscious awareness of representation learning. Skill decomposition is not obvious!
Three avenues to scaling practical learning research

1. Make every university a LearnLab
   a. Foster a culture of iterative improvement
   b. Share assessments, data & analytics

2. Scale deep content analytics
   a. Use course data & learning theory

3.
Outline

• Learning Science knowns & unknowns
  – Illusions, desirable difficulties, cognitive load
  – Cannot reliably apply LS principles across courses

• What to do: Scale science through practice
  – Social-technical infrastructure to support iterative engineering & practice-relevant theory

• Scaling deep content analytics

• Scaling iterative course improvement
Psych MOOC Analysis: What student choices associate with most learning?

Learning by doing > 6x better than learning by watching!

Koedinger et al. (2015). Learning is Not a Spectator Sport: Doing is Better than Watching for Learning from a MOOC. Proceedings of Learning at Scale.
Do these correlations generalize & are they causal?

• See LAK paper on Friday
  – Differences in doing for same students is predictive across course units
  – Doing is predictive across multiple courses

• Randomized controlled experiments needed
Three avenues to scaling practical learning research

1. Make every university a LearnLab
   a. Foster a culture of iterative improvement
   b. Share assessments, data & analytics

2. Scale deep content analytics
   a. Use course data & learning theory

3. Do process to outcome analytics across courses
Overall Argument

• You cannot see learning! Beware illusions!
  – Students watching lectures
  – Instructors watching students

• Data breaks illusions
  – Students need feedback
  – Instructors need assessment data

• Challenges for practical learning science
  – We don’t know what we know
  – Trillions of ways to learn; 1 size does not fit all

• Ed science & technology is revolutionary
  – it spreads access, personalizes, but mostly
  – it provides data to drive iterative improvement
Thank you!

http://learnlab.org

http://cmu.edu/simon

http://learnsphere.org

Ken Koedinger
koedinger@cmu.edu
1. As you listen to a lecture or read a text, do you know when you are learning and when you're not?
   a. I know when I'm learning
   b. I don't know when I'm learning
2. As you give a lecture, do you know when your students are learning and when they are not?
   a. I know when my students are learning as I lecture.
   b. I don't know when my students are learning as I lecture.
3. To effectively and efficiently redesign a course that produces better student outcomes, one should:
   a. follow your intuitions
   b. apply learning science
   c. do learning science
4. Does answering an in-video question correctly mean the student has learned the associated concept or skill?
   a. Yes, a correct answer to an in-video question does mean the student has learned.
   b. No, a correct answer to an in-video question does not mean the student has learned.
Post-assessment: My answers

http://bit.ly/1VU9qM1

1. As you listen to a lecture or read a text, do you know when you are learning and when you're not?
   a. I know when I'm learning
   b. I don't know when I'm learning

2. As you give a lecture, do you know when your students are learning and when they are not?
   a. I know when my students are learning as I lecture.
   b. I don't know when my students are learning as I lecture.

3. To effectively and efficiently redesign a course that produces better student outcomes, one should:
   a. follow your intuitions
   b. apply learning science
   c. do learning science

4. Does answering an in-video question correctly mean the student has learned the associated concept or skill?
   a. Yes, a correct answer to an in-video question does mean the student has learned.
   b. No, a correct answer to an in-video question does not mean the student has learned.
Submitted abstract

Massive scale education has emerged through online tools such as Wikipedia, Khan Academy, and MOOCs. The number of students being reached is high, but what about the quality of the educational experience? As we scale learning, we need to scale research to address this question. Such learning research should not just determine whether high quality has been achieved, but it should provide a process for how to reliably produce high quality learning. Scaling practical learning research is as much an opportunity as a problem. The opportunity comes from the fact that online courses are not only good for widespread delivery, but are natural vehicles for data collection and experimental instrumentation. I will provide examples of research done in the context of widely used educational technologies that both contribute interesting scientific findings and have practical implications for increasing the quality of learning at scale.
Long abstract

My primary message is how we all, students, instructors, course developers, and scientists, need data to determine if our decisions yield better learning. We cannot just watch. A key proposition of our KLI Framework is that learning is not visible, whether it is our own or our students. We cannot just do. Learning requires information and feedback. We must do, keep score, and compare scores. For students, watching lectures or reading, whether online or off, gives students the illusion of knowing and learning. For instructors and course developers, watching students and collecting their impressions gives us an illusion of effective instruction and of effective course design. These illusions can only be broken through frequent assessment and comparison of assessment results. The best way to find out what principles of learning work best are randomized controlled experiments and analysis across many course contexts. But, other analytic approaches and theory building are critical as well.

My second message is that the science and technology of learning has incredible potential -- what we have now is a Model-T equivalent, but jet airplanes of education are possible. The challenge is huge: There are over 200 trillion different ways to teach and clear evidence that one size does not fit all. Online learning is revolutionary not just because it spreads access and facilitates personalization, but because it provides a means to improvement -- a powerful way to get out on the field, keep score, and compare. We must, however, make sense of such data and build theory.
Extra slides ...
Thank you!

