Teaching Causal and Statistical Reasoning

Richard Scheines

and *many* others

Philosophy, Carnegie Mellon University

Outline

Today

1. The Curriculum
2. The Online Course
   - Modules
   - Case Studies
   - Support Materials
   - Causality Lab
3. Learning Studies

Tomorrow

1. The Causality Lab
   - Doing Exercises
   - Authoring Exercises
2. Pilot Studies

Motivation

- In College Curriculum :

  Statistical Methods ubiquitous – Causation rare

- Empirical Research Methods experience

- We have a theory to sell

- We got a grant – ugh.

Educational Goals

- Providing an extensive introductory treatment of the modern theory of causation and its relationship to statistical ideas
- Equipping students with the analytical tools needed to critically assess social and behavioral “studies” reported in the press
- Providing a foundation for more advanced work in Causal Bayes Networks

Causal and Statistical Reasoning Curriculum

1. Causation

2. Association and Independence

3. Causation → Association

4. Association → Causation

Causation

- **Foundations** (Events, Kinds of Events, Variables, Populations an Samples)
- **Causation Among Variables**
  - Deterministic Causation
  - Indeterministic Causation
- **Representation:**
  - Causal Graphs
  - Modeling Ideal Interventions

Direct Causation

X is a direct cause of Y relative to **S**, iff

$$\exists \mathbf{z}, x_1 \neq x_2 \quad P(Y \mid X \text{ set= } x_1, \mathbf{Z} \text{ set= } \mathbf{z})$$

$$\neq P(Y \mid X \text{ set= } x_2, \mathbf{Z} \text{ set= } \mathbf{z})$$

where **Z** = **S** - {X,Y}

$$X \longrightarrow Y$$

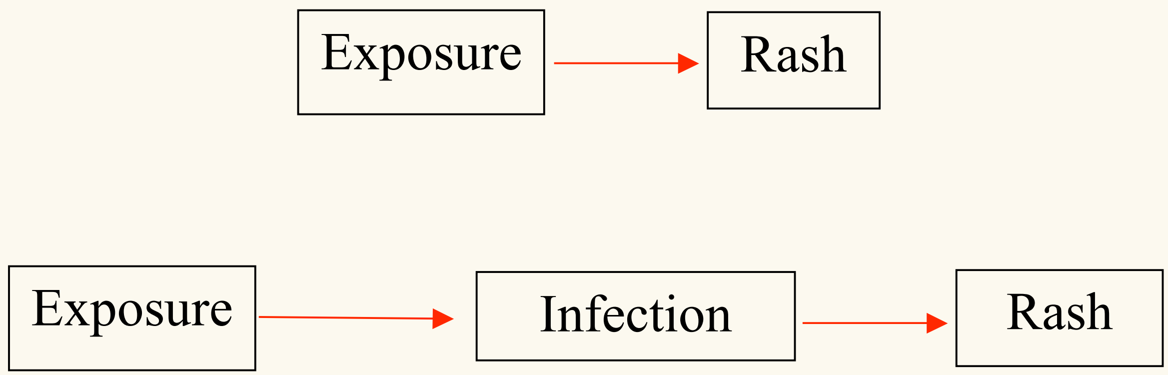
Causal Graphs

Causal Graph G = {**V**,**E**}

Each edge X → Y represents a direct causal claim:

