Causal Model Search in Educational Research

Martina Rau¹
Richard Scheines²

¹ Department of Educational Psychology, UW Madison
² Department of Philosophy, CMU
1. INTRODUCTION
“Online” Educational Data

Data Pouring in From:

- Computer Tutors
- Online courses
- Virtual Labs
- NSF center on learning science (1 of 6)
- Cognitive Tutors (Algebra, Physics, Geometry, etc.)
  - ~600,000 HS students
  - Recent independent evaluation (180 schools): twice as much algebra learned
- Datashop
  - ~500 publicly accessible datasets in standardized format
  - Analytic tools for analyzing these data

A 2000 kg car in neutral at the top of a 20.0 deg inclined driveway 20.0 m long slips its parking brake and rolls down. Assume that the driveway is frictionless.

What is the magnitude of the velocity of the car when it hits the garage door?

Answer:
Online Course

CMU: The Open Learning Initiative
www.cmu.edu/oli

- Since 2002
- 25 College courses
- Automatic data logging
- Dozens of research studies

EdX
https://www.edx.org/

- MIT, Harvard, Berkeley, UT
- > $ 50 million in start-up funding
- Data collection being made public
- Data mining being prioritized
Virtual Labs: Chem Lab
Virtual Labs: Chem Lab

- Number of engaged actions ⇒ 48% of the post-test variation
- # interactions with the virtual lab outweighed ALL other factors including gender and SAT score as the predictor of positive learning outcome.
Kinds of Data

1. Log data – time stamped events:
   - login, page request, glossary, quiz attempt, score request, video, etc

2. Assessment data -
   - Pre-test scores
   - Intermediate assessments (low stakes, high stakes)
   - Midterm score
   - Final exam scores

3. Problem Solving Data:
   a) Unstructured Virtual Labs --> Customized Data
   b) Structured Cognitive Tutors --> PSLC Data Shop
## Log Data: Edx MOOC Example

<table>
<thead>
<tr>
<th>User</th>
<th>Res</th>
<th>Time</th>
<th>Resp1</th>
<th>Resp2</th>
<th>Count1</th>
<th>Count2</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>video</td>
<td>2m 30s</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>9</td>
<td>answer</td>
<td>10m 5s</td>
<td>correct</td>
<td>correct</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>book</td>
<td>4m 41s</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>10</td>
<td>book</td>
<td>40s</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>10</td>
<td>answer</td>
<td>20s</td>
<td>incorr.</td>
<td>--</td>
<td>1</td>
<td>--</td>
</tr>
<tr>
<td>10</td>
<td>answer</td>
<td>15s</td>
<td>incorr.</td>
<td>--</td>
<td>2</td>
<td>--</td>
</tr>
<tr>
<td>10</td>
<td>answer</td>
<td>1m 8s</td>
<td>incorr.</td>
<td>incorr.</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>answer</td>
<td>28s</td>
<td>--</td>
<td>correct</td>
<td>--</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>video</td>
<td>2m 10s</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>10</td>
<td>answer</td>
<td>6s</td>
<td>correct</td>
<td>--</td>
<td>4</td>
<td>--</td>
</tr>
</tbody>
</table>
## Log Data: Fractions Tutor Example

