Interactive Support for Mathematical Reasoning and Metacognitive Judgments of Learning

Ken Koedinger
Director of the Pittsburgh Science of Learning Center
Human-Computer Interaction & Psychology
Carnegie Mellon University (CMU)
Chasm between learning sciences & educational practice

Dimensions of today’s challenge

- **Empirical dimension**
  - Lots of rigorous principle-testing lab studies
  - Lots of realistic classroom design research
  - *Too few experiments combine both*

- **Theoretical dimension**
  - *Almost as many theories as there are results!*

- **Practical dimension**
  - Great ideas in field, ignored in lab
  - Strong scientific results being ignored in field
Pittsburgh Science of Learning Center (PSLC)
Purpose Statement

Leverage cognitive theory and cognitive modeling to identify the instructional conditions that cause robust student learning
Overview

• Background: Cognitive Tutors & PSLC
• Assistance Dilemma
  – Argument for less
  – Argument for more
• Example-problem dimension -- cognitive & metacognitive issues
Real World Impact of Learning Science

*Cognitive Tutor Algebra* course

- Based on cognitive theory & AI models of student thinking & learning
- Most widely used & evaluated Intelligent Tutoring System
  > 2600 schools
  > 10 full-year field studies demonstrating better student learning

Analyze real world problem scenarios

An experimental aircraft has sunk off the coast of South Africa at a depth of 12,730 feet. The military have located the aircraft and are in the process of raising it to the surface. It is currently 7625 feet below the surface and is being raised at the rate of 165 feet per hour. (Hint: Consider the direction above sea level to be positive)

1. How deep was the aircraft five hours ago?
2. How deep will the aircraft be five hours from now?
3. When did the military start raising the aircraft?
4. When will the aircraft reach the surface?

To write an expression, define a variable for the time from now and use this variable to write a rule for the depth of the aircraft.

Use table, spreadsheet

<table>
<thead>
<tr>
<th>TIME</th>
<th>DEPTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOURS</td>
<td>FEET</td>
</tr>
<tr>
<td>Expression</td>
<td>H</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>-27,019...</td>
</tr>
</tbody>
</table>

Use graphs, graphics calculator

Use equations, symbolic calculator

-7625+185H = -12790

Add 7625
185H = -5,166
Divide by 185
H = -1,033/37

Model tracing to provide context-sensitive Instruction

Tracked by knowledge tracing

Skills

Changing axis bounds
Changing axis intervals
Correctly placing points
Write expression, any form
Find y, any term
Find x, any form
Identifying units
Relating a given
Prior achievement:
Intelligent Tutoring Systems bring learning science to schools

A key PSLC inspiration:
Educational technology as research platform to generate new learning science
Logic of Pittsburgh Science of Learning Center (PSLC)

• Support experimental studies that
  – Test fundamental principles, not whole courses
  – Are internally & externally valid
• Create a theory of “robust learning”
• Leverage technology & computational modeling
A Bet About the Future

• The key to the 21st century university:
• Technology!
• And not because of the direct benefits of technology
• But because of vast data on learning & fast feedback to instructors, designers, administrators, researchers ...
LearnLab: Like a research hospital for learning

- Tech enhanced courses in Science, Math & Language
- Agreements with schools
  - *Instrument* for continuous embedded assessment
  - *In Vivo Experiments* test principles for achieving robust learning
Example *In Vivo* Experiment on “Self-Explanation”

- Self-explanation: Have students explain steps in solutions

- *In vivo* experiments: Tightly controlled principle-testing experiment embedded in a real course

Problem Solving Condition
(Ecological control: Tutor as it was)

Given: \( \angle \text{EC} \) is a right angle.

- \( \text{m} \angle \text{SOR} = 90 \)
- \( \text{m} \angle \text{OSC} = 90 \)
- \( \text{m} \angle \text{SRN} = 45 \)
- \( \text{m} \angle \text{ESR} = 135 \)

Find the measure of \( \angle \text{SRM} \).

Messages
Some reasons dealing with parallel lines are highlighted in the Glossary. Which of these reasons is appropriate?
You can click on each reason in the Glossary to find out more.

Glossary
- Converse of Isosceles Triangle (Theorem)
- Isosceles Right Triangle
- Triangle Sum (Theorem)
- Linear Pair
- Linear Trio
- Parallel Lines --- Corr. Angles Are Congruent
- Parallel Lines --- Alt. Int. Angles Are Congruent
- Parallel Lines --- Alt. Ext. Angles Are Congruent
- Parallel Lines --- Int. Angles on the Same Side Are Congruent

If two parallel lines are intersected by a transversal, then alternate interior angles are congruent.

Example: \( L_1 \) and \( L_2 \) are parallel lines, intersected by transversal \( T \). \( \angle 1 \) and \( \angle 2 \) are alternate interior angles. If \( m \angle 1 = 37^\circ \), then \( m \angle 2 \) is also \( 37^\circ \).
Problem solving answers

Explanation by reference

Some reasons dealing with parallel lines are highlighted in the Glossary. Which of these reasons is appropriate?