X is a direct cause of Y relative to **V**

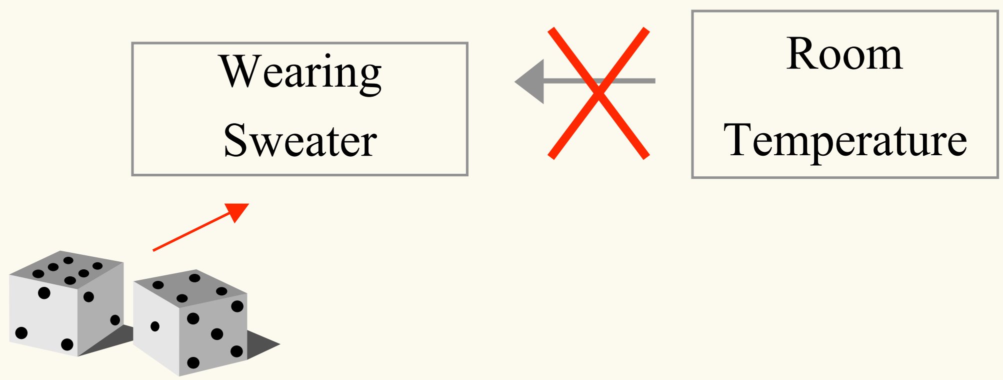
Chicken Pox



Modeling Ideal Interventions

Interventions on the Effect

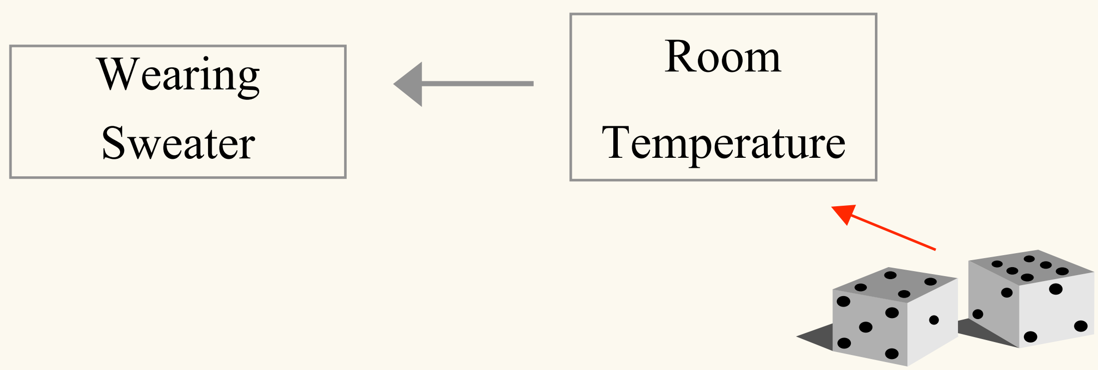
Post experimental System



Modeling Ideal Interventions

Interventions on the Cause

Post -experimental System



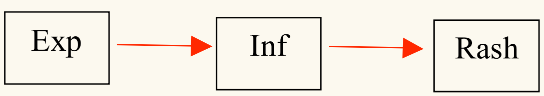
Interventions & Causal Graphs

- Model an ideal intervention by adding an "intervention" variable outside the original system
- Erase all arrows pointing into the variable intervened upon

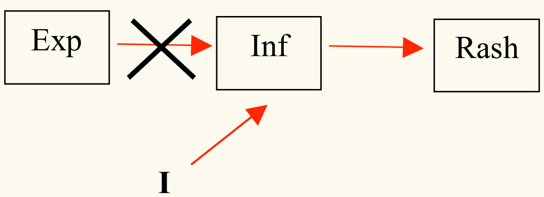
---

Intervene to change Inf

Pre-intervention graph



Post-intervention graph?