<table>
<thead>
<tr>
<th>Student Id</th>
<th>Time</th>
<th>Duration (sec)</th>
<th>Student Response Type</th>
<th>Tutor Response Type</th>
<th>Problem Name</th>
<th>Step Name</th>
<th>KC Model</th>
<th>Attempt At Step</th>
<th>Outcome</th>
<th>Selection</th>
<th>Action</th>
<th>Input</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student1</td>
<td>5/14/13</td>
<td>14:09</td>
<td>1 ATTEMPT</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NtpDate</td>
<td>NtpTimeCheck</td>
</tr>
<tr>
<td></td>
<td>5/14/13</td>
<td>14:09</td>
<td>32 ATTEMPT RESULT</td>
<td></td>
<td>fract1_numMultiply1</td>
<td>equivMultipl yNum</td>
<td>1 Correct</td>
<td>fract1_numMultiply1</td>
<td>UpdateTextAr ea</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5/14/13</td>
<td>14:10</td>
<td>4 ATTEMPT RESULT</td>
<td></td>
<td>fract1_denomMultiply1</td>
<td>equivMultipl yDenom</td>
<td>1 Correct</td>
<td>fract1_denomMultiply1</td>
<td>UpdateTextAr ea</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5/14/13</td>
<td>14:10</td>
<td>4 ATTEMPT RESULT</td>
<td></td>
<td>_root goToStep</td>
<td></td>
<td>1 Correct</td>
<td>_root goToStep</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5/14/13</td>
<td>14:10</td>
<td>18 ATTEMPT RESULT</td>
<td></td>
<td>fract3_numUpdateTextAr ea</td>
<td>equivNameN umFract</td>
<td>1 Correct</td>
<td>fract3_numUpdateTextAr ea</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5/14/13</td>
<td>14:10</td>
<td>3 ATTEMPT RESULT</td>
<td></td>
<td>fract3_denomUpdateTextAr ea</td>
<td>equivNameDen omFract</td>
<td>1 Correct</td>
<td>fract3_denomUpdateTextAr ea</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5/14/13</td>
<td>14:10</td>
<td>6 ATTEMPT RESULT</td>
<td></td>
<td>fract4_numUpdateTextAr ea</td>
<td>equivNameN umFract</td>
<td>1 Correct</td>
<td>fract4_numUpdateTextAr ea</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
“Online” Educational Data

Questions/Challenges:

- Raw Log Data $\rightarrow$ Meaningful Variables
- Which curricular or tutorial interventions cause learning?
- Which (influencible) student behaviors facilitate learning?
- By what mechanisms do successful interventions cause learning?
Motivation

1. Experiment i

2. Causal Model Search

3. Experiment i++

Closing the loop

- Open questions
2. MODEL SEARCH: ONLINE COURSE BEHAVIORS
Causal and Statistical Reasoning

Suppose you are traveling in a car and you want to make several calls on your cell phone. In the simulation below, click on the "SEND" button to place a call, and click on the "END" button to end a call (you must click on END before you can try another call). The phone on the right will ring if your call got through. Attempt at least 10 calls.

Question 2

Which of the choices best represents the causal system in the simulation?

A. (Phone Button [send,end])
B. (Phone Button [send,end], Attempts[0,10])
C. (Phone Button [send,end], Call Connected [yes, no])
Student & Log Data

- Pre-test (%)
- Midterm1 (%)
- Gender
- Race
- Computer-comfort
- Final Exam (%)
- Logged in time
- Voluntary-exercise completion (%)
- Quiz Scores (avg. %)
- Print-requests (% of modules)
- 12 others
Interaction $\rightarrow$ Learning

Year 1

Print requests $\rightarrow$ voluntary interactive exercises $\rightarrow$ quiz $\rightarrow$ final

Pretest $\rightarrow$ Print requests $\rightarrow$ voluntary interactive exercises $\rightarrow$ quiz $\rightarrow$ final

Year 2

Print requests $\rightarrow$ voluntary interactive exercises

Pretest $\rightarrow$ Print requests $\rightarrow$ voluntary interactive exercises $\rightarrow$ final

Correlation coefficients:
- $r_{pretest, print requests} = .302$
- $r_{print requests, voluntary exercises} = .75$
- $r_{voluntary exercises, quiz} = .353$
- $r_{pretest, quiz} = .323$
- $r_{quiz, final} = .353$
- $r_{pretest, final} = .323$
- $r_{print requests, final} = .41$
- $r_{voluntary exercises, final} = .41$
- $r_{pretest, print requests} = .323$
- $r_{print requests, voluntary exercises} = .75$
- $r_{voluntary exercises, quiz} = .353$
- $r_{pretest, quiz} = .323$
- $r_{quiz, final} = .353$
- $r_{pretest, final} = .323$
- $r_{print requests, final} = .41$
- $r_{voluntary exercises, final} = .41$

CS - Pedagogy
3. SEARCHING FOR MECHANISMS/MEDIATORS
What are the Mediators?

Experimental Condition

Student Behavior/Understanding

Pre-test

Student Properties

Post-Test

Unmeasured Student Properties
What are the Mediators?

