Causal Model Search in Educational Research

Martina Rau[1]

Richard Scheines[2]

[1] Department of Educational Psychology, UW Madison
[2] Department of Philosophy, CMU

Carnegie Mellon University

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

1. INTRODUCTION

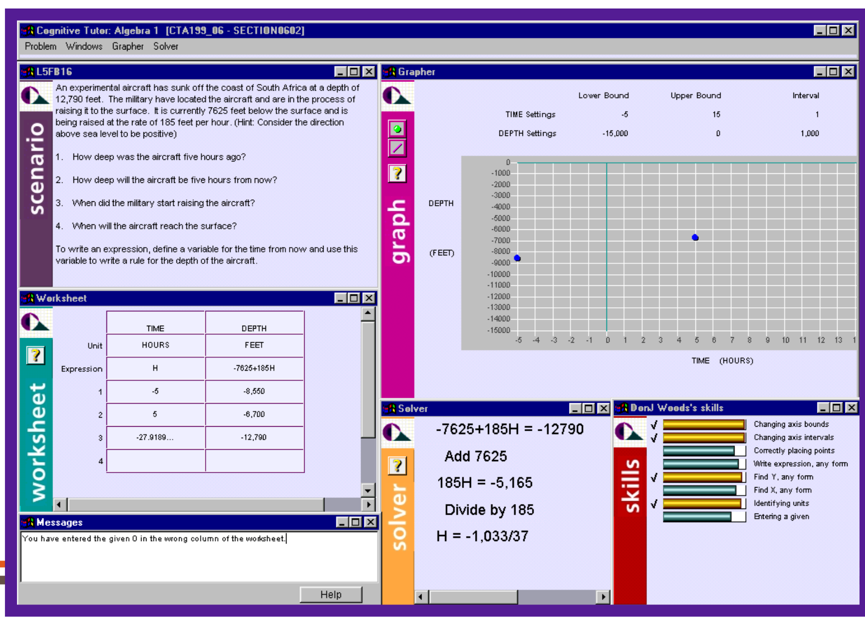
“Online” Educational Data

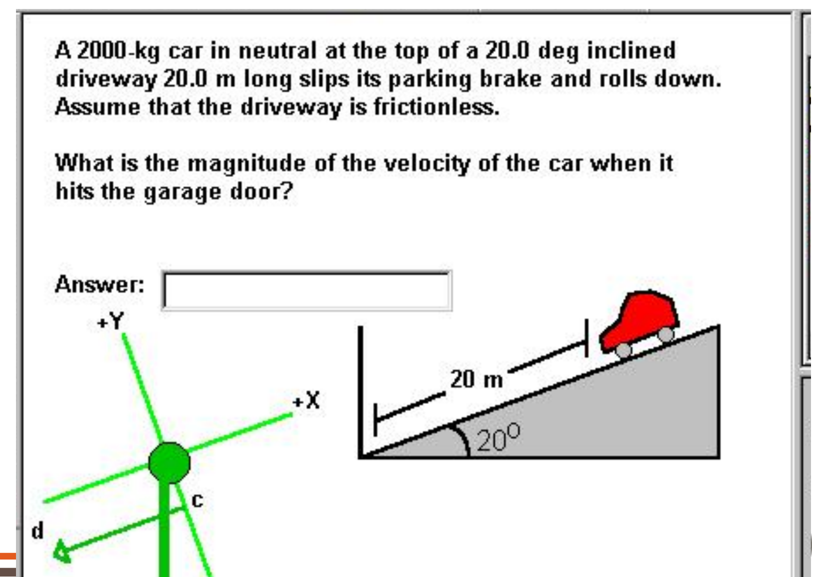
Data Pouring in From:

- Computer Tutors
- Online courses
- Virtual Labs

Pittsburgh Science of Learning Center
http://www.learnlab.org

- 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





Online Course

CMU: The Open Learning Initiative

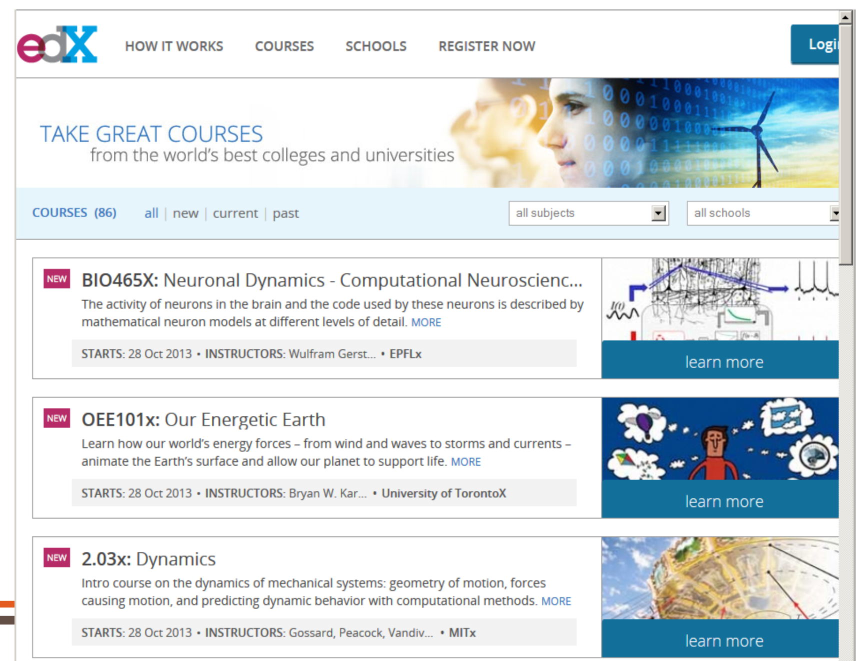
www.cmu.edu/oli

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

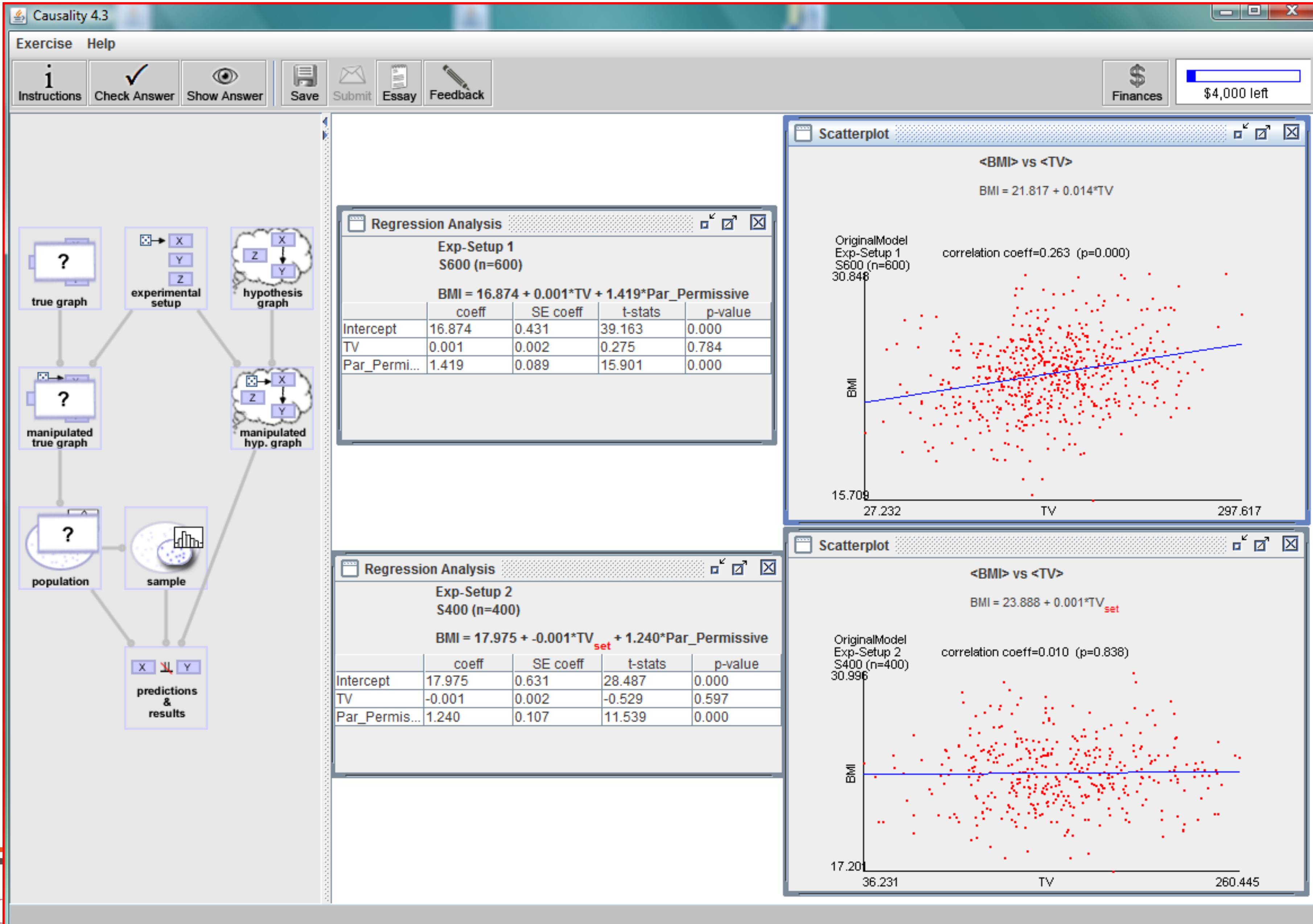
EdX

https://www.edx.org/

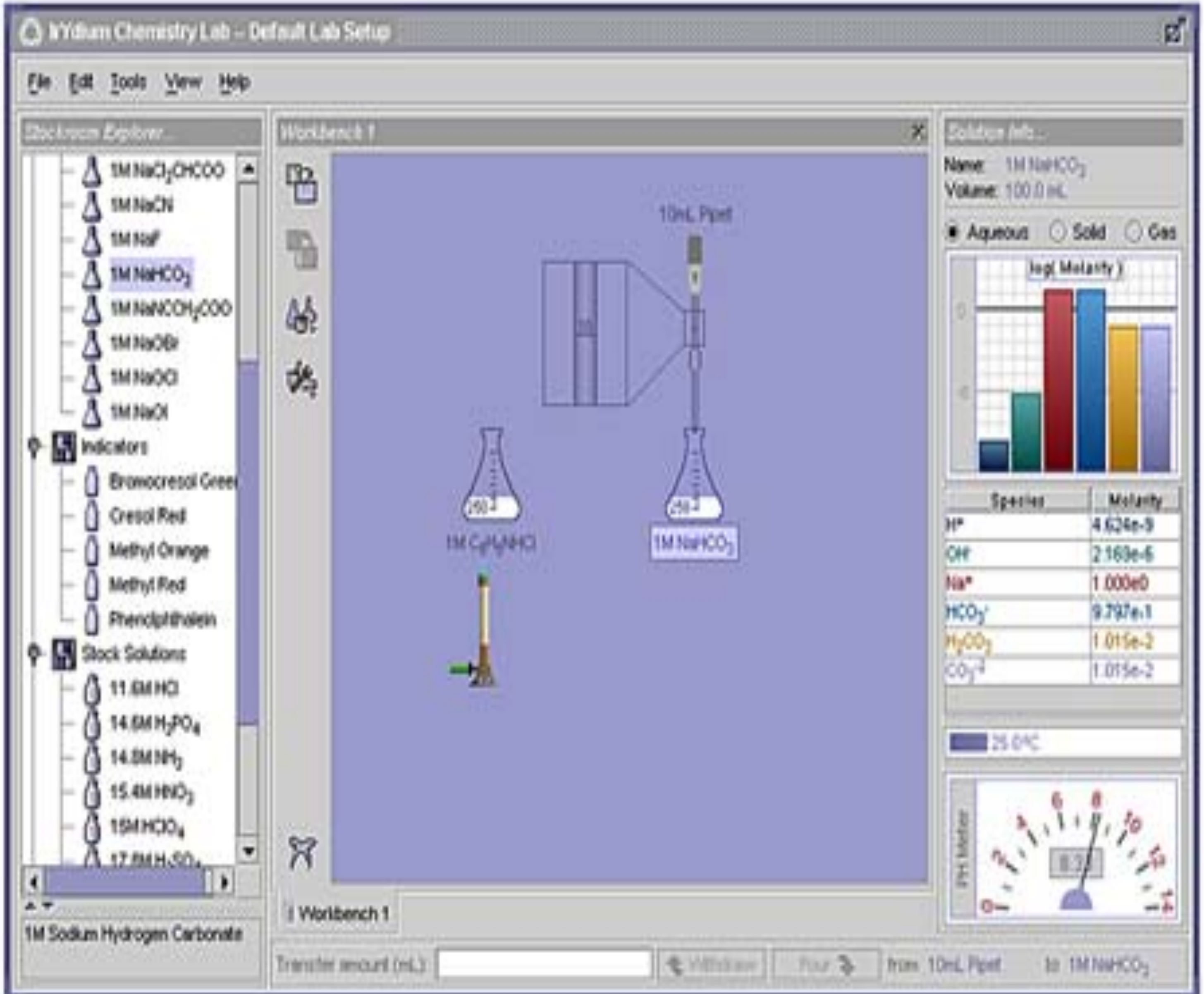
- MIT, Harvard, Berkeley, UT
- > $ 50 million in start-up funding
- Data collection being made public
- Data mining being prioritized



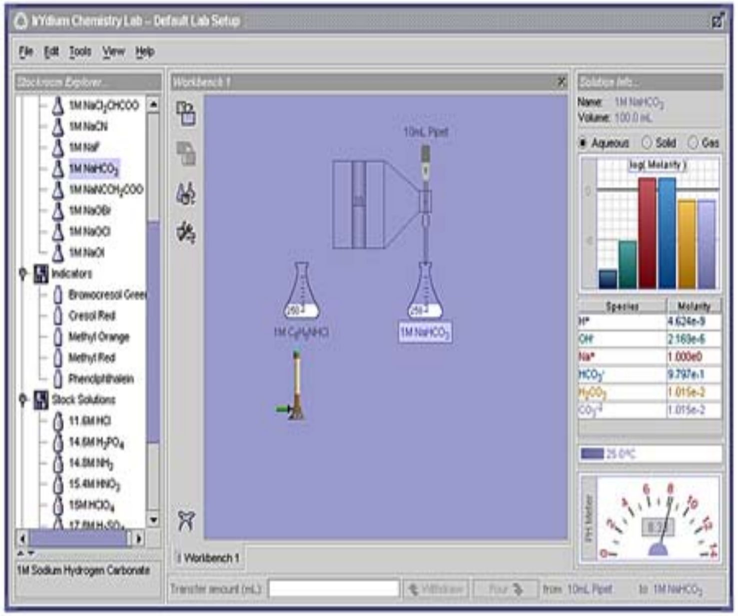
Virtual Labs: Causality Lab



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

User	Res	Time	Resp1	Resp2	Count1	Count2
9	video	2m 30s	- -	- -	- -	- -
9	answer	10m 5s	correct	correct	1	1
10	book	4m 41s	- -	- -	- -	- -
10	book	40s	- -	- -	- -	- -
10	answer	20s	incorr.	- -	1	- -
10	answer	15s	incorr.	- -	2	- -
10	answer	1m 8s	incorr.	incorr.	3	1
10	answer	28s	- -	correct	- -	2
10	video	2m 10s	- -	- -	- -	- -
10	answer	6s	correct	- -	4	- -