You can click on each reason in the Glossary to find out more.

If two parallel lines are intersected by a transversal, then alternate interior angles are congruent.

Example: \( L_1 \) and \( L_2 \) are parallel lines, intersected by transversal \( T \). \( \angle 1 \) and \( \angle 2 \) are alternate interior angles. If \( m\angle 1 \) is 37°, then \( m\angle 2 \) is also 37°.
Self-explanation yields better transfer

![Graph showing the comparison between problem solving and explanation conditions on post-test performance.](image)
The “Assistance Dilemma”

Instructional designer’s dilemma:

When should instruction *provide* students with assistance

vs.

When should it *withhold* assistance & elicit student knowledge construction?

- Fundamental unsolved problem in the learning sciences
  - Defines a design space for instruction
  - Experiments & theory to find areas of space where robust learning is maximized
### Assistance Dilemma: Whether to give or to receive?

<table>
<thead>
<tr>
<th>Giving information or assistance</th>
<th>Benefits</th>
<th>Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency of communication</td>
<td>Shallow processing</td>
<td></td>
</tr>
<tr>
<td>Does not engage LTM</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Withholding information or assistance</th>
<th>Benefits</th>
<th>Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation effect Engages &amp; structures LTM</td>
<td>Cost of errors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Floundering, confusion, wasted time</td>
<td></td>
</tr>
</tbody>
</table>
The no pain, no gain argument => more assistance

Many researchers & results argue for lower assistance
place greater demands on students

- “Desirable difficulties” (Bjork)
  • Interleaved practice, delayed feedback, wide spacing
- Abstract examples (Kaminski & Sloutsky)
- Prompting for self-explanation
- Invention as preparation (Schwartz)
Testing effect

• “Tests enhance later retention more than additional study of the material, even when tests are given without feedback. This surprising phenomenon is called the testing effect”
  – Roediger & Karpicke, 2006
“Tests enhance later retention more than additional study of the material” (R&K, 06)
Testing Effect Example 2
Thompson, Wenger, & Bartling (1978)

- "Tests enhance later retention more than additional study of the material" (R&K, 06)

Memory task: Study & recall a list of 40 words
F trials were tests where failure yielded feedback: a study trial

Yes? and no
Spacing Effect: Wider is better (lower assistance) ...

But too wide yields poorer long-term retention
& (not shown) re-study requires more instructional time

Is it “no pain, no gain”?

Or “less pain, more gain”?

• Arguments for higher assistance ...
The less pain, more gain argument => more assistance

Other researchers & results argue for higher assistance

- Direct instruction (Behaviorist, Klahr, ...)
- Cognitive Load Theory (Sweller, Mayer ...)
  - Reduce “extraneous” cognitive load
  - Mayer’s multimedia principles
    - Modality, contiguity principles ....
  - Sweller et al.
    - No-goal problems, worked examples ...
- Provocative paper:
  - Why minimal guidance during instruction does not work:
    An analysis of the failure of constructivist, discovery, problem-based, ... (Kirschner, Sweller, & Clark, 2006).
Worked Examples
Sweller & Cooper

- “a worked example constitutes the epitome of strongly guided instruction” (K, S, C, 06)

Comparison
- 8 Problems
  Solve \((a+b)/c = d\) for \(a\)
  Vs.
- Example-problem pairs
  Example:
  \((a+b)/c = d\)
  \(a+b = dc\)
  \(a = dc - b\)

Yes & maybe?
Can worked example principle beat a strong control?

- Prior work: Added worked examples to *untutored* problem solving
- New: Add examples to *tutored* problem solving
PSLC *In Vivo* Studies of Worked Examples in Intelligent Tutors

- Chemistry tutor studies
  - Replacing half problems with worked examples yields more efficient learning — same outcome in ~20% less time
- Algebra
  - Less time, better long-term retention
- Geometry tutor studies
  - Not only less time, but better conceptual transfer
Ecological Control = Standard Cognitive Tutor
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?
Treatment condition: Half of steps are given as examples.

Student still has to self explain worked out step.

Worked out steps with calculation shown by Tutor.

Given is circle A with arc BD.

If the measure of arc BD is 34.7°, what is the measure of arc BFD?

\[ m \text{Arc BFD} - m \text{Arc BD} = 360 \text{ degrees} \]
\[ m \text{Arc BFD} = 360 \text{ degrees} - m \text{Arc BD} \]
\[ m \text{Arc BFD} = 360 - 34.7 \]
\[ m \text{Arc BFD} = 325.3 \]

**Rule:**
Result: Better conceptual transfer

\[ d = 0.73 \ast \]
Think aloud data: Indicate different modes of thinking

- Problem group: Explained how
  - “one can compute the measure of arc EF by subtracting 33.3 from 360”
  => *more procedural learning*

- Example group: Explained why
  - “This is a major and minor arc, this means the sum of both is 360 degrees”
  => *more conceptual principle learning*
Theory: Does Cognitive Load Theory Account for Result?