Exp. Condition \(\perp\parallel\) Post-Test | \{Pre-test, Student Properties, Engagement, Correct Rep\}

Exp. Condition \(\perp\parallel\) Post-Test | \{Pre-test, Student Properties, Time, Correct Rep\}
Fractions Tutor

Naming Fractions

A. Let's name a fraction to compare it to another!
   1. How many blue sections are there? 3
   2. Use the arrows to make all sections the same size.
   3. How many blue sections do you need to fill the gray diagram? 6
   4. What fraction are the blue sections of the unit? 3/6

B. Let's name a second fraction to compare it to the first!
   1. How many purple sections are there? 3
   2. Use the arrows to make all sections the same size.
   3. How many purple sections do you need to fill the gray diagram? 4
   4. What fraction are the purple sections of the unit? 3/4

C. Are these fractions still the same?
   1. How many total sections are in the blue diagram? 6
   2. The blue sections are equal to the purple section.
   3. However, the diagrams show different fractions, because their units have different shapes.
   4. How many total sections are in the purple diagram? 4

Hint: Continue
## Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Single graphical representation</th>
<th>Multiple graphical representations</th>
</tr>
</thead>
<tbody>
<tr>
<td>No self-explanation</td>
<td><img src="image" alt="Graphical representation" /></td>
<td><img src="image" alt="Graphical representation" /></td>
</tr>
<tr>
<td>prompts</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Graphical representation" /></td>
<td><img src="image" alt="Graphical representation" /></td>
</tr>
<tr>
<td>Self-explanation</td>
<td><img src="image" alt="Graphical representation" /></td>
<td><img src="image" alt="Graphical representation" /></td>
</tr>
<tr>
<td>prompts</td>
<td><img src="image" alt="Question box" /></td>
<td><img src="image" alt="Question box" /></td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Option box" /></td>
<td><img src="image" alt="Option box" /></td>
</tr>
</tbody>
</table>

\[N = 110\] 6th-grade students, 2.5h

[Rau et al., AIED 2009, best student paper]
## Experiment 2

<table>
<thead>
<tr>
<th>Self-explanation prompts</th>
<th>Single graphical representation</th>
<th>Multiple graphical representations</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Diagram" /></td>
<td><img src="image2" alt="Diagram" /></td>
<td><img src="image3" alt="Diagram" /></td>
</tr>
<tr>
<td><img src="image4" alt="Diagram" /></td>
<td><img src="image5" alt="Diagram" /></td>
<td><img src="image6" alt="Diagram" /></td>
</tr>
</tbody>
</table>

$N = 290$ 4th- and 5th-grade students, 5h  
[Rau et al., ICLS 2012]
Learning with Multiple Representations

- Multiple representations $\Rightarrow$ Learning $\sqrt{}$
- Mechanisms?
- Standard in ITS (Intelligent Tutor Systems):
  - Error-rate
  - Hint-use
  - Time-spent
Model Search: Experiment 1

\[ \chi^2 = 22.11, \ df = 19, \ p = .29 \]

No mediation of multiple representations on learning.
Model Search: Experiment 2

Error-rate mediates negative effect

Positive direct effects

χ² = 6.89, df = 10, p = .74
Mediator Variables: Non-monotonic?
Transforming/Defining Variables

Raw data (non-monotonic) → Transformation → Transformed Var. (monotonic)

- Raw measures of error-rate, hint-use, time-spent,
- Identification of ‘optimal level’ of error-rate, hint-use, time-spent
- Transformation: distance (squared distance) from optimum
Transforming Variables: No help

- Result: raw variables no worse, perhaps better
- Models using the raw variables explained slightly more variance than models with the transformed variables

[Rau & Scheines, EDM 2012]
Multiple representations increase learning

Standard Variables: *Time*, *Error*, and *Hints* do *not* seem to be mechanisms through which multiple representations *increase* learning.
4. INFORMED MEDIATORS
Motivation

- **Learning processes** [Koedinger et al., 2012]:
  - **Understanding**: sense-making processes
    \[ a = 1 \times a \]
    
    \[ \ldots \]

  - **Fluency**: fluency-building processes
    \[ a + a + a = 3 \times a \]
    \[ \ldots \]
Comparing Fractions

A. Let's review a rectangle as an example to compare two fractions!

1. The two rectangles below show $\frac{3}{5}$ and $\frac{3}{7}$.

   - **Rectangle A:**
   - **Rectangle B:**

2. The sections in rectangle A are **larger than** the sections in rectangle B, because in rectangle A, there are **fewer** sections than in rectangle B.