Log Data: Fractions Tutor Example

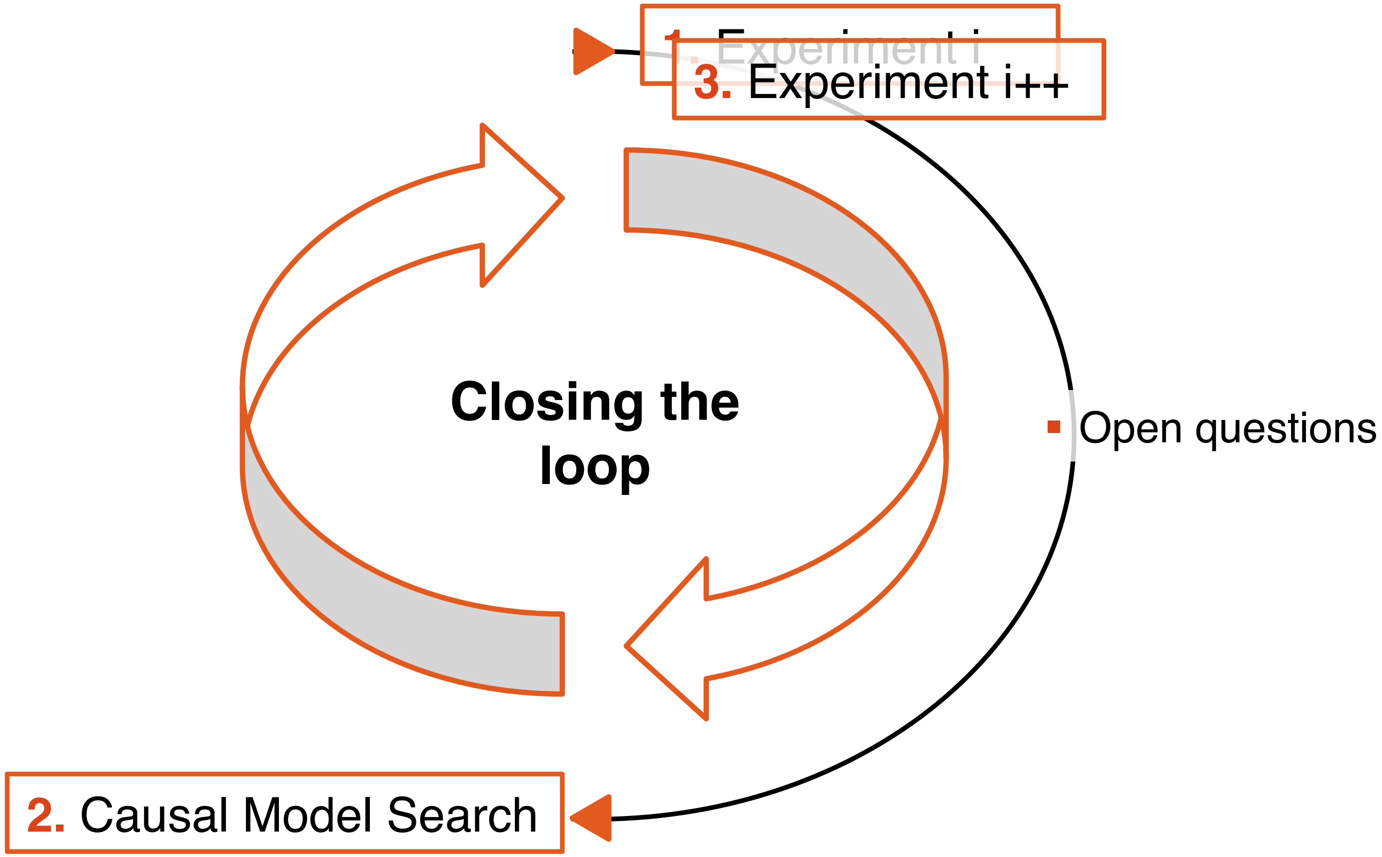
Student Id	Time	Duration (sec)	Student Response Type	Tutor Response Type	Problem Name	Step Name	KC Model	Attempt At Step	Outcome	Selection	Action	Input
Student1	5/14/13 14:09	1	ATTEMPT		1					NtpDate	NtpTimeCheck	2013-05-14 10:09:23.551 -0400
Student1	5/14/13 14:09	32	ATTEMPT	RESULT	1	fract1_numMultiply1 UpdateTextArea	equivMultiplyNum	1	Correct	fract1_numMultiply1	UpdateTextArea	3
Student1	5/14/13 14:10	4	ATTEMPT	RESULT	1	fract1_denomMultiply1 UpdateTextArea	equivMultiplyDenom	1	Correct	fract1_denomMultiply1	UpdateTextArea	3
Student1	5/14/13 14:10	4	ATTEMPT	RESULT	1	_root goToStep		1	Correct	_root	goToStep	2
Student1	5/14/13 14:10	18	ATTEMPT	RESULT	1	fract3_num UpdateTextArea	equivNameNumFract	1	Correct	fract3_num0	UpdateTextArea	1
Student1	5/14/13 14:10	3	ATTEMPT	RESULT	1	fract3_denom UpdateTextArea	equivNameDenomFract	1	Correct	fract3_denom0	UpdateTextArea	3
	5/14/13 14:10	6	ATTEMPT	RESULT	1	fract4_num UpdateTextArea	equivNameNumFract	1	Correct	fract4_num0	UpdateTextArea	3

“Online”Educational Data

Questions/Challenges:

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

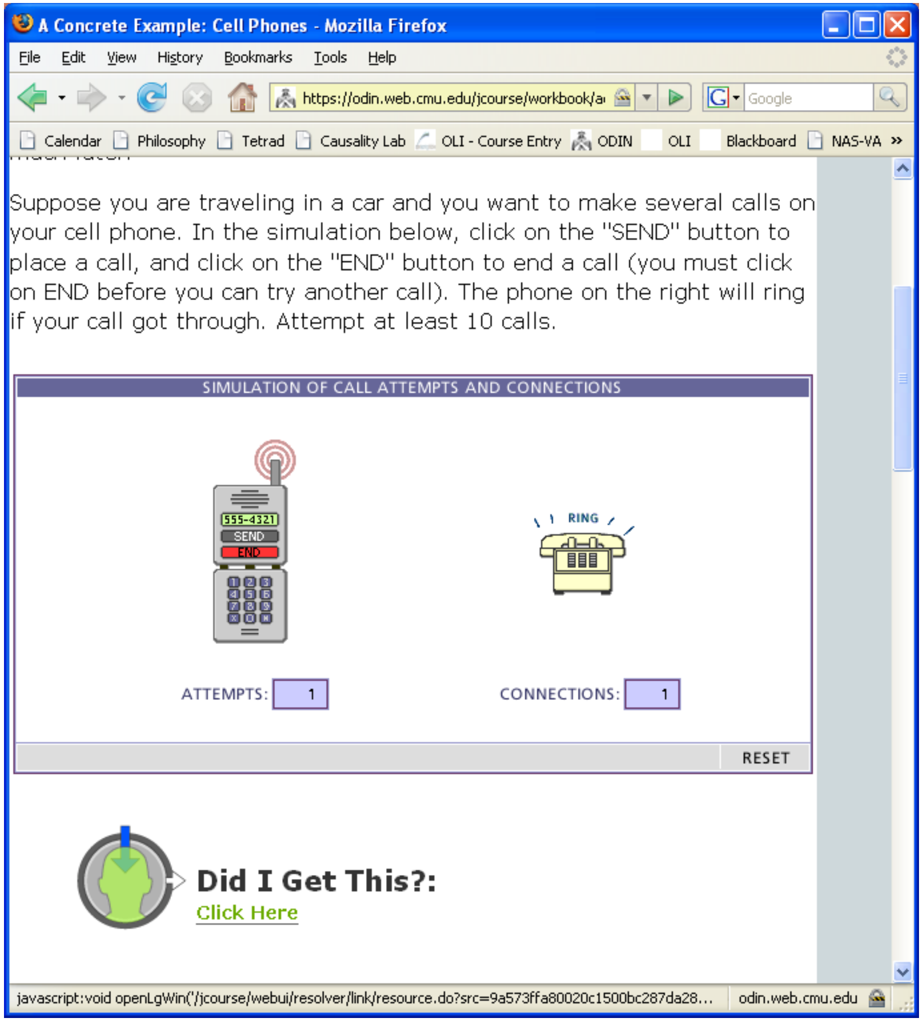
Motivation

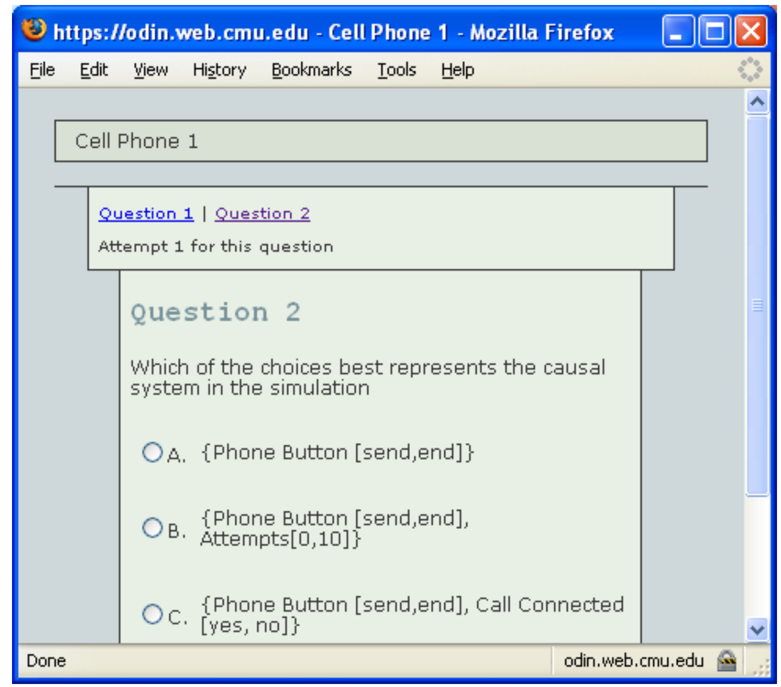


1. Experiment i
3. Experiment i++
Closing the loop
Open questions
2. Causal Model Search

2. MODEL SEARCH: ONLINE COURSE BEHAVIORS

Causal and Statistical Reasoning



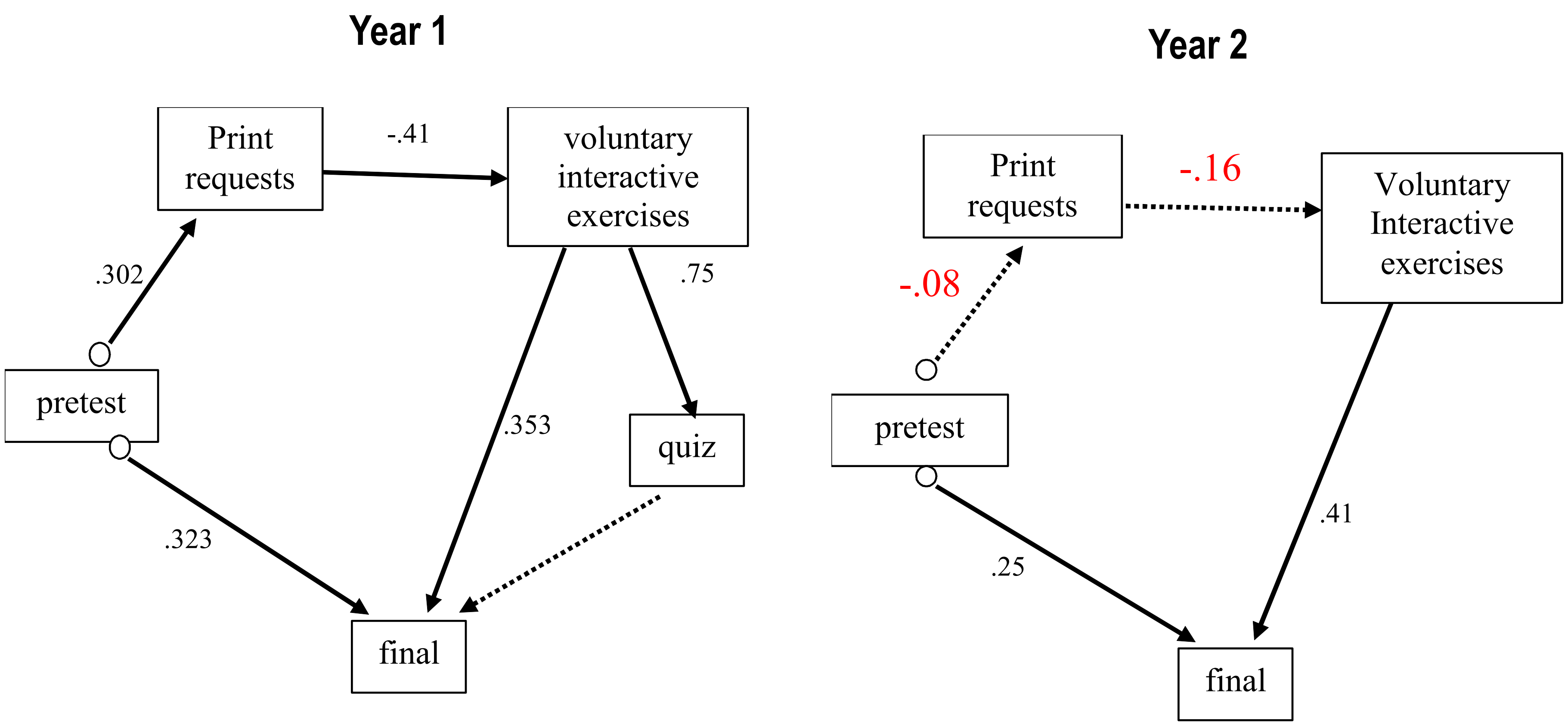


Student & Log Data

- Pre-test (%)
- Midterm1 (%)
- Gender
- Race
- Computer-comfort
- Final Exam (%)

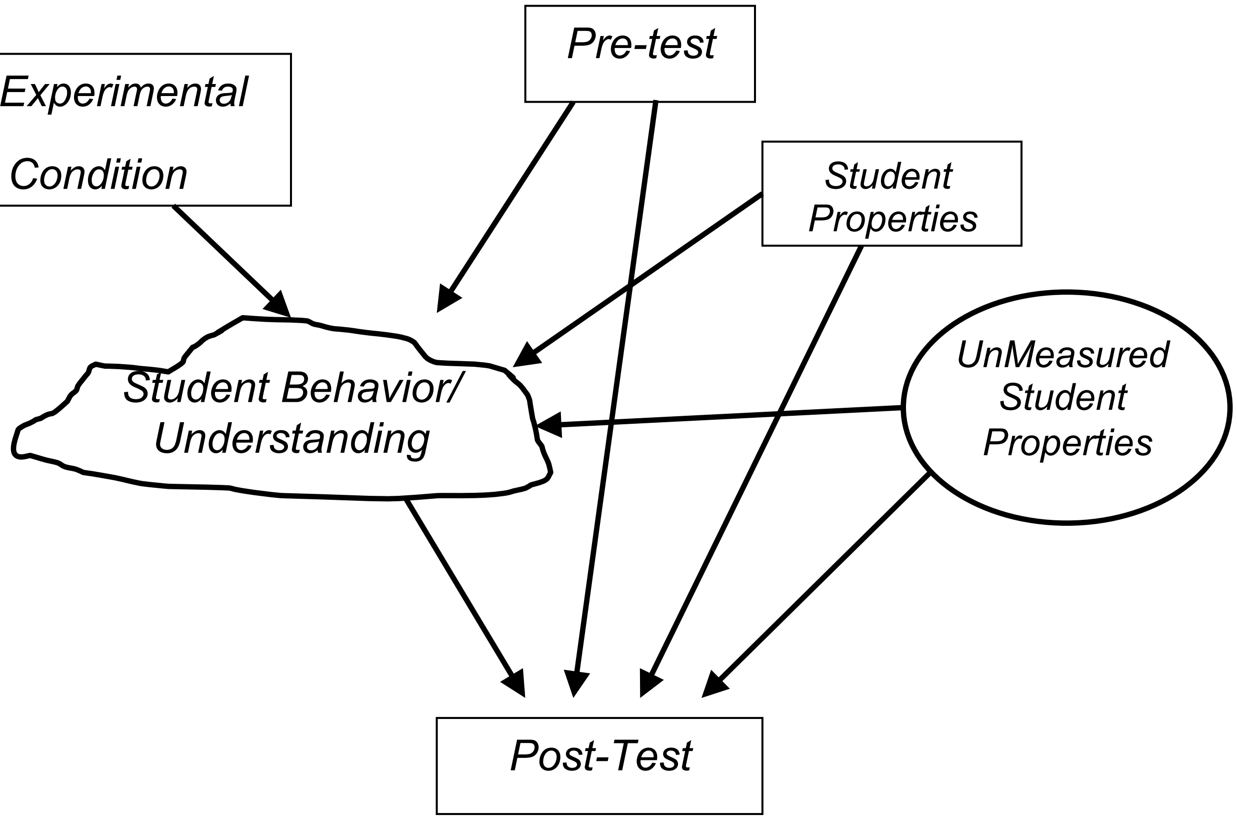
- Logged in time
- Voluntary-exercise completion (%)
- Quiz Scores (avg. %)
- Print-requests (% of modules)
- 12 others

Interaction → Learning

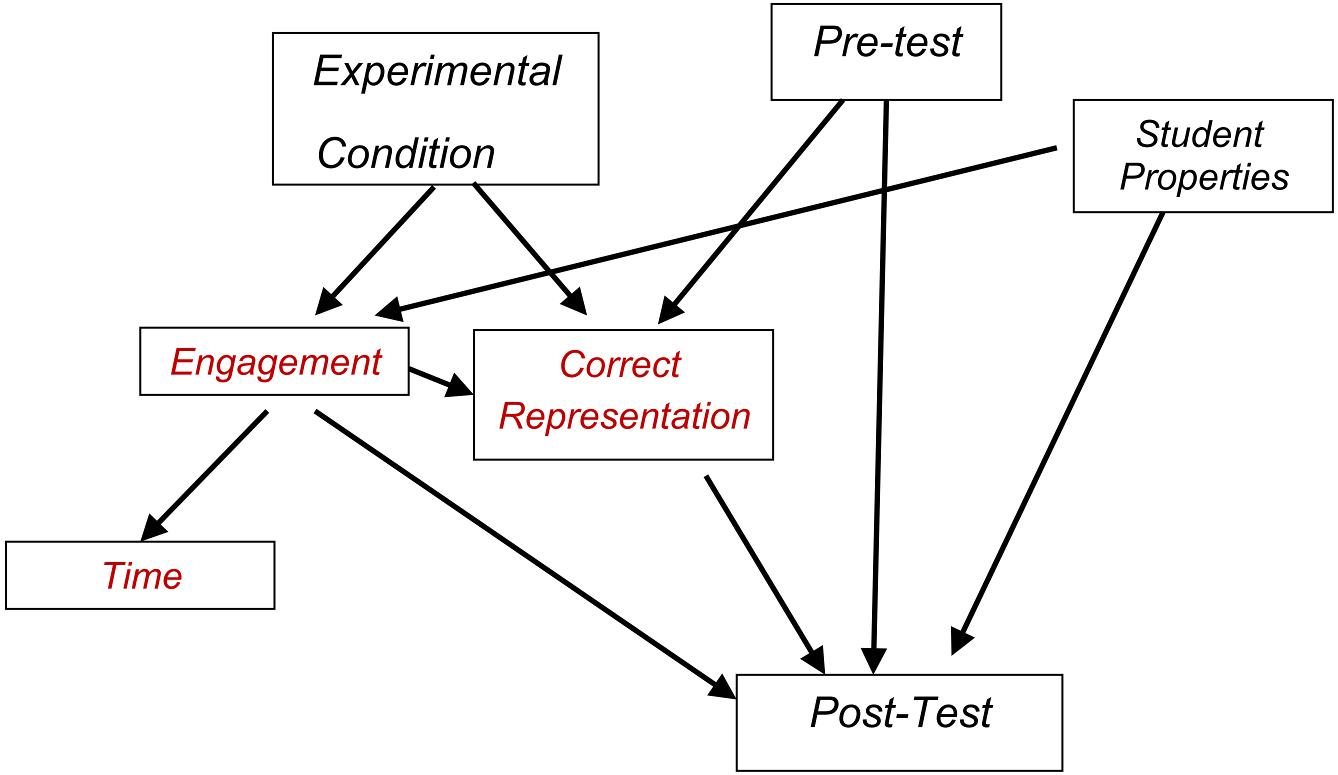


3. SEARCHING FOR MECHANISMS/MEDIATORS

What are the Mediators?