Come e-visit ...

Ed tech data & analytics
- learnlab.org/datashop
- Gradebook, quiz reliability: www.cmu.edu/simon/datalab
- MOOC Analytics: LearnSphere.org

Ed tech courses & authoring tools
- Tutors for math, science, language: ctat.pact.cs.cmu.edu
- Essay grading & dialog tutors: oliwww.cs.cmu.edu/~cprose/Projects.html
- Intro college online courses: oli.cmu.edu

Ken Koedinger
koedinger@cmu.edu
A community data infrastructure to support online learning improvement.

Existing Resources

Why LearnSphere?

World class repository of education data.
Building on DataShop, the world’s largest open repository of transactional data, and MOOCdb, a database design and supporting framework created to harness the vast amounts of data being generated by MOOCs, LearnSphere will integrate existing and new educational data infrastructures to offer a world class repository of education data.

Large distributed data infrastructure
LearnSphere will facilitate a distributed method of data storage and access control. LearnSphere offers a central portal for the sharing, storage, and analysis of public and private datasets. For private datasets, a local storage option allows researchers to share tools and results while maintaining ownership of their data. The distributed data sharing model of LearnSphere allows other researchers to find the data by

Data-driven course design.
LearnSphere will enable new opportunities for learning education researchers, course developers, and instructors to better evaluate causal claims, leading to improved teaching and learning. This data driven course redesign is possible both through better analytics of relational data and through online platform support of controlled experimentation.

Analytics sharing & use
LearnSphere will transform scientific discovery and innovation in education through a scalable data infrastructure designed to enable educators, learning scientists, and researchers to easily collaborate over shared data using the latest tools and technologies. While a standard set of analysis tools allow researchers to quickly start gathering information, users can also leverage user-contributed workflows and
Welcome to DataShop, the world's largest repository of learning interaction data.

Create an account or Log in to start analyzing data.

What can I do with DataShop?

Upload a dataset

Project Add this dataset to ...
new project existing project choose later

Project Name Psychology MOOC data

Data Collection Type
Not specified
Not human subjects data (not original)
Study data collected under an IRB

Dataset Name 2013 Psych

Description (optional)

800+ data sets

math, science, language ...

K12 & college

http://learnlab.org/datashop
Cognitive model
=> adaptive support of learning by doing

• Cognitive Model: A system that can solve problems in the various ways students can

\[3(2x - 5) = 9\]
\[6x - 15 = 9\]
\[2x - 5 = 3\]
\[6x - 5 = 9\]

• Model Tracing: Follows student through their individual approach to a problem -> context-sensitive instruction
**Cognitive model**

=> adaptive support of learning by doing

=> theory-coded data stream

- **Cognitive Model**: A system that can solve problems in the various ways students can

  If goal is solve $a(bx+c) = d$
  Then rewrite as $abx + ac = d$

  **Hint message**: “Distribute $a$ across the parentheses.”

  Known? = 85% chance

  $3(2x - 5) = 9$

  $6x - 15 = 9$

  $2x - 5 = 3$

  **Model Tracing**: Follows student through their individual approach to a problem -> context-sensitive instruction

  **Bug message**: “You need to multiply $c$ by $a$ also.”

  Known? = 45%

- **Knowledge Tracing**: Assesses student's knowledge growth -> individualized activity selection and pacing
Data-driven adaptation in the design loop

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

Open Learning Initiative (OLI)

CMU lead: Norman Bier was Candace Thille
Stanford lead: Candace Thille
Open Learning Initiative (OLI)
30 full highly interactive courses

Open Access & For Credit

- American English Speech
- Arabic for Global Exchange
- Anatomy & Physiology
- Argument Diagramming
- Biochemistry
- Elementary French I
- Elementary French II
- Engineering Statics
- Elementary Spanish I
- Elementary Chinese I
- Evidence-Based Practice in Management and Consulting
- Health Information Technology Foundations
- Introduction to Biology
- Introduction to Chemistry

- Introduction to Psychology
- Introduction to Visual Design
- Logic & Proofs
- Media Programming
- Modern Biology
- NSC Cyber Technology Program
- Principles of Computing
- Probability & Statistics
- Responsible Computing
- Statistical Reasoning
- STEM Foundations
- STEM Readiness
OLI Use Since 2006

Course Use
- Used by 1809 Instructors in 1050 Institutions
- > 1.2 million Independent Learner Enrollments (Registered and Anonymous)

Development
- 44 Academic and 9 CMU service courses
- By 104 contributing Faculty from 55 Institutions

OLI Academic Enrollments

- AY2009
- AY2010
- AY2011
- AY2012
- AY2013
- AY2014

Spring
Fall
Identify Specific Learning Challenges:
Practice Synthesizing and Applying Skills & Knowledge
“Improvement in post-secondary education will require converting teaching from a solo sport to a community-based research activity.”

Herbert Simon
Nobel Laureate & CMU Professor
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.
New professional masters!

www.metals.cs.cmu.edu