3. Since there are **3** colored sections, in each rectangle, the fraction in rectangle A is **larger than** the fraction in rectangle B.

B. Let's use a number line to compare two fractions!

1. The two number lines below show $\frac{3}{5}$.

   - **Number line A:**
   - **Number line B:**

2. The sections in number line A are **larger than** the sections in number line B, because in number line A, there are **fewer** sections than in number line B.

3. Since there are **3** sections between 0 and the dot in each number line, the fraction in number line A is **larger than** the fraction in number line B.

C. What did we learn about the rectangle and the number line?

1. Rectangle A and number line A each have **5** total sections. Rectangle B and number line B each have **7** total sections.

2. The more sections the rectangle and the number line are cut into, the **smaller** the size of the sections.

3. All rectangles and number lines have the same numerator, so the rectangle with the larger sections shows the **larger** fraction.
Which representations are equivalent?
## Background: Experiment 3

<table>
<thead>
<tr>
<th>Sense-making problems</th>
<th>no</th>
<th>yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluency-building problems</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- \( N = 599 \text{ 4}^{\text{th}}-/5^{\text{th}}\)-graders

[Rau et al., ITS 2012]
Mediator hypotheses

- How do sense-making processes and fluency-building processes interact?
  - Understanding hypothesis:

  ![Diagram](image)

  Sense-making support enables benefit?

  Fluency-building support

Mediation hypothesis

- Pre
- Post
- Delayed

**Mediators**

- Performance on fluency-problems
- Performance on sense-problems

**Hypothesis:**

- Sense + fluency (vs. fluency-only) (vs. sense-only)
Mediation Hypotheses

- How do sense-making processes and fluency-building processes interact?
  - Understanding hypothesis:
    - Sense-making support enables benefit?
    - Fluency-building support

- Fluency hypothesis:
  - Sense-making support enables benefit?
  - Fluency-building support
Mediation Hypotheses

- Mediators: Performance on fluency-problems
- Mediators: Performance on sense-problems

- pre
- post
- delayed

- sense + fluency (vs. fluency-only)
- (vs. sense-only)
Variable identification

- Search among large number of potential variables [Rau et al., EDM 2012]
- Based on knowledge component model
Knowledge Component Model

Equivalent Fractions

A Let's review rectangles to see what makes fractions equivalent!

1 The blue and the purple rectangle show different fractions. What fraction does each rectangle show?

2 Are these two fractions equivalent? yes

3 By what numbers must you multiply to get the equivalent fraction?

B Let's use number lines to see what makes fractions equivalent!

1 The two number lines show different fractions. What fraction does each number line show?

2 Are these two fractions equivalent? yes

3 By what numbers must you multiply to get the equivalent fraction?

C What did we learn about the rectangle and the number line?

1 You can find equivalent fractions by multiplying numerator and denominator by the same number.

2 Multiplying the numerator and the denominator by the same number is like cutting the areas into more sections without changing the amount.

3 Rectangles and number lines that show the same amount with different numbers of sections show equivalent fractions.
Variable identification

- Search among large number of potential variables
  [Rau et al., EDM 2012]
- Based on knowledge component model
  - Significant predictors of posttest performance
  - Significant differences between conditions
Understanding hypothesis

Mediators

Performance on fluency-problems

nameCircleMixError

equivalenceError

Performance on sense-problems

improperMixError

pre → post → delayed

sense + fluency (vs. fluency-only) (vs. sense-only)
Understanding hypothesis

- sense + fluency (vs. fluency-only)
- (vs. sense-only)

χ² = 30.88, df = 9, p < .0001; N = 131
Fluency hypothesis

- sense + fluency (vs. fluency-only)
- (vs. sense-only)

Mediators

- Performance on fluency-problems

Mediators

- SE-Error
- Performance on sense-problems
- place1Error
Fluency hypothesis

\[ \chi^2 = 49.14, \text{ df } = 6, \ p < .0001; \ N = 117 \]
Possible alternative models