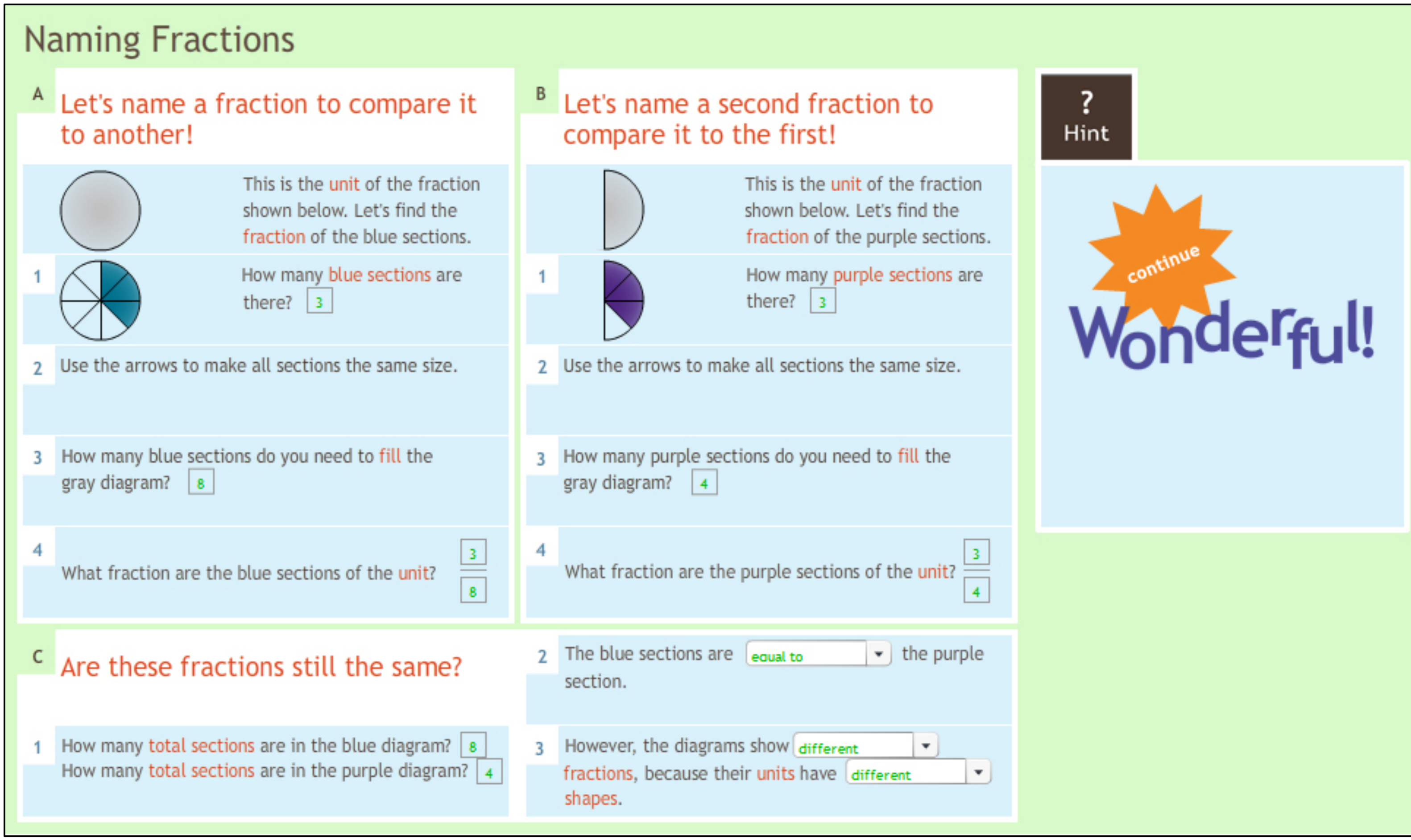
What are the Mediators?



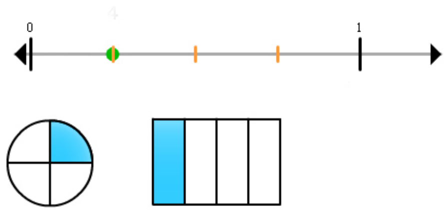
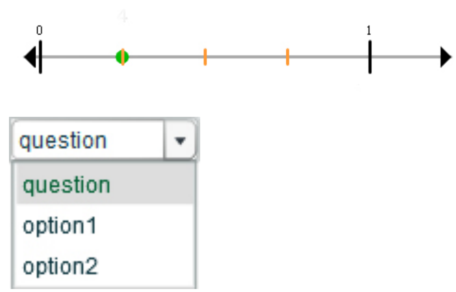
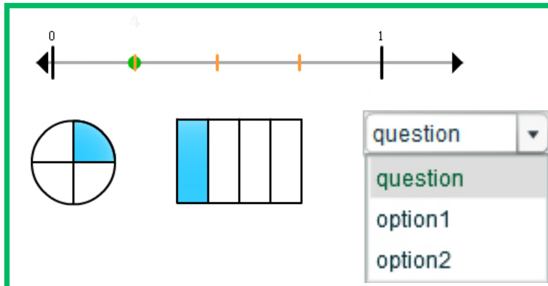
Exp. Condition _||_ Post-Test | {Pre-test, Student Properties, Engagement, Correct Rep}

Exp. Condition ~~_||_~~ Post-Test | {Pre-test, Student Properties, Time, Correct Rep}

Fractions Tutor

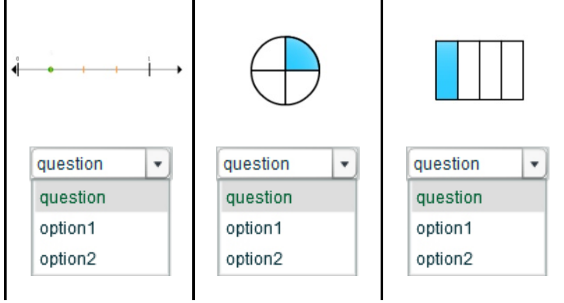
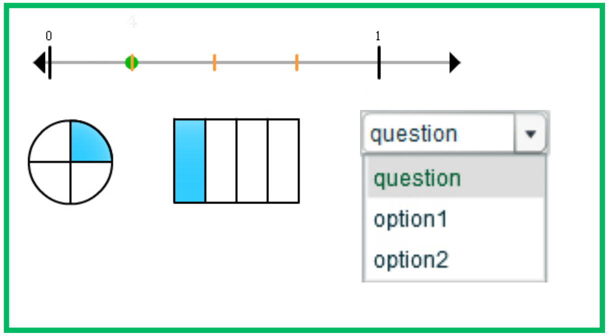


Experiment 1

	Single graphical representation	Multiple graphical representations
No self-explanation prompts		
Self-explanation prompts		

N = 110 6th-grade students, 2.5h
[Rau et al., AIED 2009, best student paper]

Experiment 2

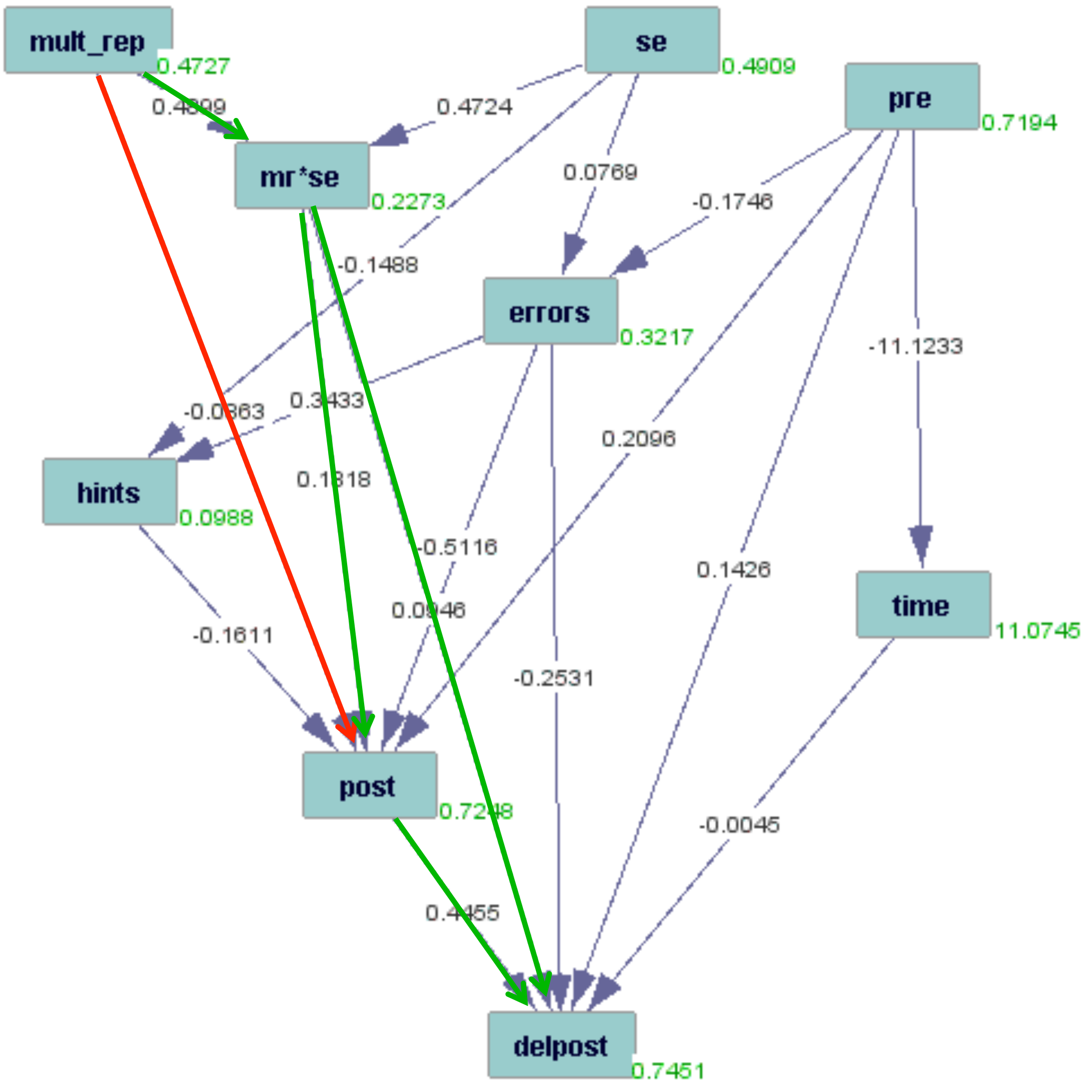
	Single graphical representation	Multiple graphical representations
Self-explanation prompts		

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

Learning with Multiple Representations

- Multiple representations $\Rightarrow$ Learning √
- Mechanisms?
- Standard in ITS (Intelligent Tutor Systems):
  - *Error-rate*
  - *Hint-use*
  - *Time-spent*

Model Search: Experiment 1

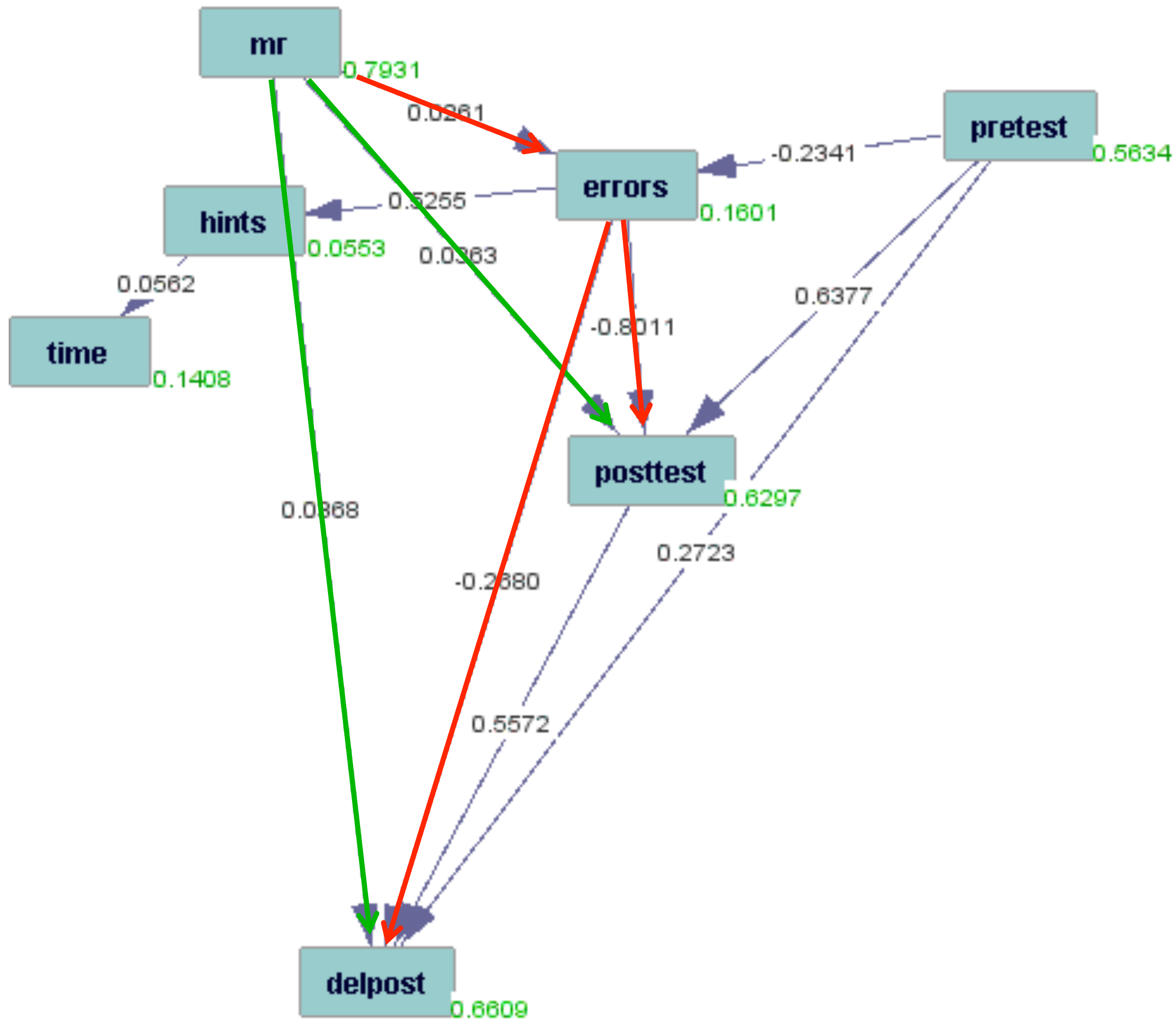


No mediation of multiple representations on learning

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

Model Search: Experiment 2

Error-rate mediates negative effect



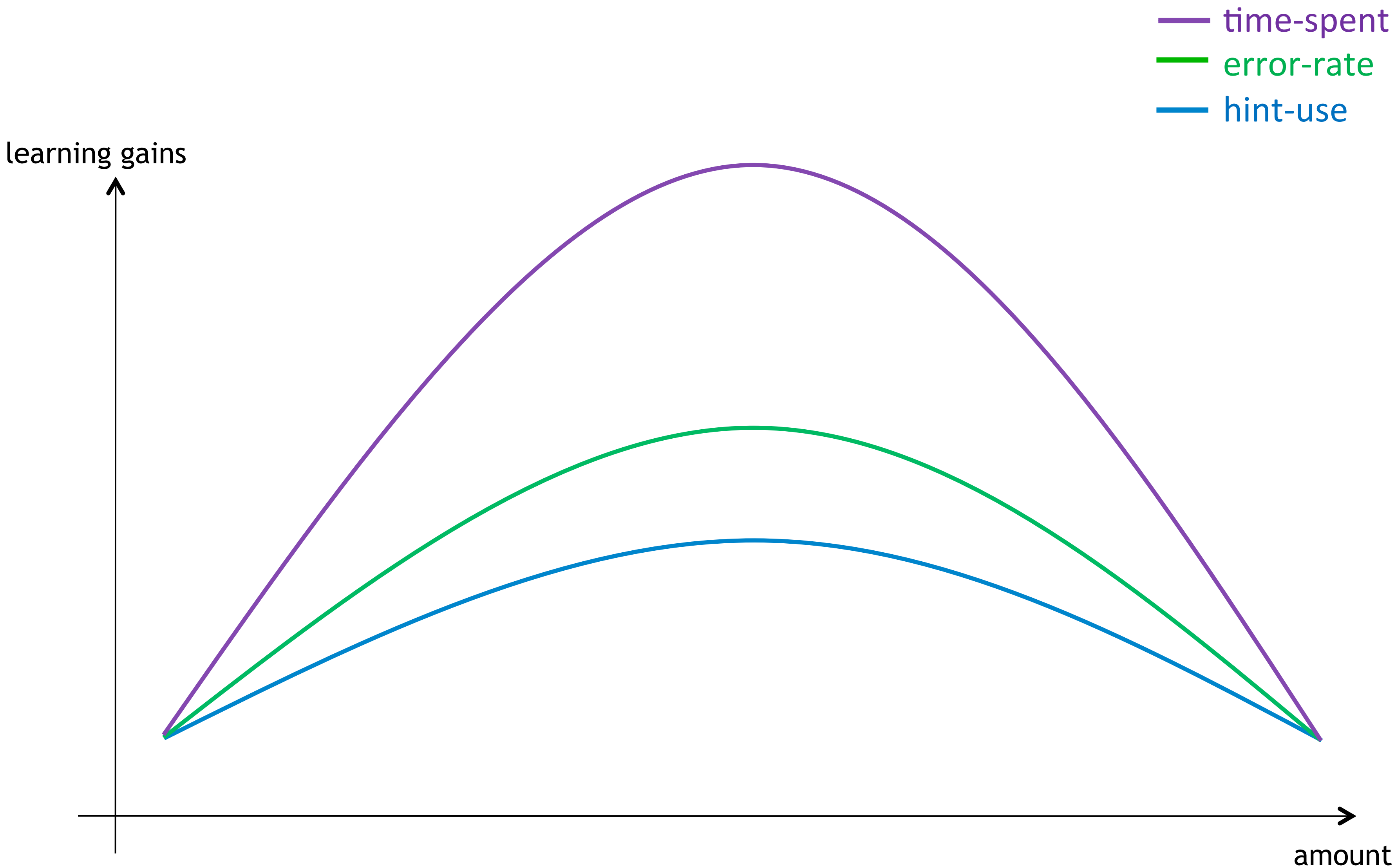
Positive direct effects

Carnegie Mellon University

$\chi^2 = 6.89$, df = 10, $p = .74$

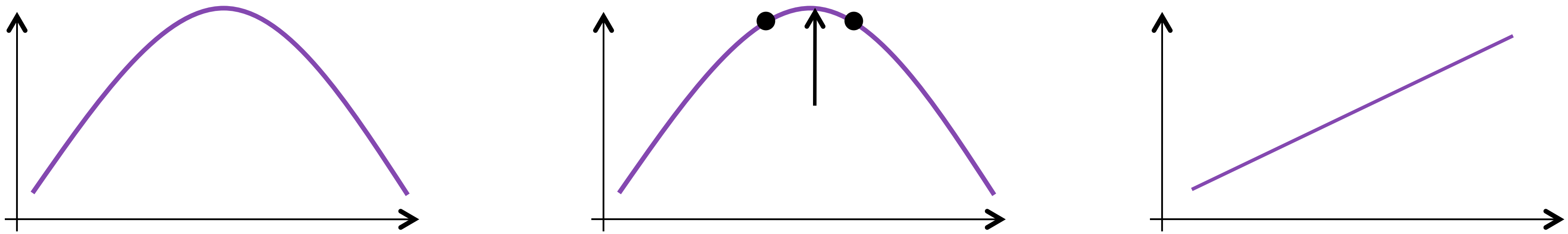
WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Mediator Variables: Non-monotic?