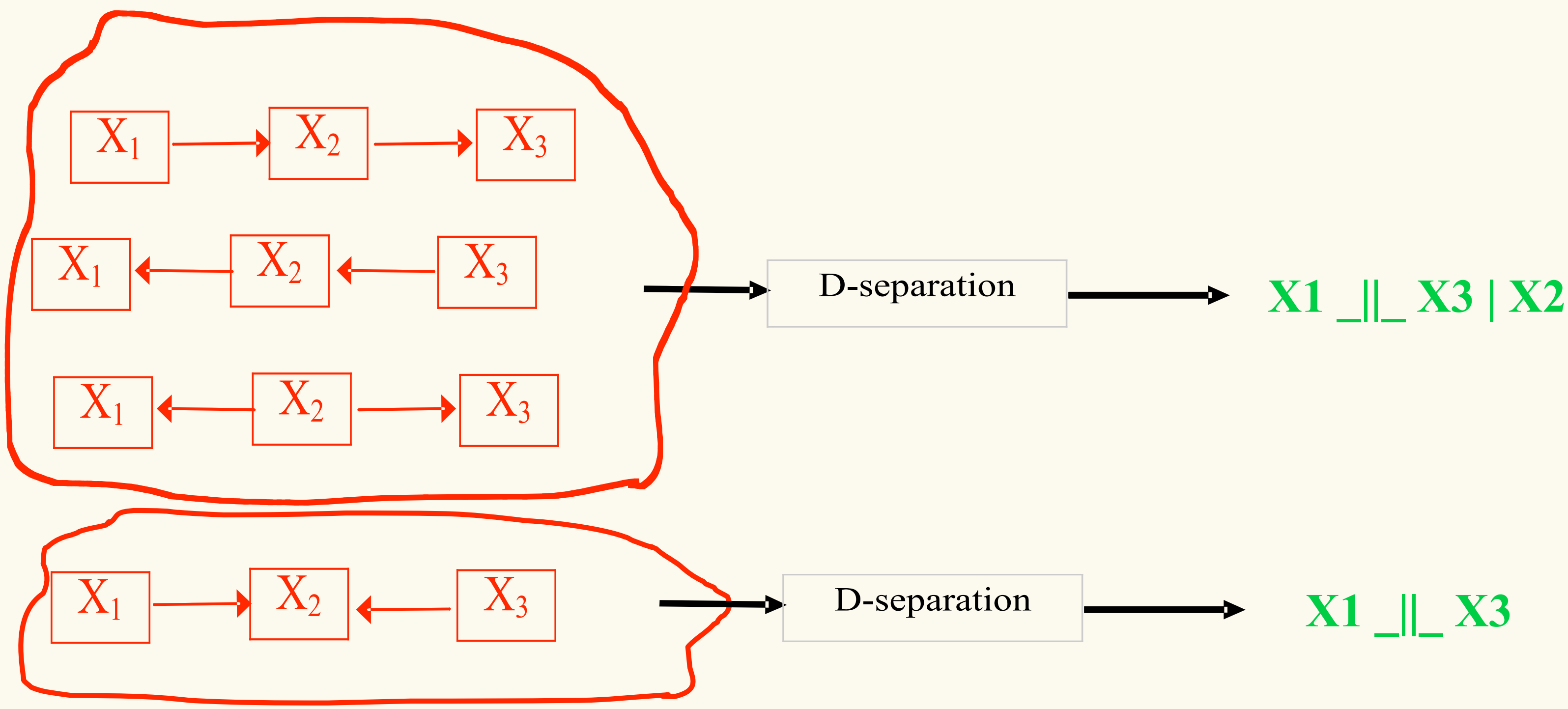
Association and Independence

- Relative Frequency
- Conditional Relative Frequency
- Independence
- Conditional Independence

Causation → Association

D-separation

Equivalence



Association → Causation

Problems for Causal Discovery:

- Underdetermination
- Confounding
- Measurement Error
- Sampling Variability (Statistics!)

Strategies for Causal Discovery

- Experiments (Interventions)
- Statistical Control (multiple regression, etc.)
- Search