Possibilities: $> 2^{20}$
Model Search Results: Understanding model

\[ \chi^2 = 16.10, \text{ df } = 6, p = .013; N = 131 \]
Model Search Results: Understanding model

\[ \chi^2 = 16.10, \text{ df} = 6, p = .013; N = 131 \]
Model Search Results: Fluency model

\[ \chi^2 = 8.32, \text{ df} = 5, \ p = .14; \ N = 117 \]
Model Search Results: Fluency model

\[ \chi^2 = 8.32, \text{ df } = 5, \ p = .14; \ N = 117 \]
Mediation hypothesis

Mediators

Performance on fluency-problems

Performance on sense-problems

pre

post

delayed

sense + fluency (vs. fluency-only) (vs. sense-only)
Taking Stock

- Results are in line with understanding hypothesis, but not with fluency hypothesis
  - Sense-making support reduces errors students make on fluency-building problems
    [Rau, Scheines et al., EDM 2013, best paper]

- Limitations
  - Bound to fixed sequence: sense – fluency
  - Different results possible with sequence fluency – sense

- Makes testable predictions:
  - Sense-making support should be provided before fluency-building support
Experiment 4: Results

Which process should instruction support first?

- Understanding hypothesis:
  - Sense-making support enables benefit
  - Fluency-building support

- Fluency hypothesis:
  - Fluency-building support reduces benefit
  - Sense-making support

[Rau et al., AIED 2013]
Experiment 4: Model Search Results

Fluency-building errors

χ² = 4.58, df = 4, p = .33
Experiment 4: Model Search Results

Sense-making errors

$\chi^2 = 3.38$, df = 3, $p = .38$
5. CONCLUSION
Conclusion

- Both sense-making processes and fluency-building processes need to be supported
- Sense-making enhances fluency-building
- Sense-making support should be provided before fluency-building support
- Closing the loop!

1. Experiment 3

2. Causal path analysis

3. Experiment 4
Conclusion

- Overall measures of problem-solving behaviors were not successful at establishing mediation
- Informed mediators explained interaction between different learning processes
- Model search helped identify plausible models for our hypotheses
- Results from mediation analysis made testable predictions
- Results from follow-up experiment were in line with these predictions
Thanks!

REESE-21851-1-1121307, IES R305A120734, Pittsburgh Science of Learning Center, funded by the National Science Foundation (award number SBE-0354420), IES Program for Interdisciplinary Education Research

Vincent Aleven Nikol Rummel, Richard Scheines, Ken Koedinger, and Brian Junker


Students and teachers who have participated in my research
BACKUP SLIDES
Sense-only condition

Pretest

~30 min

Regular tutor problems

Tutor, ~10h

Sense-making problems

Posttest & Survey

~35 min

Delayed posttest

~30 min
Fluency-only condition

<table>
<thead>
<tr>
<th>Pretest</th>
<th>~30 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic 1</td>
<td></td>
</tr>
<tr>
<td>Topic 2</td>
<td></td>
</tr>
<tr>
<td>Topic 3</td>
<td></td>
</tr>
<tr>
<td>:</td>
<td></td>
</tr>
<tr>
<td>Regular tutor problems</td>
<td></td>
</tr>
<tr>
<td>Fluency-building problems</td>
<td></td>
</tr>
<tr>
<td>:</td>
<td></td>
</tr>
<tr>
<td>Posttest &amp; Survey</td>
<td>~35 min</td>
</tr>
<tr>
<td>:</td>
<td></td>
</tr>
<tr>
<td>Delayed posttest</td>
<td>~30 min</td>
</tr>
</tbody>
</table>
Sense + fluency condition

Pretest

~30 min

Topic 1

Regular tutor problems

Sense-making problems

Fluency-building problems

~10h Tutor

~30 min

~35 min

Posttest & Survey

~30 min

Delayed posttest
### Fluency-only vs. sense + fluency

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fluency-only</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sense + fluency</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Fluency-building problems**
- **Sense-making problems**
## Sense-only vs. Sense + fluency

<table>
<thead>
<tr>
<th>Sense-only</th>
<th>Sense + fluency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic 1</td>
<td></td>
</tr>
<tr>
<td>Topic 2</td>
<td></td>
</tr>
<tr>
<td>Topic 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sense-making problems

Fluency-building problems