Tranforming/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 distrance) 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]

Experiments 1&2 Conclusions

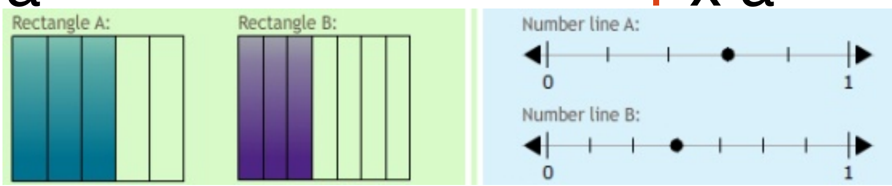
- Multiple representations increase learning
- Standard Variables: *Time*, *Error*, and *Hints* do ***not*** seem to be mechanisms through which multiple representations *increase* learning

4. INFORMED MEDIATORS

Motivation

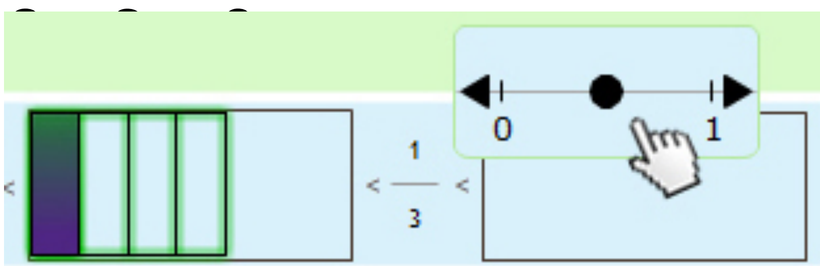
- Learning processes [Koedinger et al., 2012]:
  - Understanding: sense-making processes

a = 1 x a



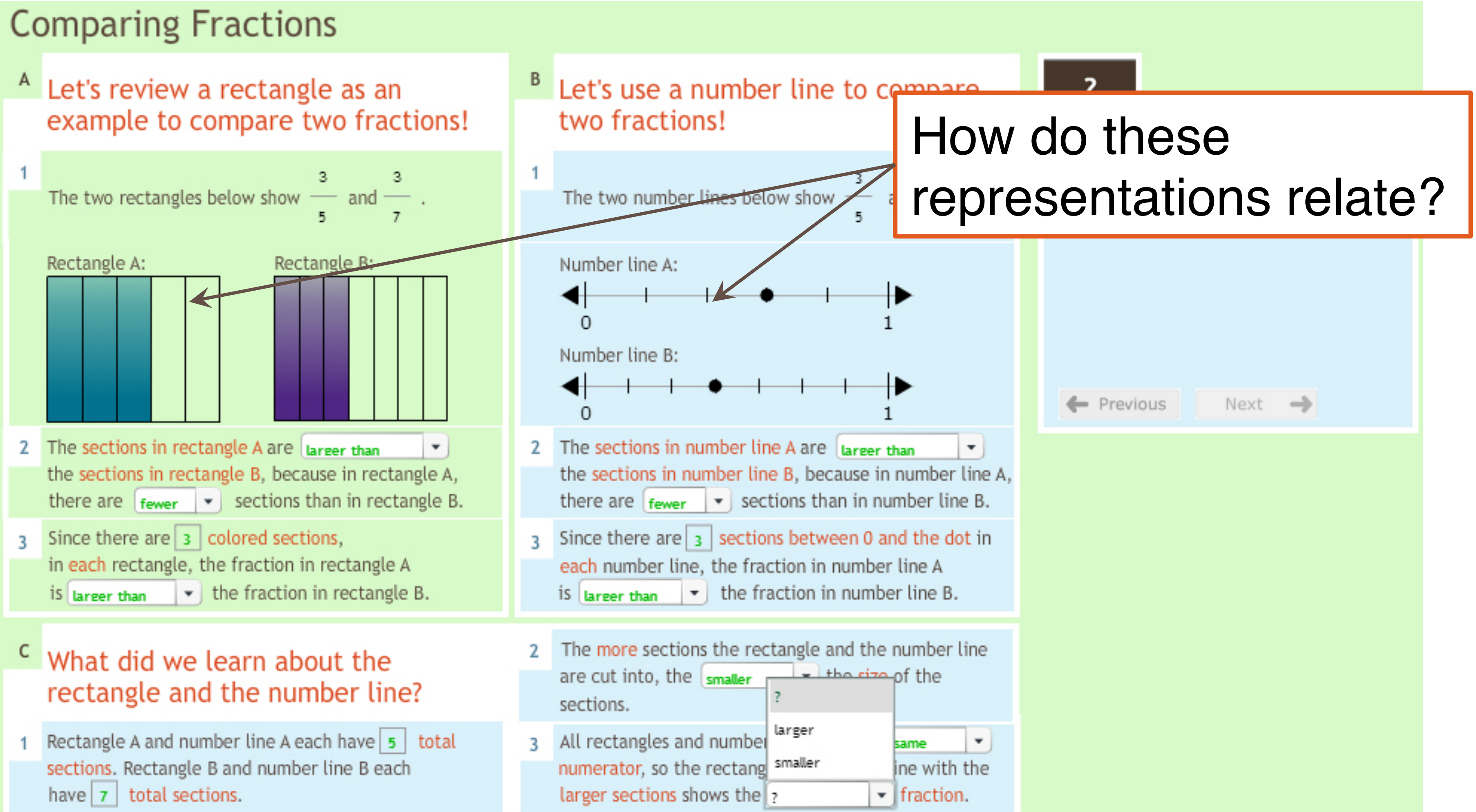
… …

  - Fluency: fluency-building processes

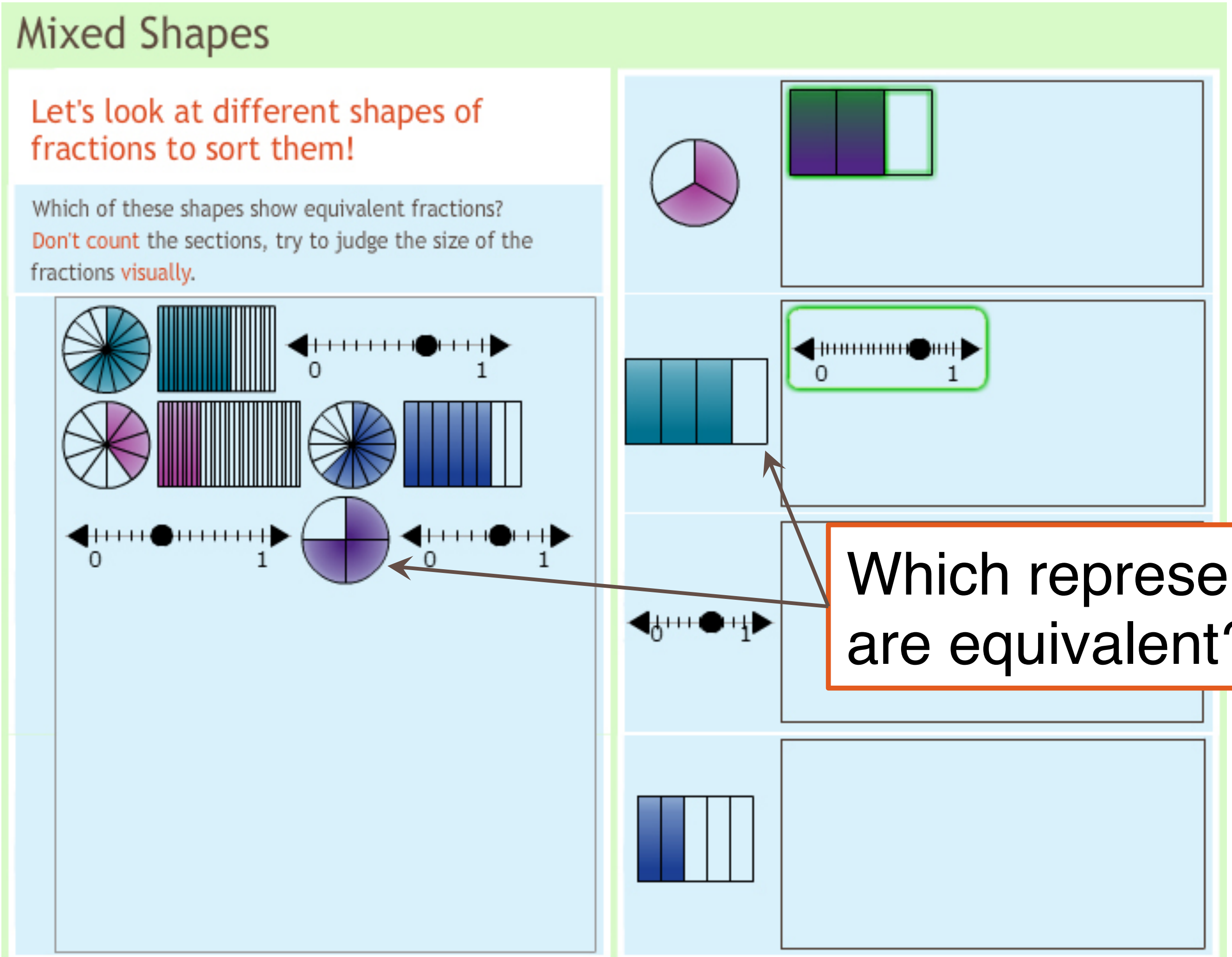


…

Fractions Tutor: Sense-making

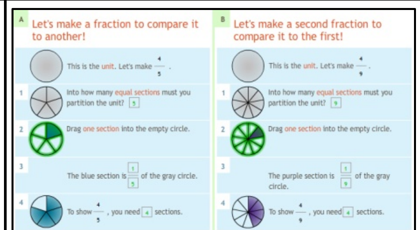
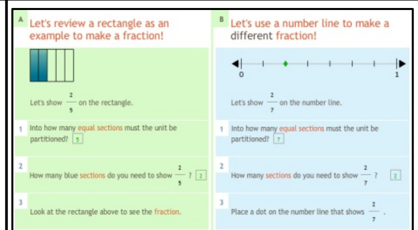
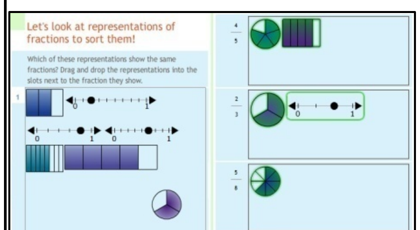
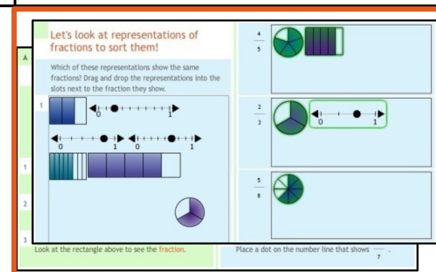
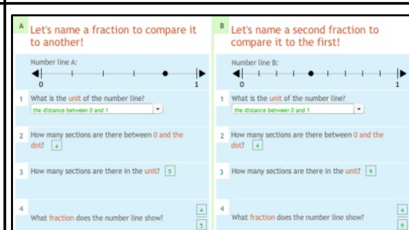


Fractions Tutor: Fluency-building



Which representations are equivalent?

Background: Experiment 3

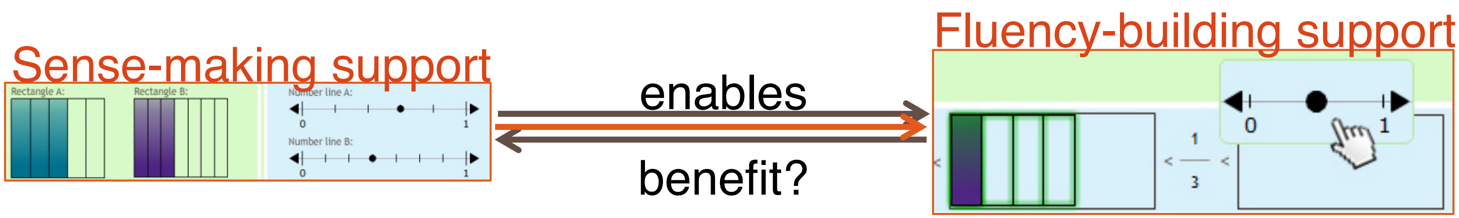
		Sense-making problems	
		no	yes
Fluency-building problems	no		
	yes		
Control			

- N = 599 4th-/5th-graders

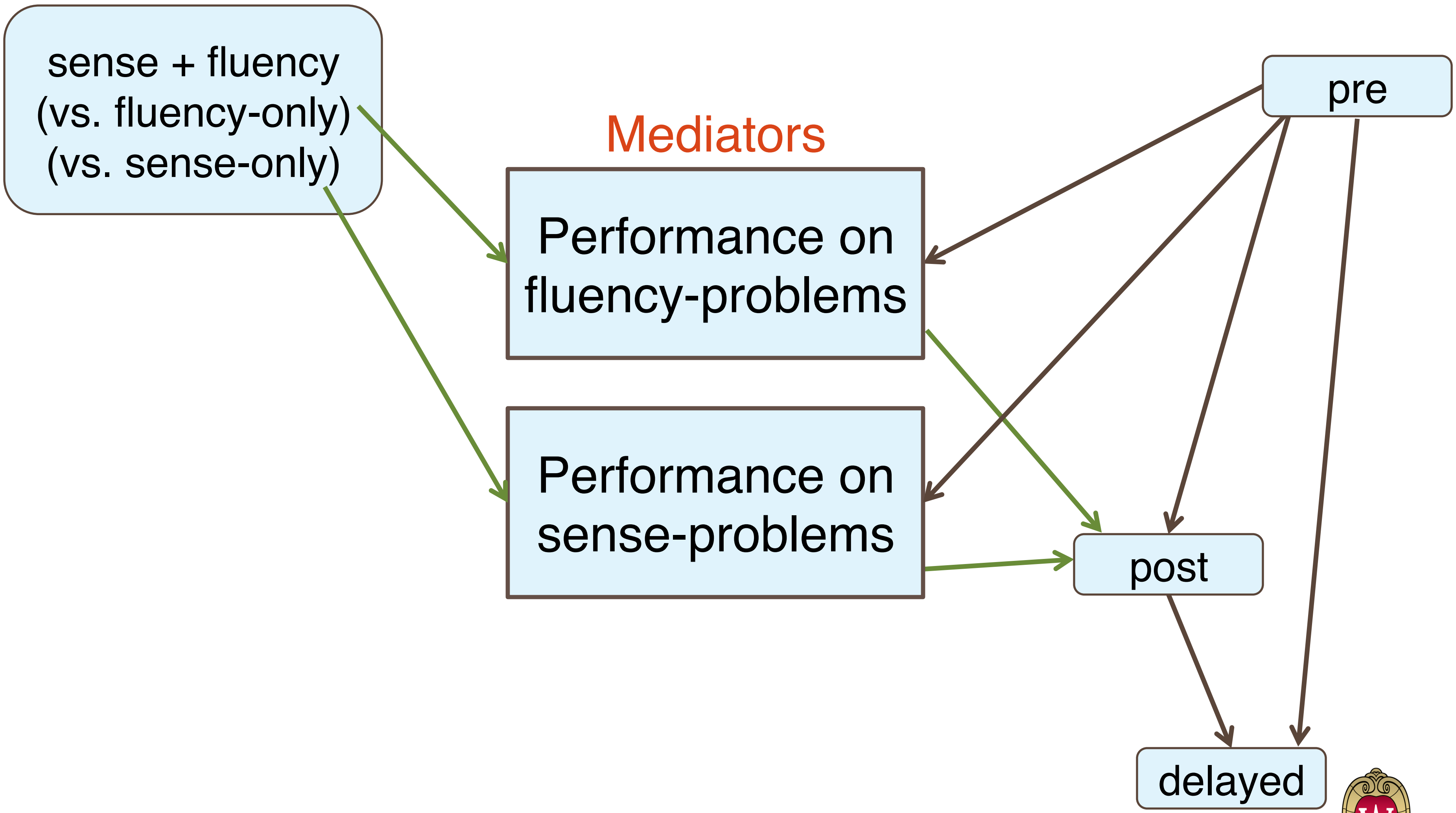
[Rau et al., ITS 2012]

Mediator hypotheses

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



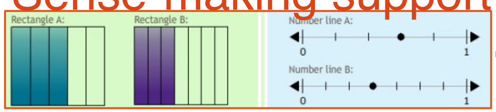
Mediation hypothesis



Mediation Hypotheses

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

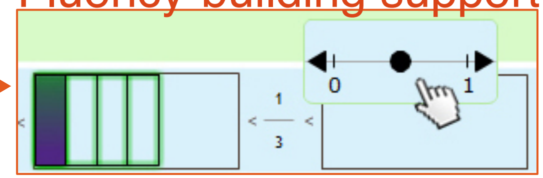
Sense-making support



enables

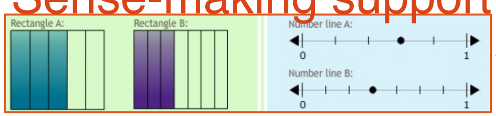
benefit?