Causal and Statistical Reasoning Online

www.phil.cmu.edu/projects/csr

Open Learning Initiative

http://oli.web.cmu.edu

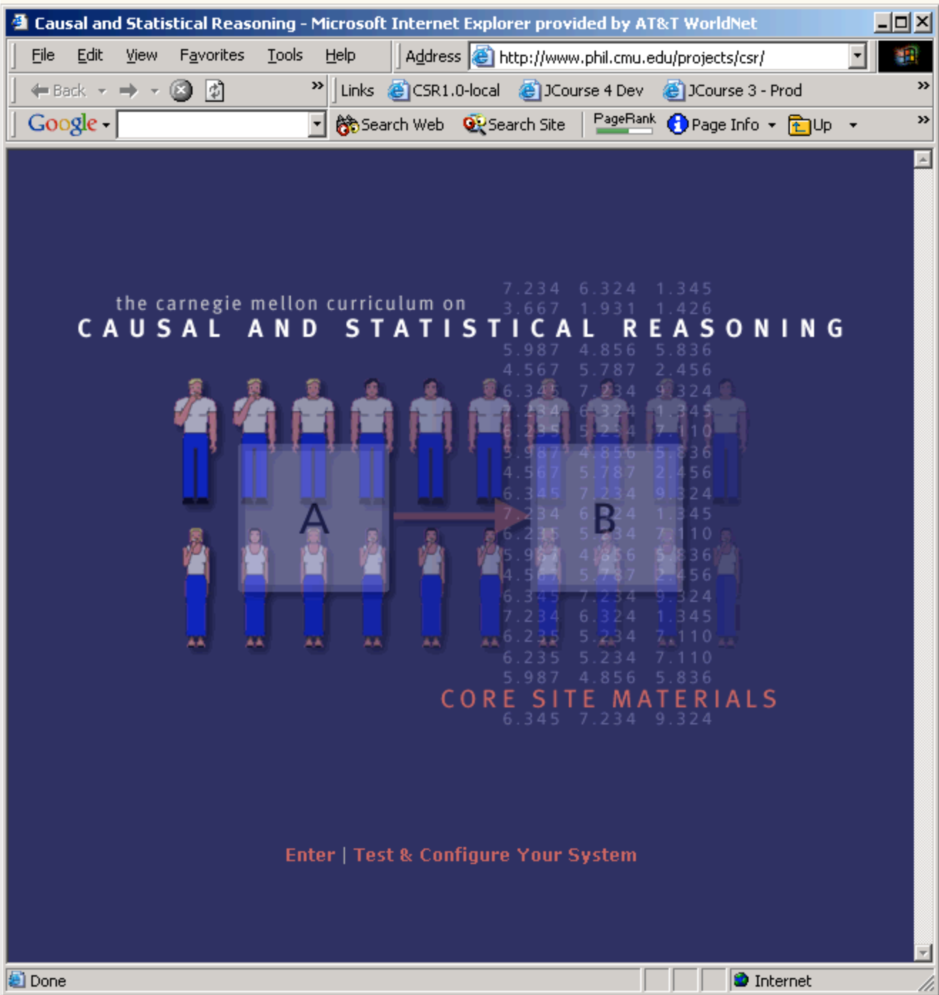
- 16 Content Modules (full semester course)
- > 100 “Case Studies”
- Causality Lab
- Support Materials
  - Recitation Lessons
  - Tests