• Tutors are designed to minimize load
  – Tutor provides step-by-step feedback, often gives goal structure to students
    • Mediates goal processing load attributed to problem solving

• Nevertheless students seem to benefit
  – Better to present examples before rather than within problems

• Why?
Why examples before rather than within problems?

• Cognitive Load Theory not adequate
• Alternative: Metacognitive frame of mind evoked by problem solving
  – More engineering than science mode of thinking (Schauble; Miller)
  – More performance oriented than learning oriented (Dweck; Elliot)
Tutor Data Mining Evidence

• What happens when students get a “bottom-out” hint?
• ...

“Bottom-out” hint in Cognitive Tutors => on-demand example
Individual differences in self-explanation

After bottom-out example:

- Some students enter answer & move on
- Others seem to reflect, engage in self-explanation
  - Measure: spend more time than usual

- *Students who reflect more, learn more!*  
  \((R = 0.48)\)
A step toward resolving assistance dilemma ... with very broad impact implications!

Instructor options

<table>
<thead>
<tr>
<th>Give</th>
<th>Elicit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lecture</strong></td>
<td><strong>Self-explanation</strong></td>
</tr>
<tr>
<td><strong>Worked example</strong></td>
<td><strong>Homework</strong></td>
</tr>
</tbody>
</table>

Explicit rules

Implicit: Example solutions

- **Current instruction**: Gives rules & elicits solutions
- **Better instruction**: Gives solutions & elicits rules/concepts
Final Thoughts

- Cognitive Tutors bring learning science to schools, but also ...
- Ed Tech as “Hubble telescope” for learning research!
- Assistance Dilemma is fundamental unsolved problem in learning science
  - Inverted-U function worked out for some dimensions of assistance
  - Using in adaptive systems that optimize for robust learning
END 1
### Assistance Dilemma Summary

<table>
<thead>
<tr>
<th>Instructional Support</th>
<th>Poor learning outcome</th>
<th>Good learning outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>High assistance</td>
<td>crutch</td>
<td>scaffold</td>
</tr>
<tr>
<td>(less demanding)</td>
<td>undesirable difficulty; extraneous load</td>
<td>desirable difficulty; germane load</td>
</tr>
<tr>
<td>Low assistance</td>
<td>undesirable difficulty; extraneous load</td>
<td>desirable difficulty; germane load</td>
</tr>
<tr>
<td>(more demanding)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Need predictive theory => engineer more effective instruction!**
Formulating trade-off => adaptive technology to optimize learning

\[
eff_m = \frac{p_m b_{\text{suc}} g_m + (1-p_m) b_{\text{fail}} g_m}{p_m (t_m + \text{fixedsuccosts}) + (1-p_m) \text{fixedfailcosts}}
\]

- \( \text{eff}_m \) = efficiency of robust learning
- \( p_m b_{\text{suc}} g_m \) = learning from success
- \( (1-p_m) b_{\text{fail}} g_m \) = learning from failure
- \( p_m (t_m + fsc) \) = success time
- \( (1-p_m) ffc \) = failure time

- \( m \) = activation of fact
- \( p_m \) = probability of recall success
- \( b_{\text{suc}} \) = gain from success
- \( b_{\text{fail}} \) = gain from review after failure
- \( g_m \) = long-term increase in activation
- \( t_m \) = time of recall
- \( fsc \) = time for success
- \( ffc \) = time for failure
General form of “assistance formula”

For each *learning event*:

Robust learning efficiency gain = 
\[ p \times \text{benefit-of-success} + (1-p)\times\text{benefit-of-failure} \]
\[ p \times \text{cost-of-success} + (1-p)\times\text{cost-of-failure} \]

\[ p = \text{Probability of success during instruction} \]
Pavlik’s Tutor for Fact Practice

Different practice tasks:
- English->Pinyin
- Audio->English
- Hanzi->English

Practice trial

+ feedback when correct

Review when incorrect
Adaptive Fading of Examples

- Fading based on quality of self explanations of worked out value steps (assessed by Tutor)
- Students who self explain well receive fewer examples than students who self explain poorly
Results

- adaptive fading examples > fixed conditions
**Results: In Vivo study**

- **Delayed Post-Test**

  ![Bar chart showing performance in % for different conditions]

  - **problem solving**
  - **fixed fading**
  - **adaptive fading**

  ➢ Result is robust in classroom environment: adaptive fading examples > problem solving
Examples <> Study trials  
Problems <> Test trials

What’s the difference?

• Testing effect studies focus on facts
  – Learning process is memory

• Worked example studies are about learning general rules & procedures
  – Robust learning requires category induction as well as memory
Progress on Principles of Learning to Guide Practice

- PSLC wiki: See learnlab.org

- IES: Cognition Practice Guide

- APS group & web site