Fluency-building support



  - Fluency hypothesis:

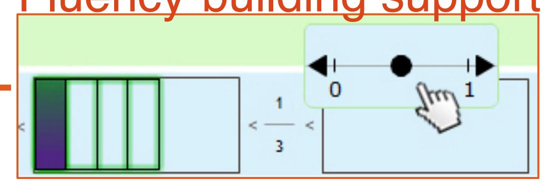
Sense-making support



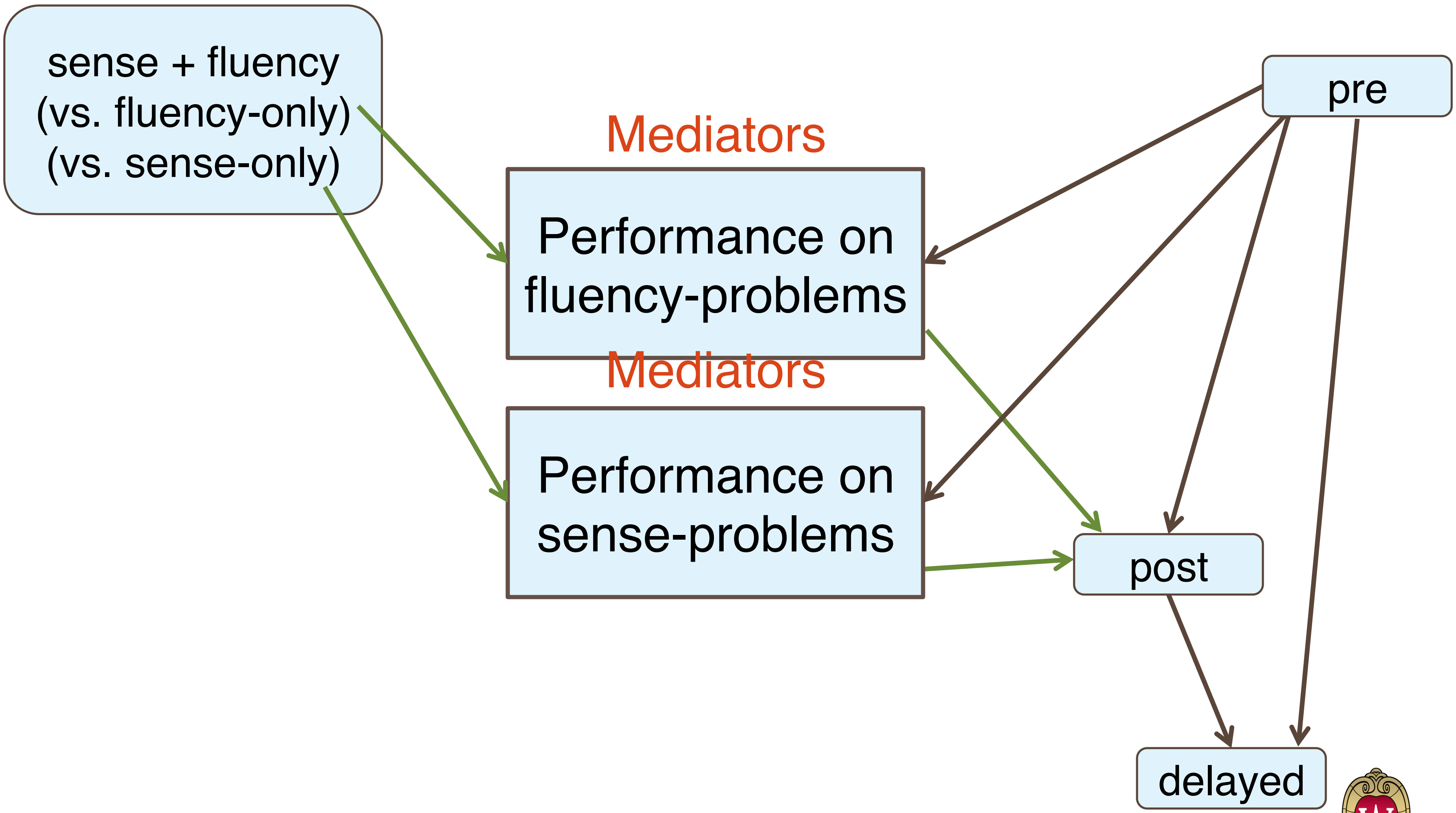
enables

benefit?

Fluency-building support



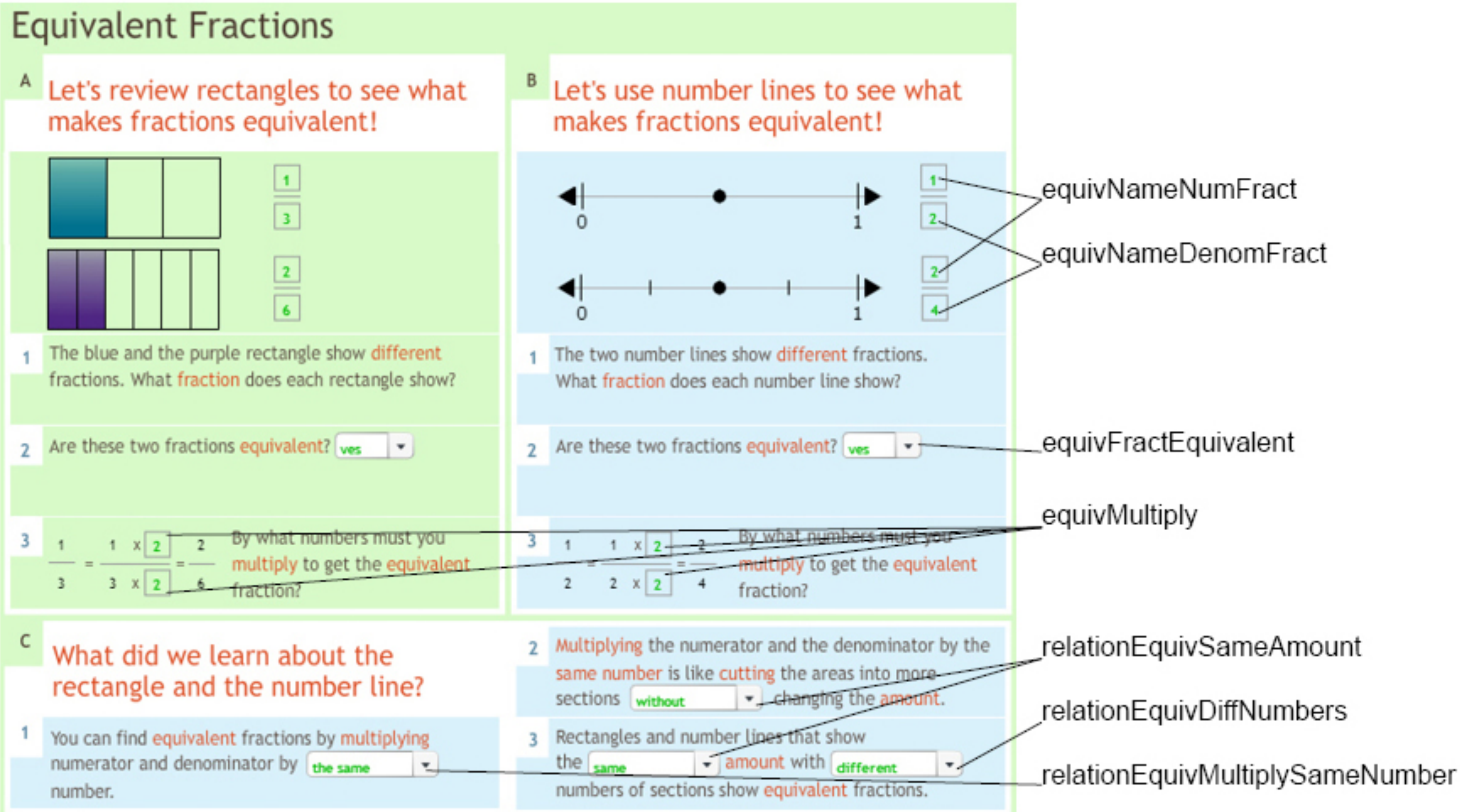
Mediation Hypotheses



Variable identification

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

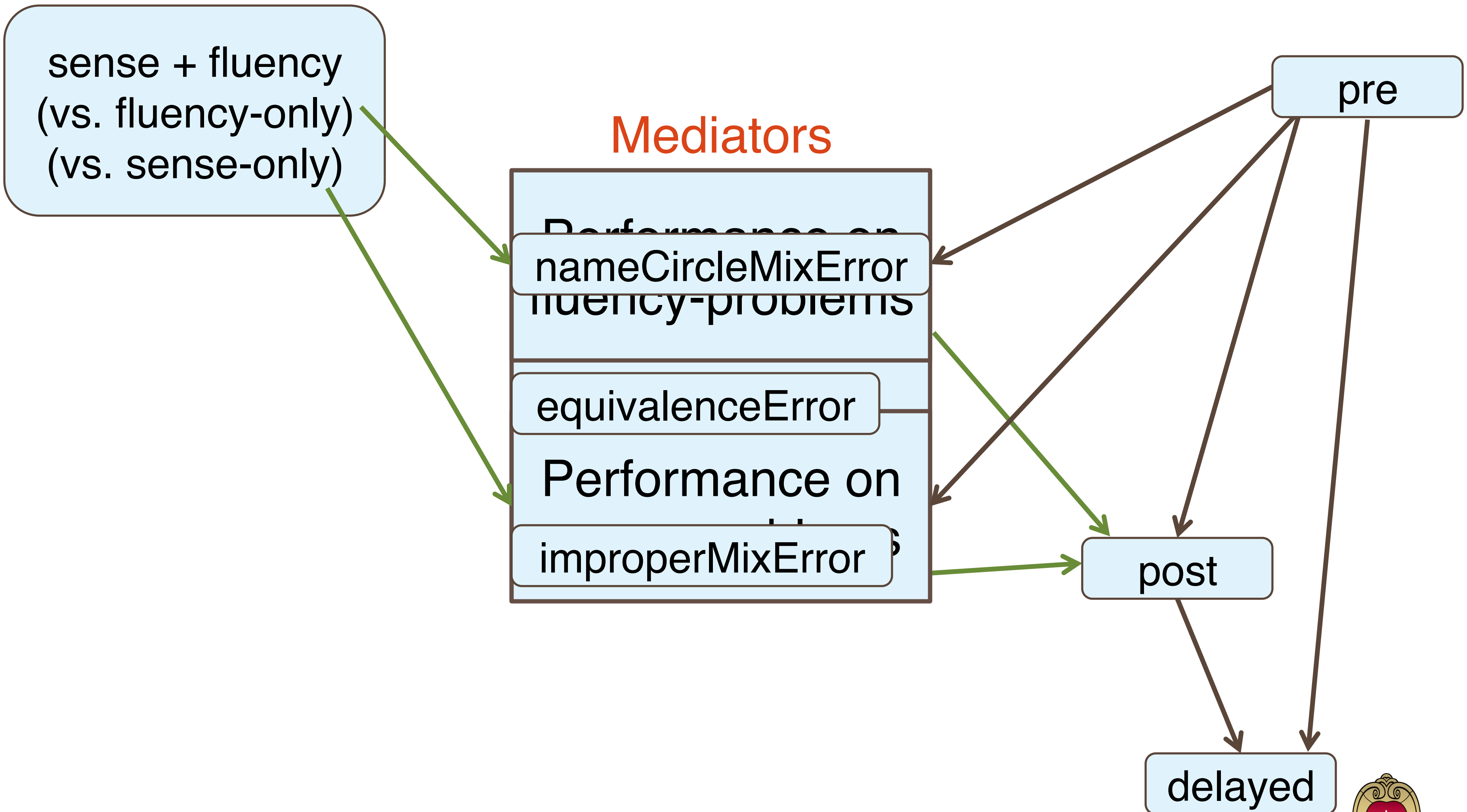
Knowledge Component Model



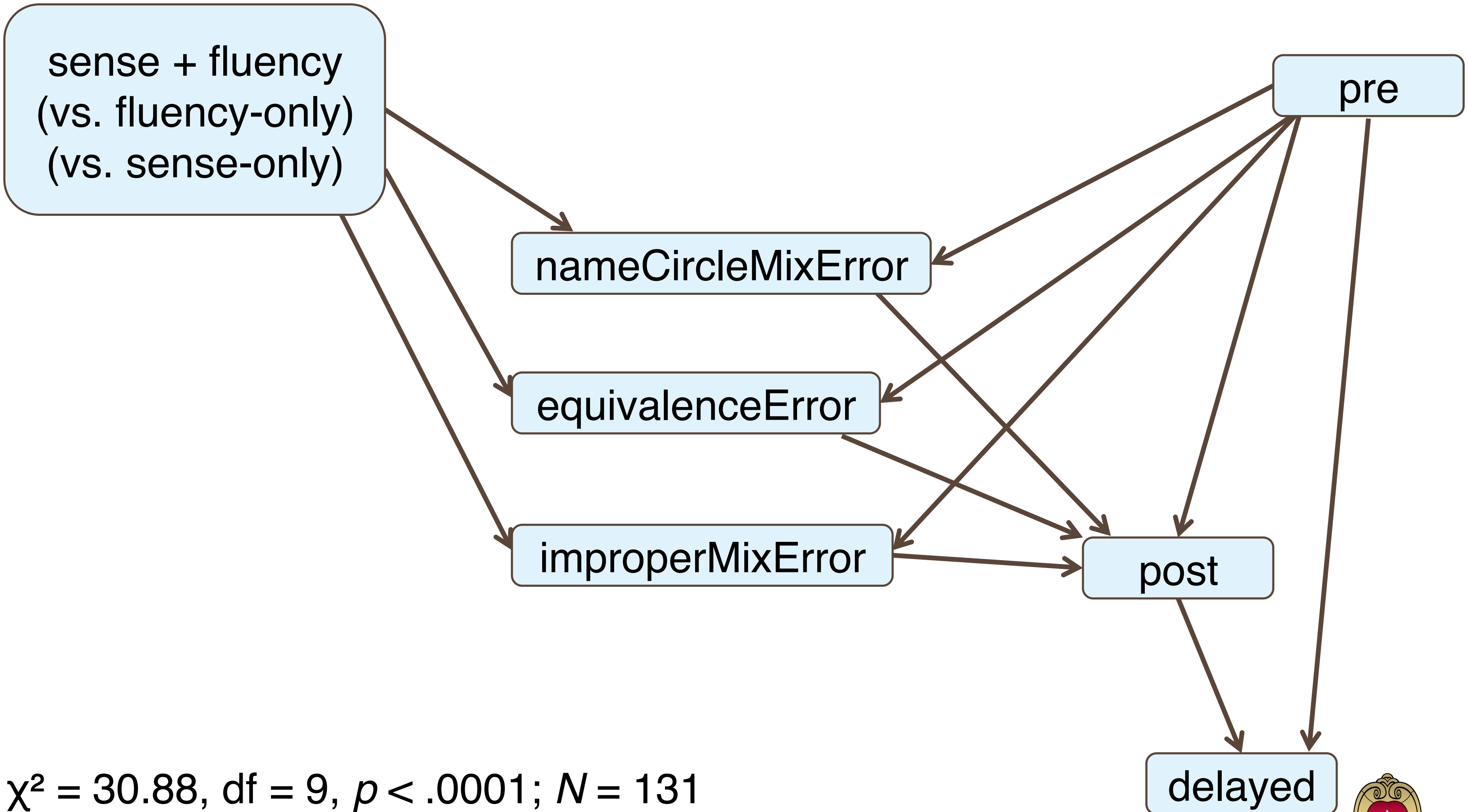
Variable identification

- 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
Mediation hypothesis

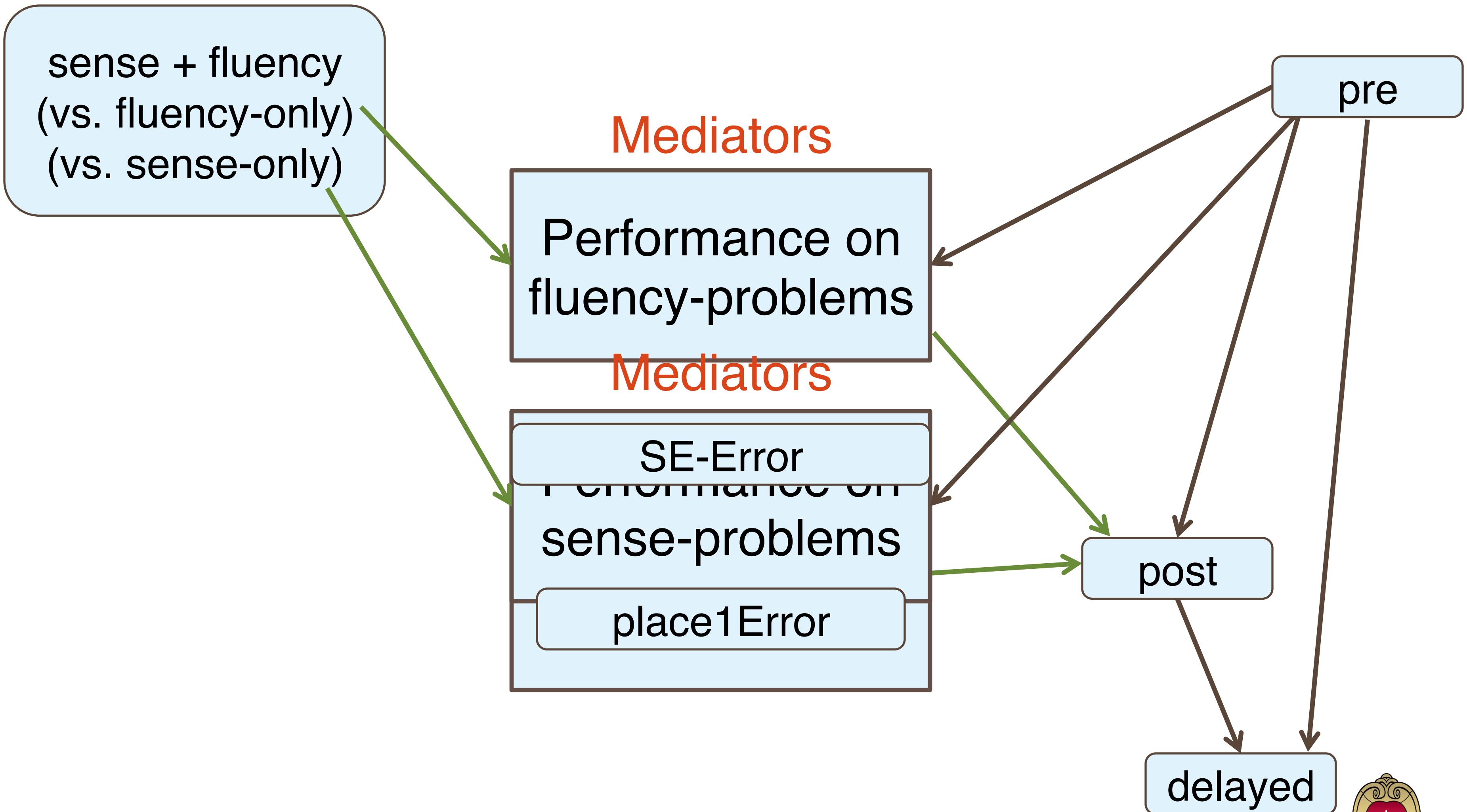


Understanding hypothesis

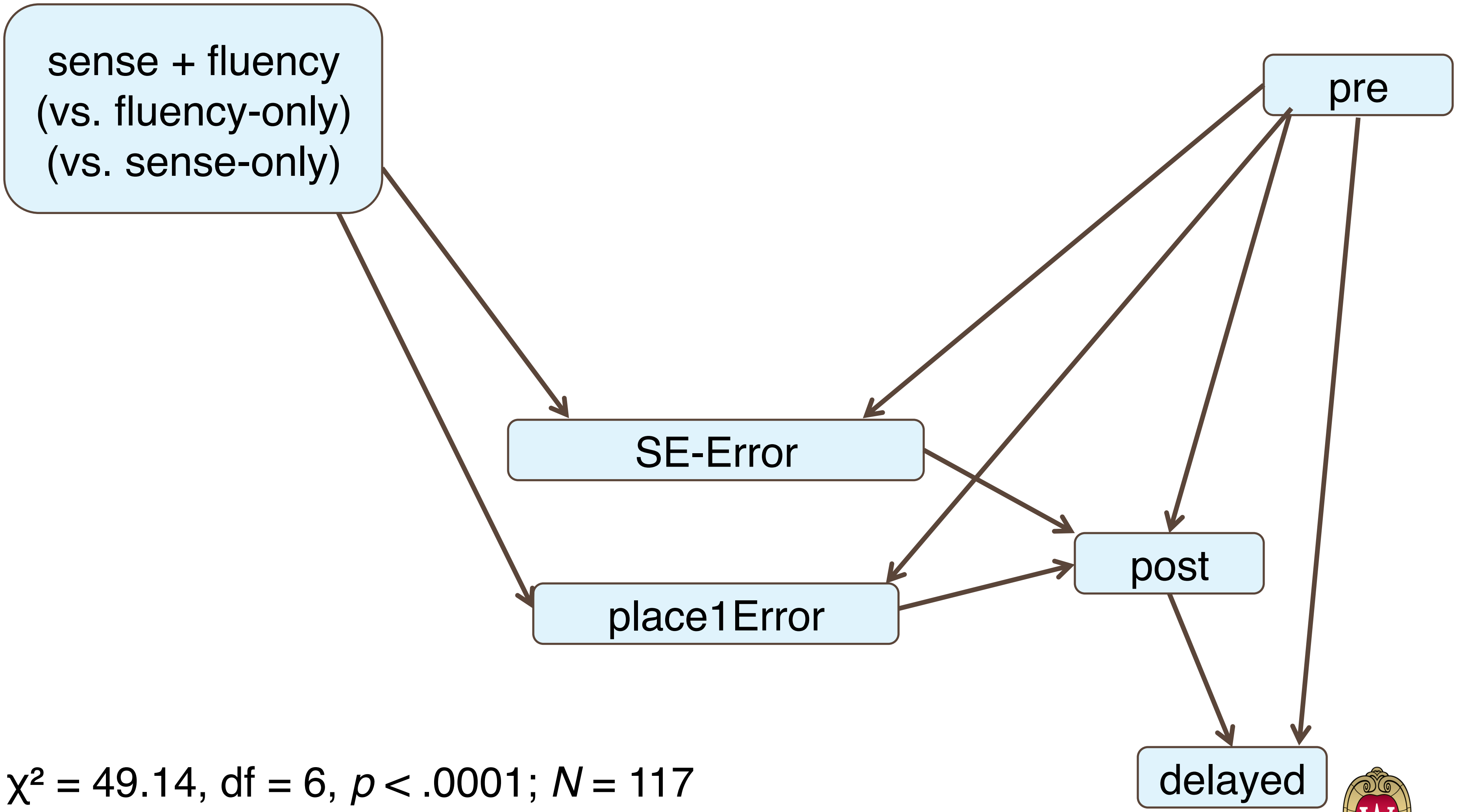


$\chi^2 = 30.88$, df = 9, $p < .0001$; $N = 131$

Fluency hypothesis

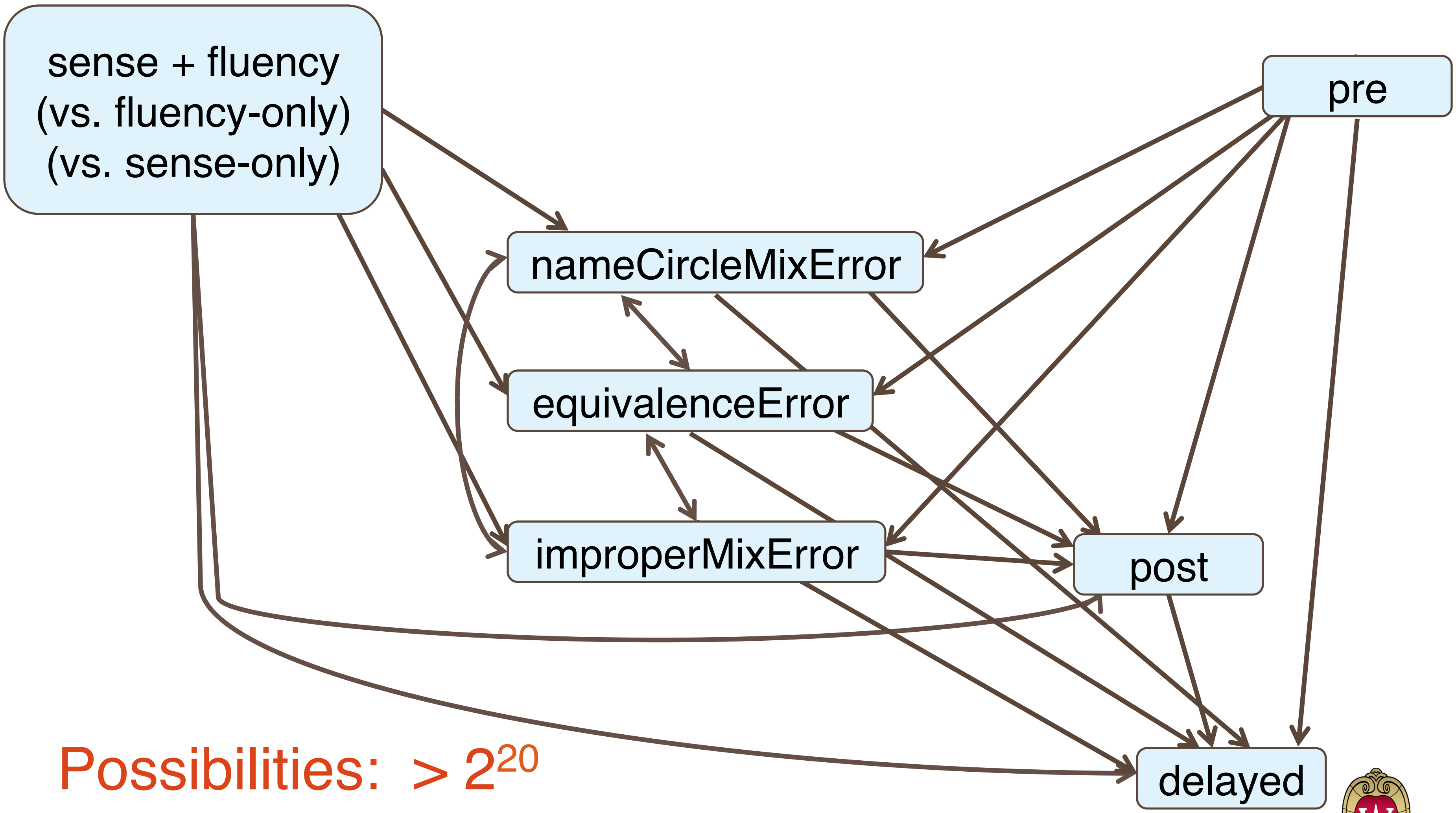


Fluency hypothesis



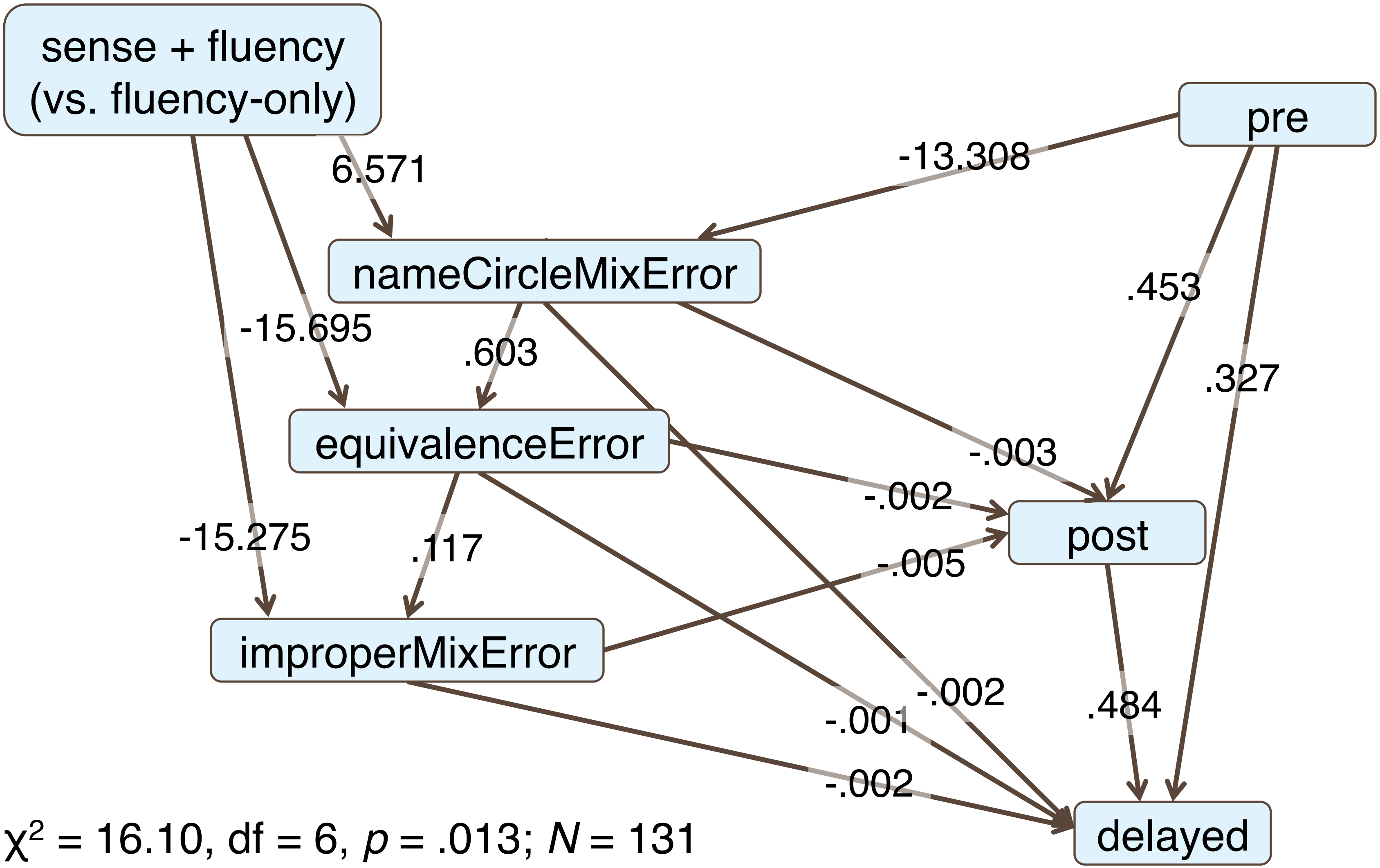
$\chi^2 = 49.14$, df = 6, $p < .0001$; $N = 117$

Possible alternative models



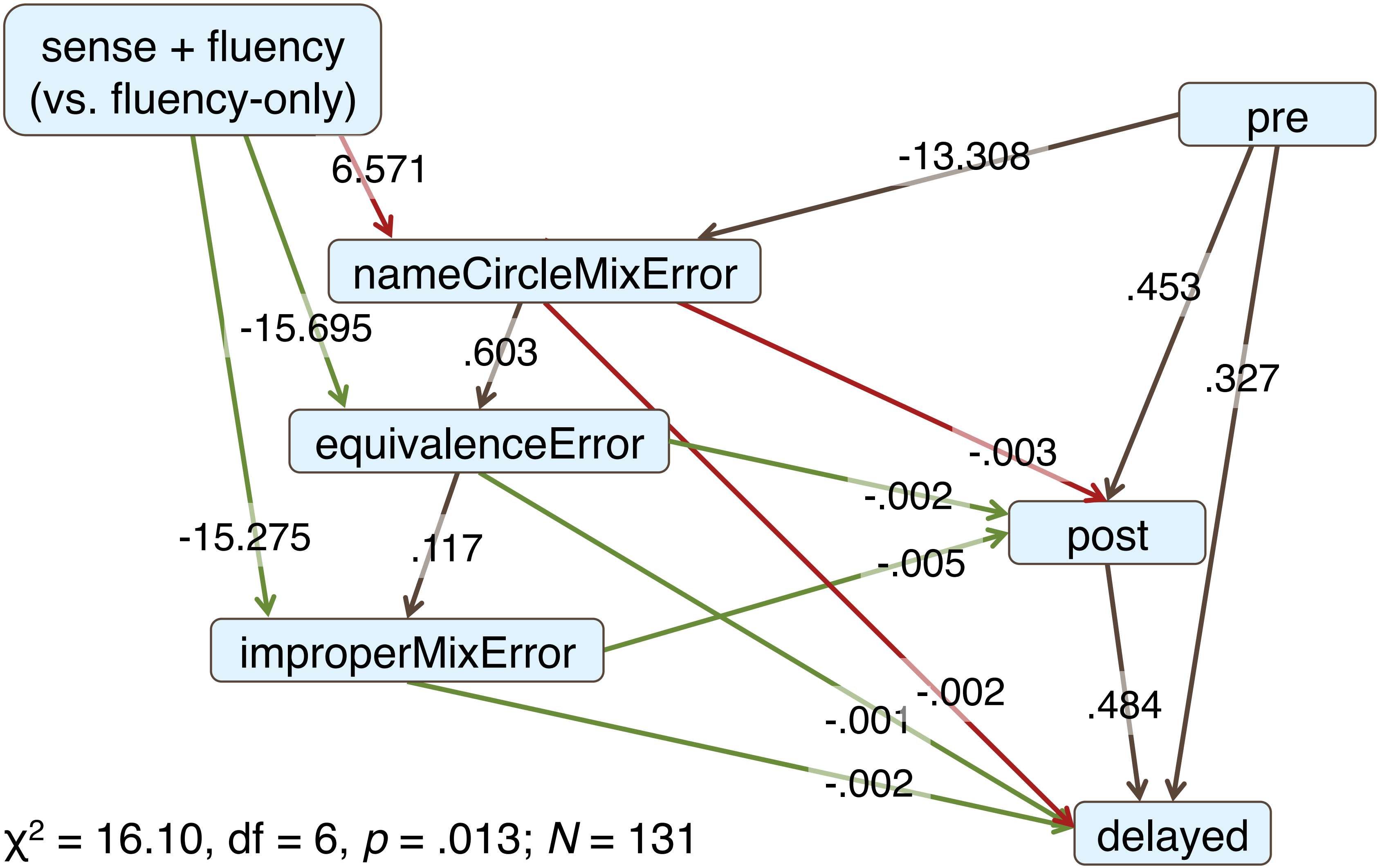
Possibilities: > 2^{20}

Model Search Results: Understanding model



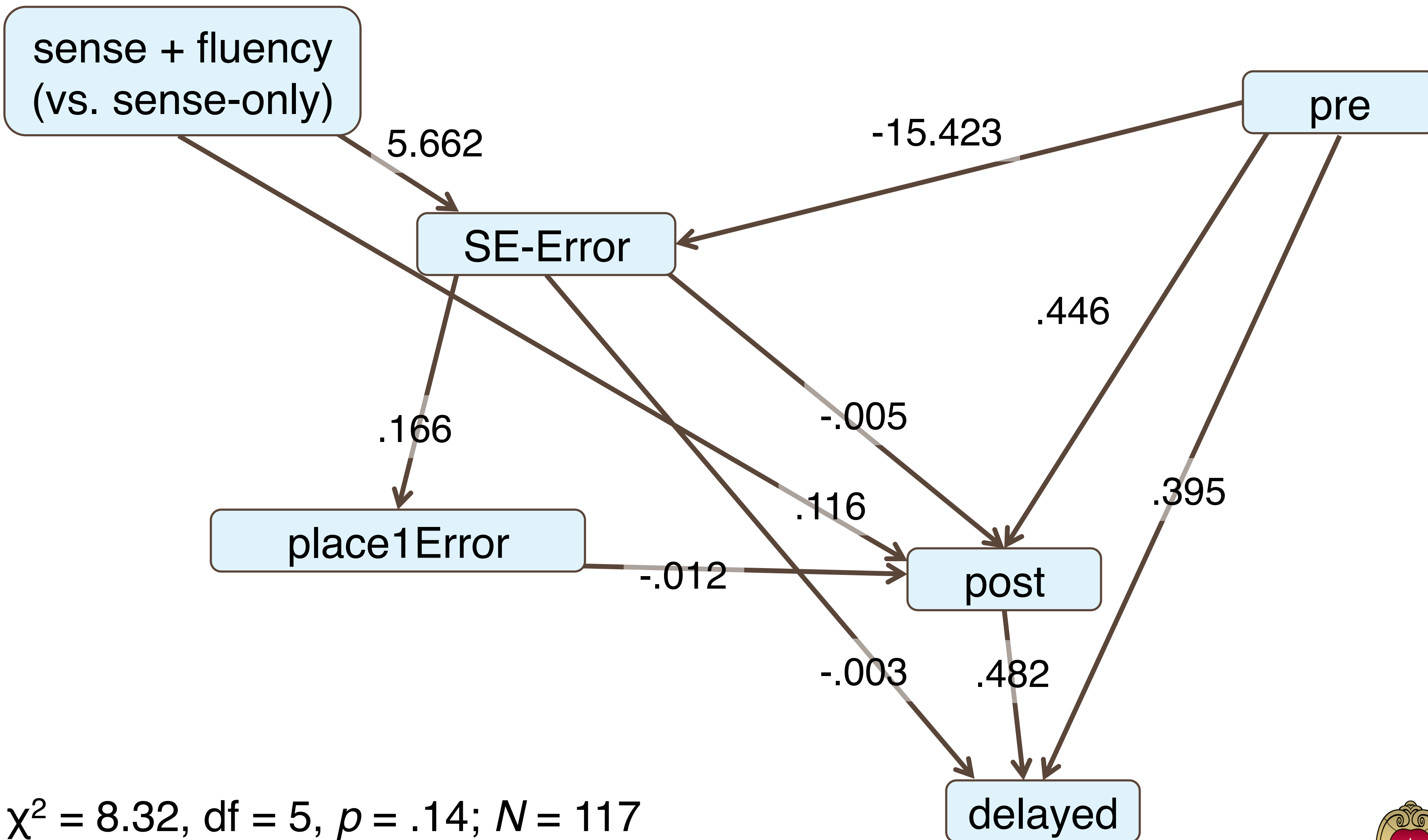
$\chi^2 = 16.10$, df = 6, $p = .013$; $N = 131$

Model Search Results: Understanding model



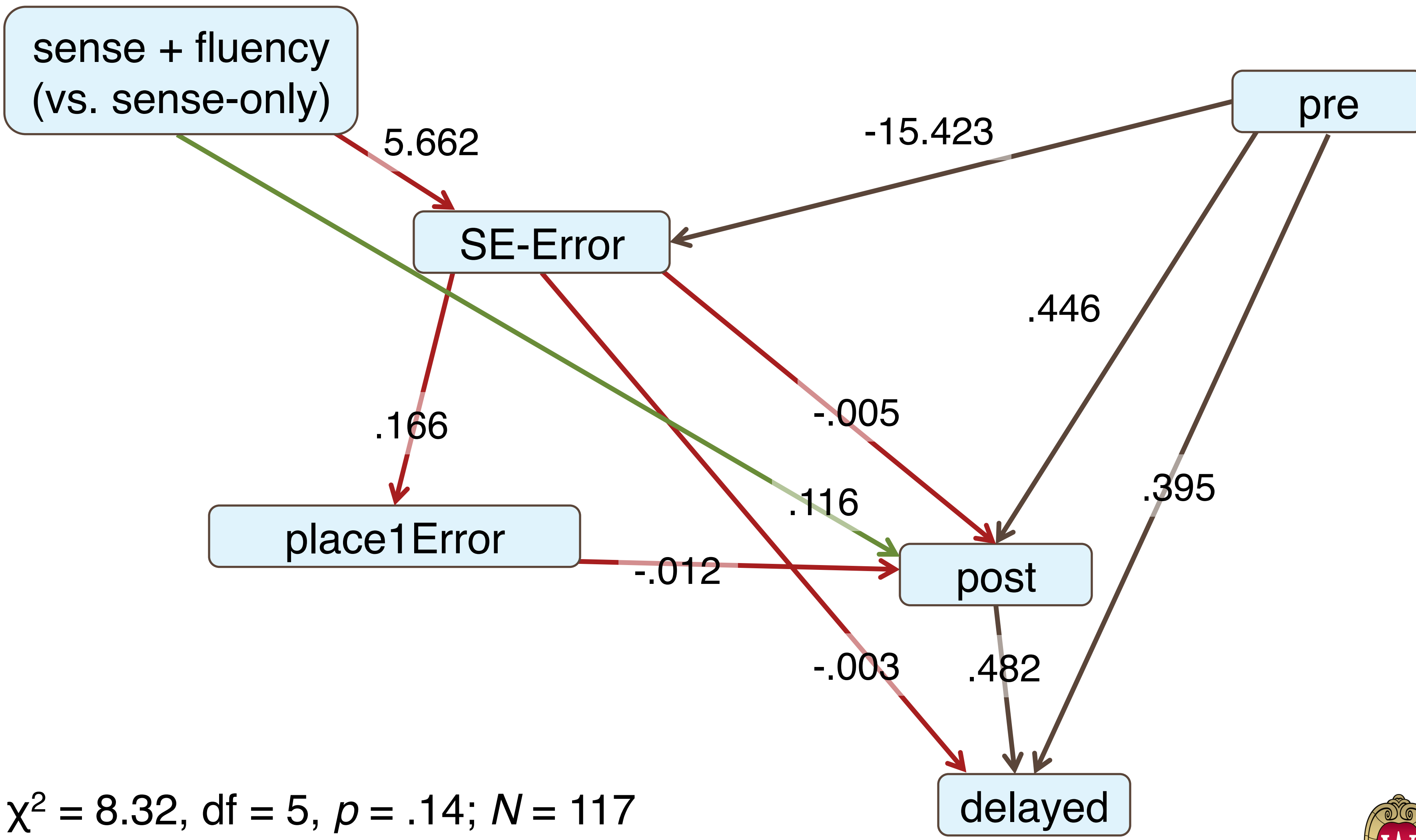
$\chi^2 = 16.10$, df = 6, $p = .013$; $N = 131$

Model Search Results: Fluency model



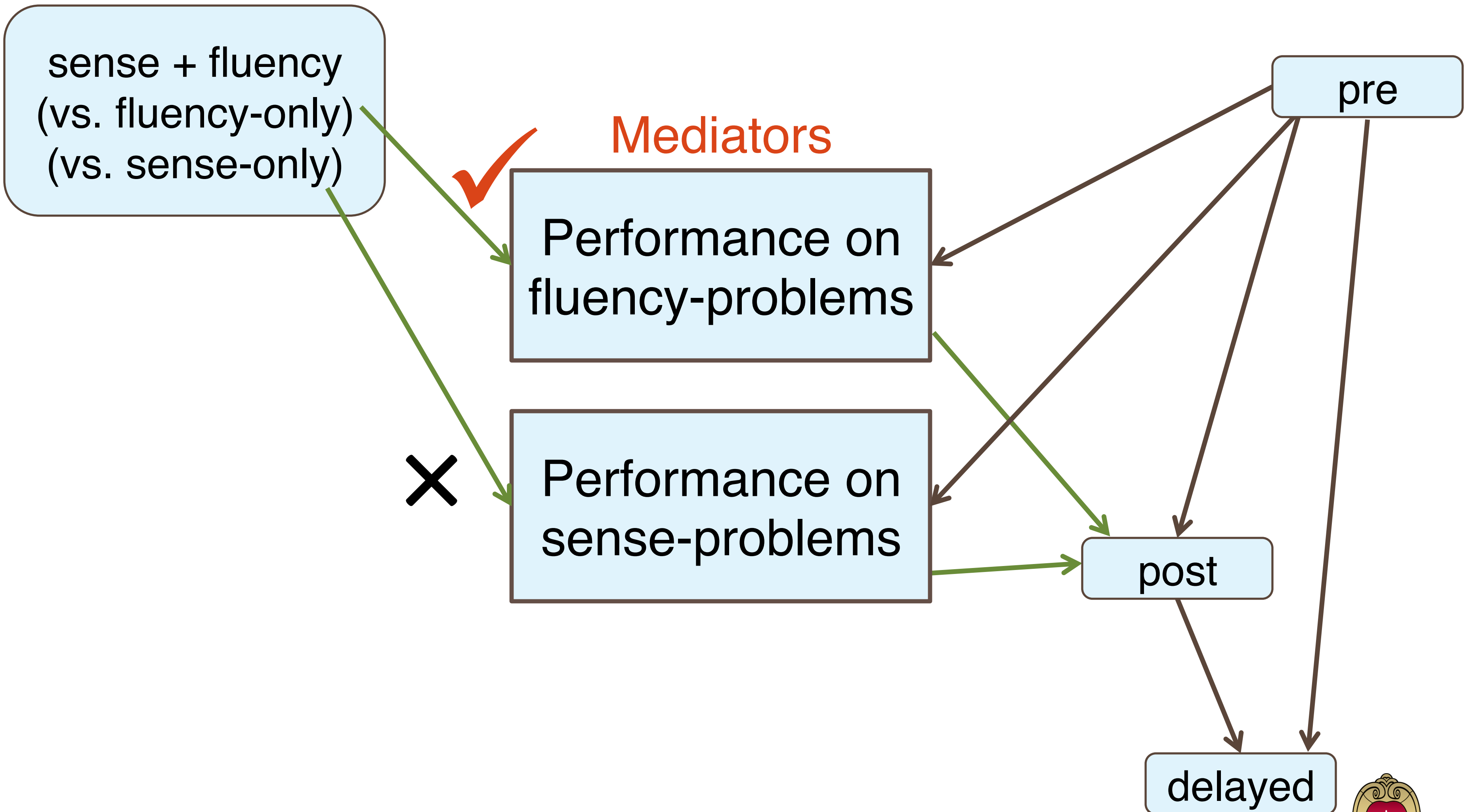
$\chi^2 = 8.32$, df = 5, $p = .14$; $N = 117$

Model Search Results: Fluency model



$\chi^2 = 8.32$, df = 5, $p = .14$; $N = 117$

Mediation hypothesis



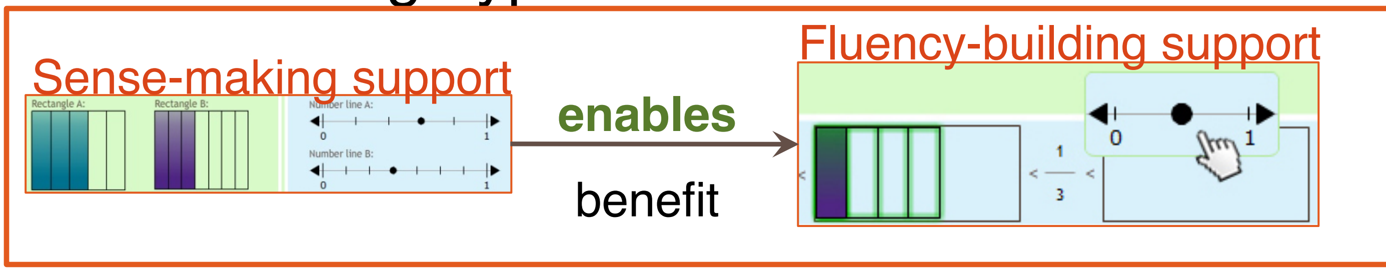
Taking Stock

- Results are in line with understanding hypothesis, but not with fluency hypothesis
  - Sense-making support reduces errors students make on fluency-building problems

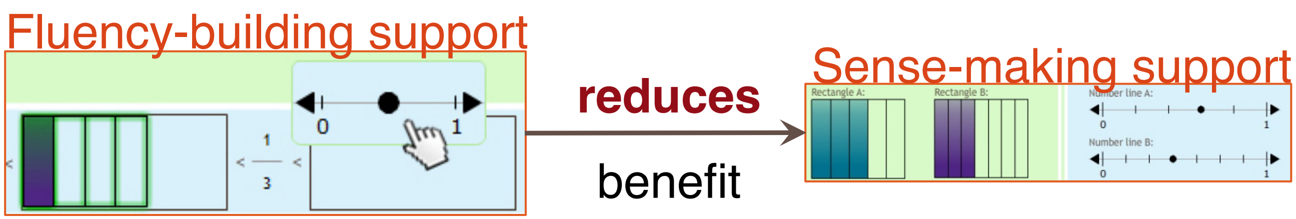
    [Rau, Scheines et al., EDM 2013, best paper]
- Limitations
  - Bound to fixed sequence: sense – fluency
  - Different results possible with sequence fluency – sense
- Makes testable predictions:
  - Sense-making support should be provided *before* fluency-building support

Experiment 4: Results

- Which process should instruction support first?
  - Understanding hypothesis:



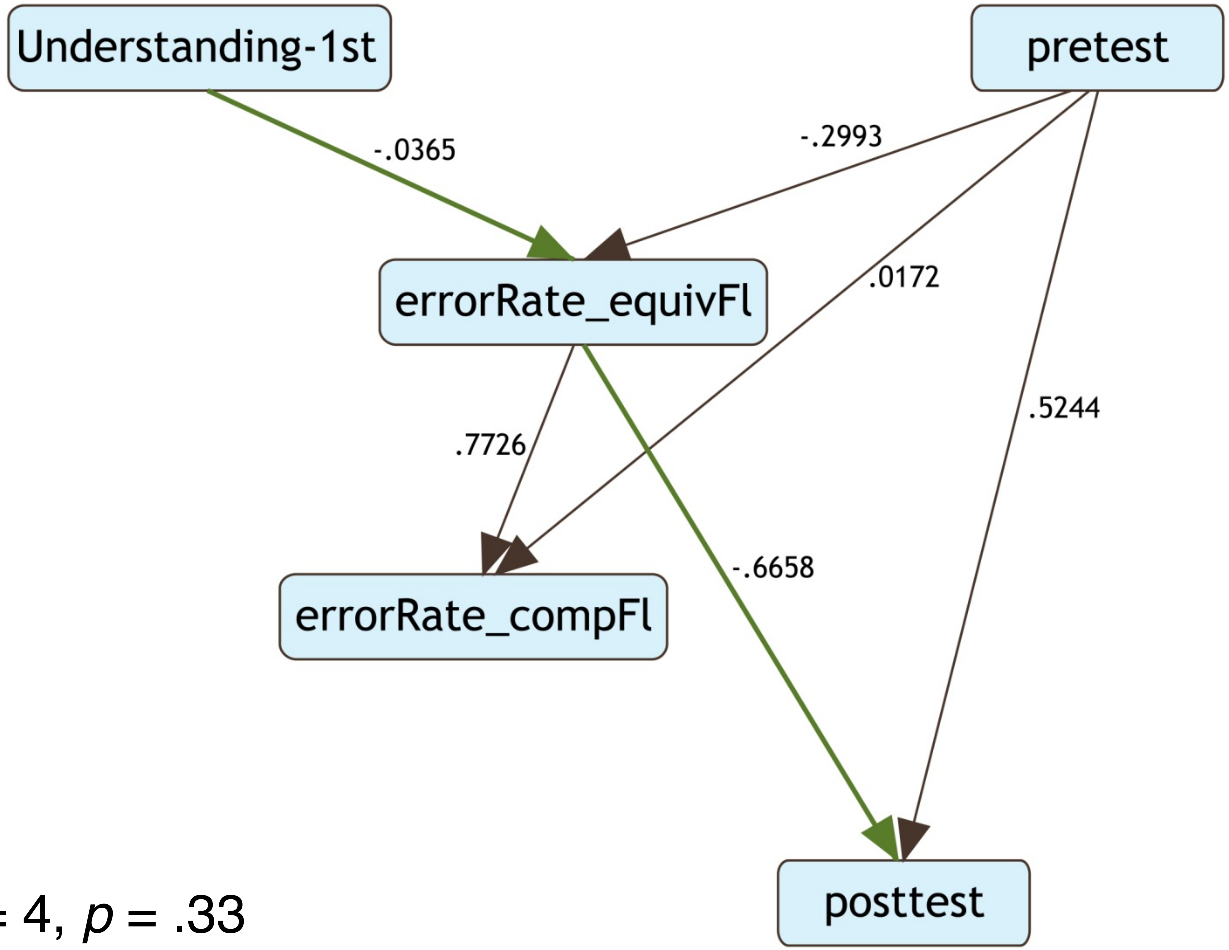
  - Fluency hypothesis:



[Rau et al., AIED 2013]

Experiment 4: Model Search Results

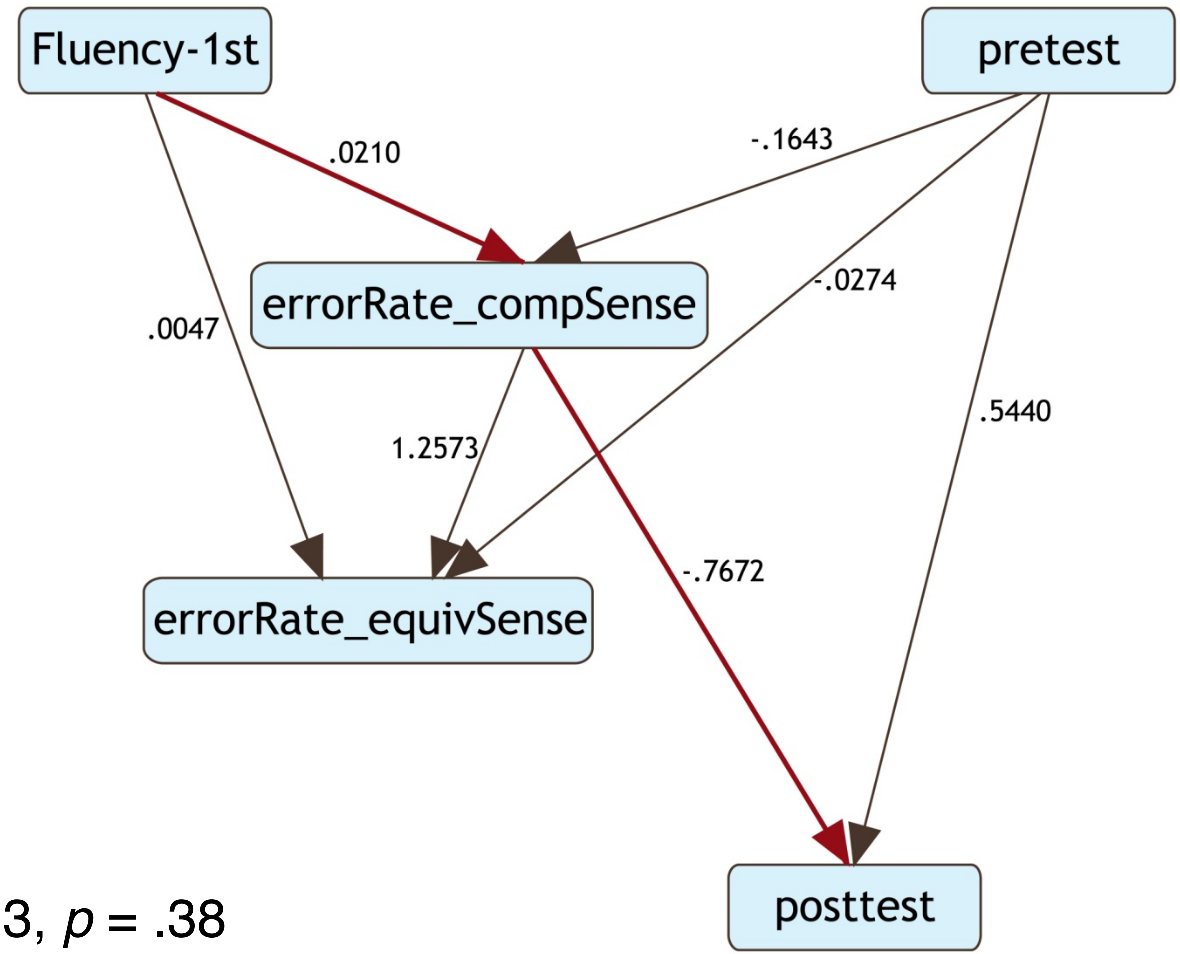
Fluency-building errors



$\chi^2 = 4.58$, 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

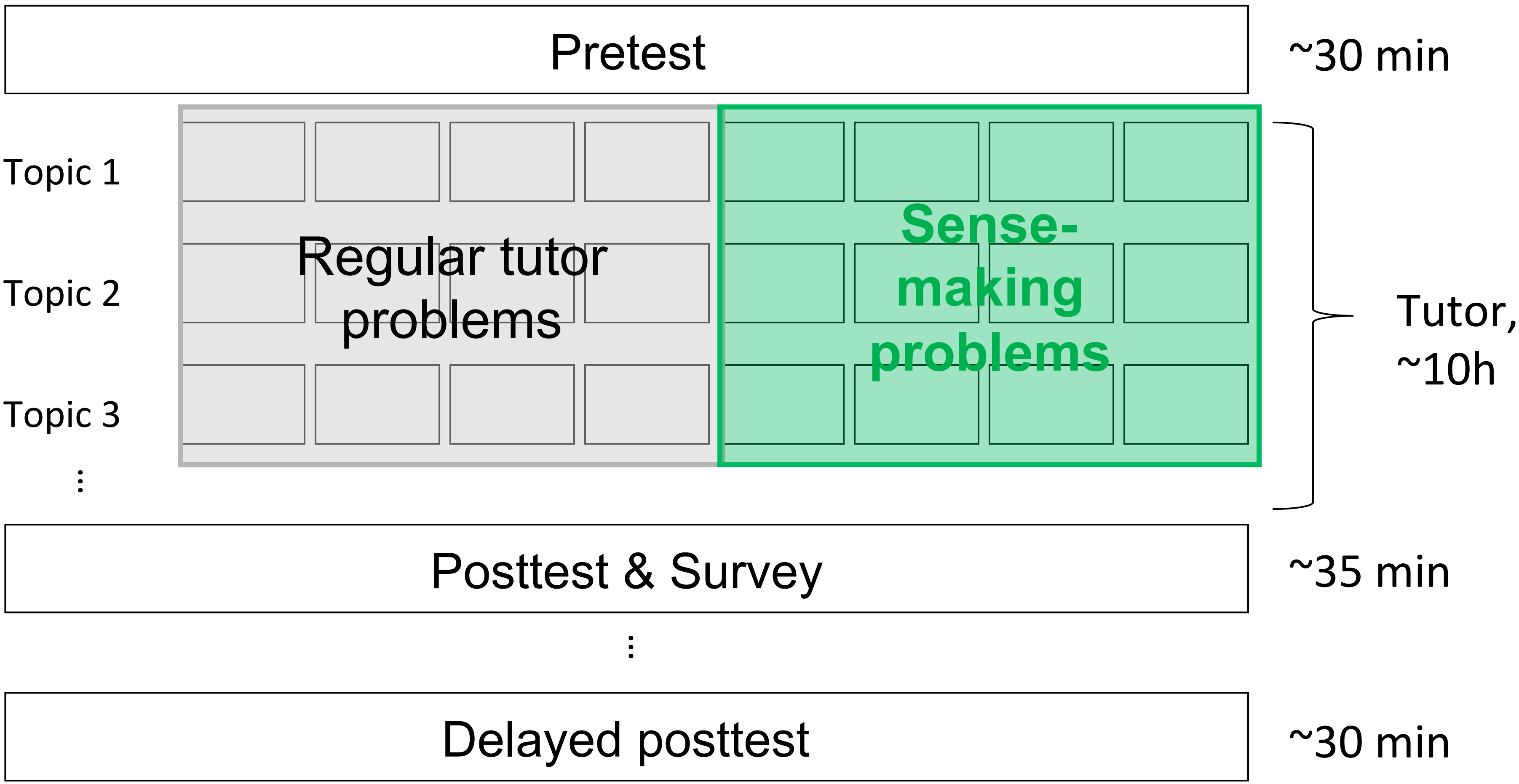
Vincent Aleven Nikol Rummel, Richard Scheines, Ken Koedinger, and Brian Junker

Jay Raspat, Michael Ringenberg, Stacie Rohrbach, Zelha Tunc-Pekkan, Laurens Feenstra, Cathy Kramer, Mitch Nathan, Peg Smith, Cassandra Studer, Jenny Olson, Dan Belenky, Ryan Baker, Neil Heffernan, Shaaron Ainsworth, Janice Gobert, David Pritchard, David Klahr, Sharon Carver, Howard Seltman, Jack Mostow, Anna Fisher, Audrey Russo, Michael Bett, Gail Kusbit, Jo Bodnar, Queenie Kravitz, Jo Bodnar, Lisa Vento, Mark Penney, Katelyn Sittler, Laura Butler, Anthony DeMauro, Adriana Joazeiro, Julika Lomas, Corinne Rockoff, Shawn Snyder, Orly Stampfer, Cassandra Studer, Angela Wagner, Jonathan Sewall, Alida Skogsholm, Brett Leber, Octav Popescu, Cindy Tipper, John Stamper, Kyle Cunningham, Jessica Kalka, Mike Komisin, Antonia Becher, Laura Pacilio, Lyubov Zeylikman, Lisa Kwon, Angela McCarthy, Lavender Yi, Siyan Zhao, Eiji Hayashi, Iris Howley, Sunyoung Kim, Yanjin Long, Gabi Marcu, Jenn Marlow, Bryan Matlen, Rony Patel, Kelly Rivers

Students and teachers who have participated in my research

BACKUP SLIDES

Sense-only condition



Fluency-only condition

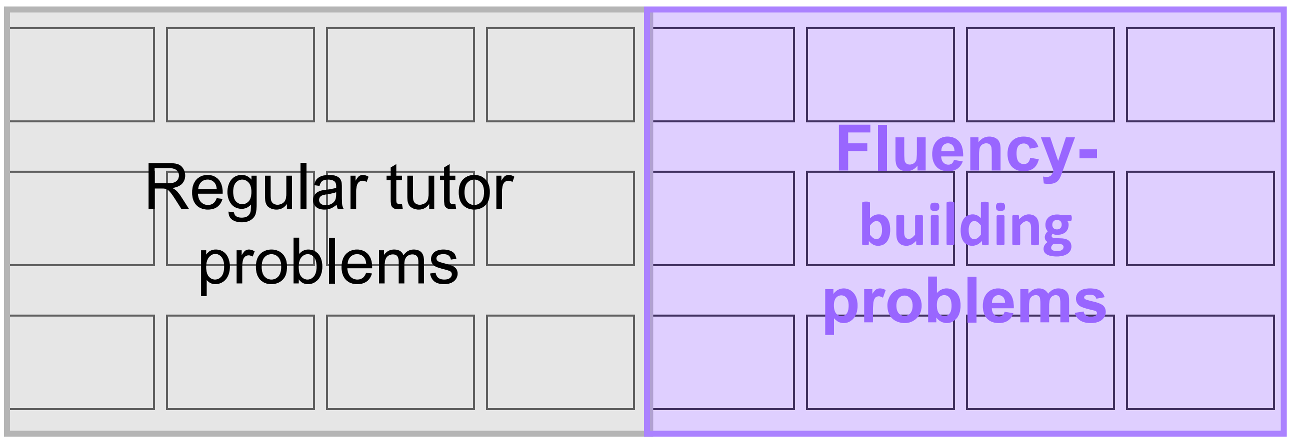
Pretest ~30 min

Topic 1

Topic 2

Topic 3

⋮



Tutor, ~10h

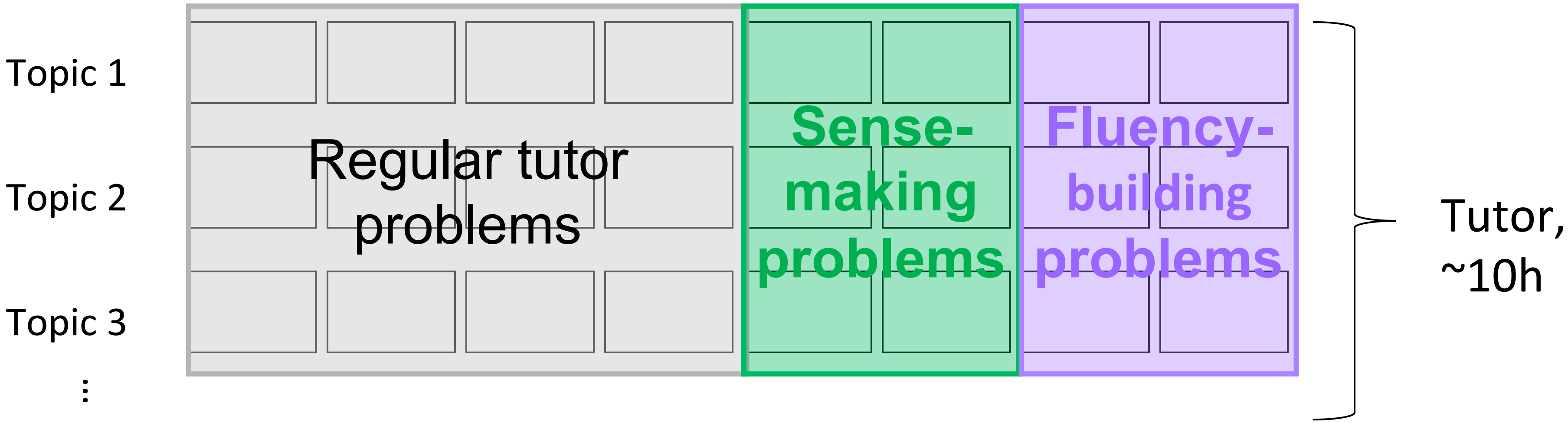
Posttest & Survey ~35 min

⋮

Delayed posttest ~30 min

Sense + fluency condition

Pretest ~30 min

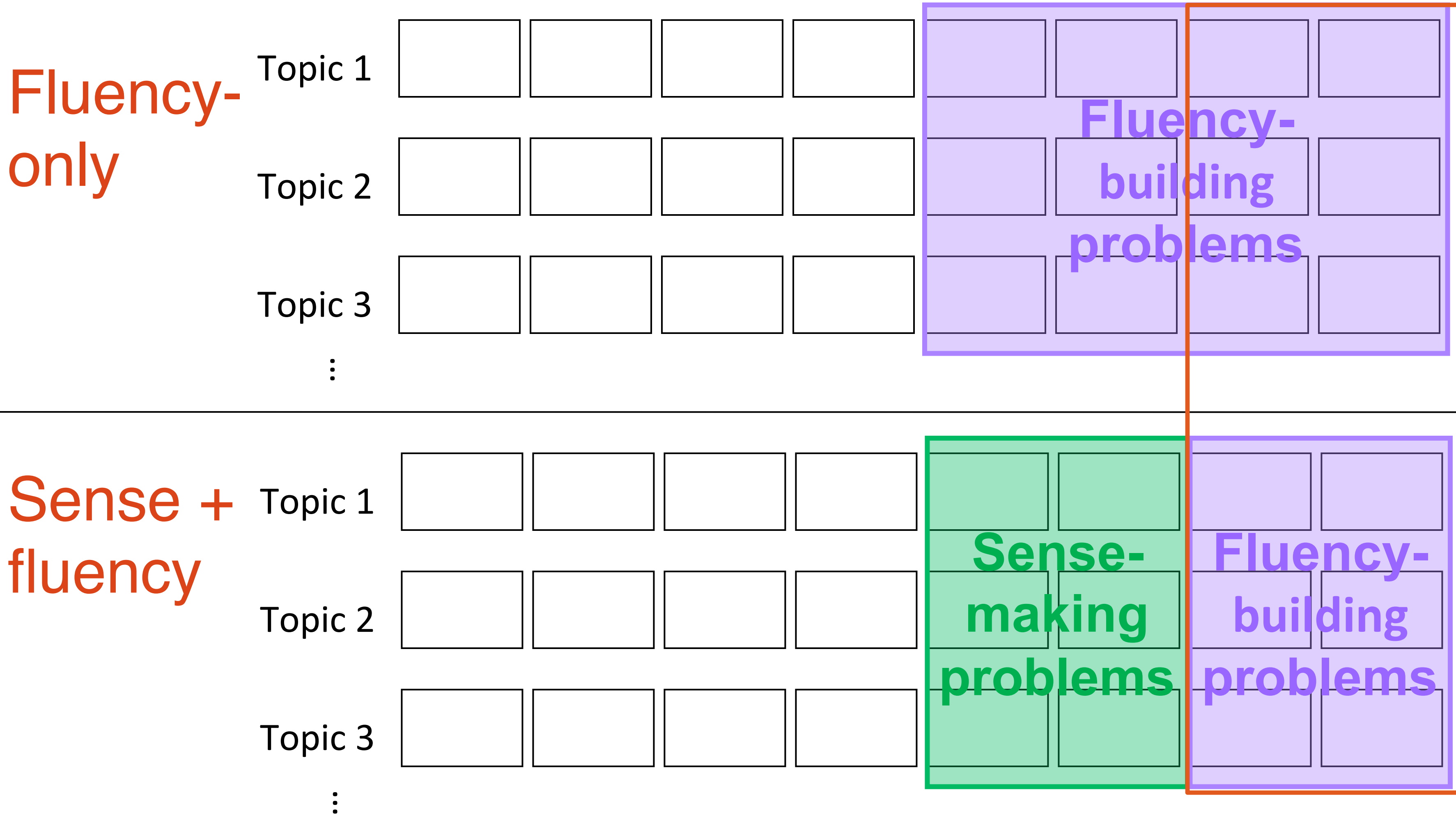


Posttest & Survey ~35 min

⋮

Delayed posttest ~30 min

Fluency-only vs. sense + fluency



Sense-only vs. sense + fluency