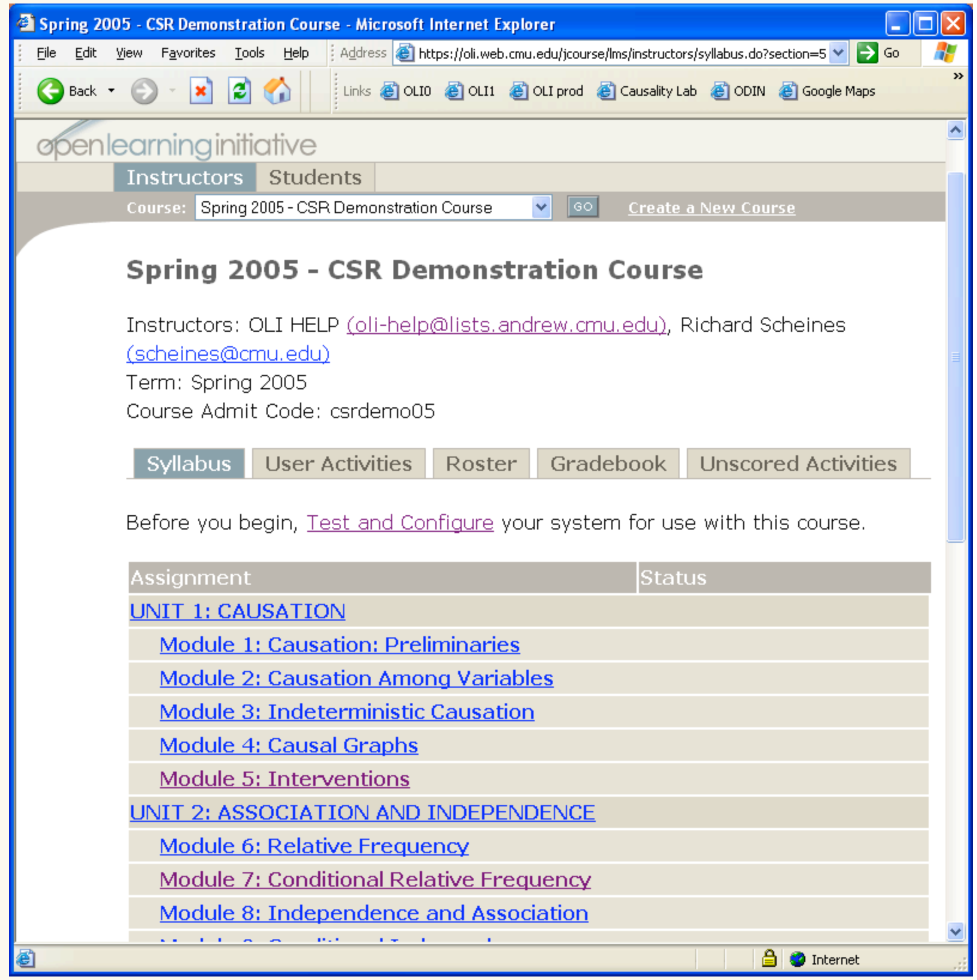
CSR

www.phil.cmu.edu/projects/csr

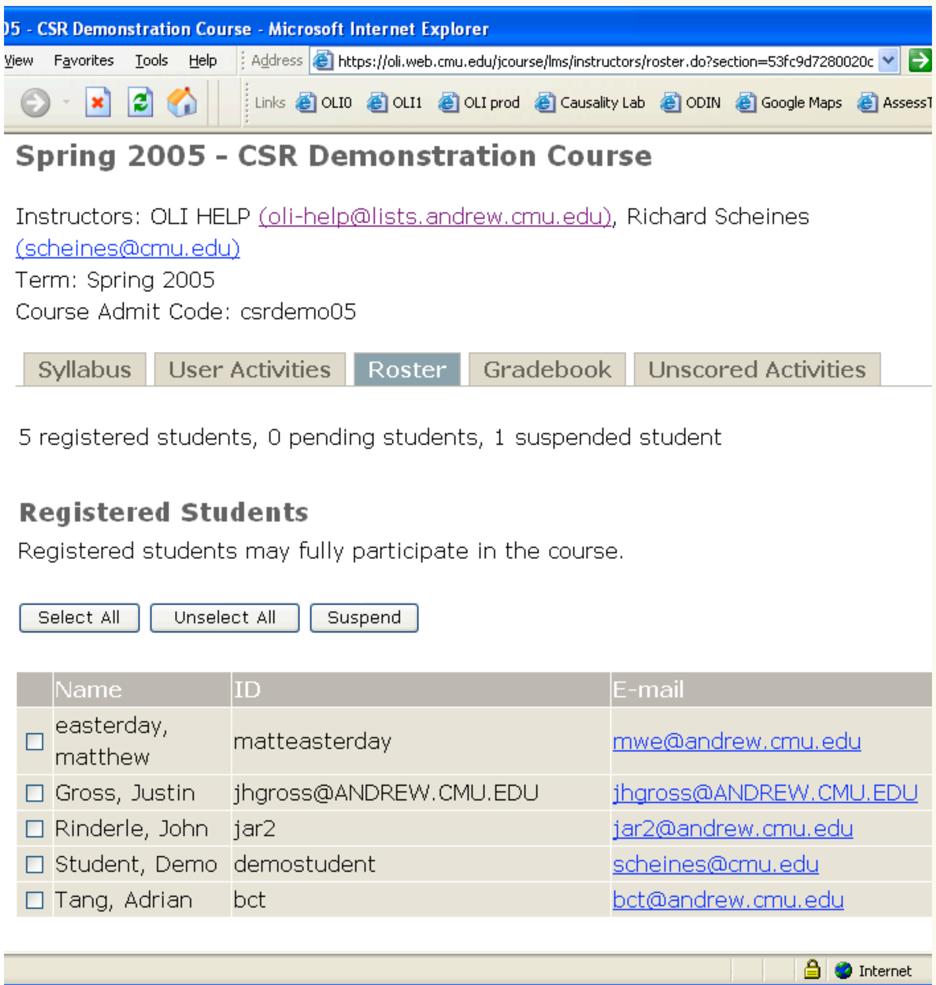
Demo

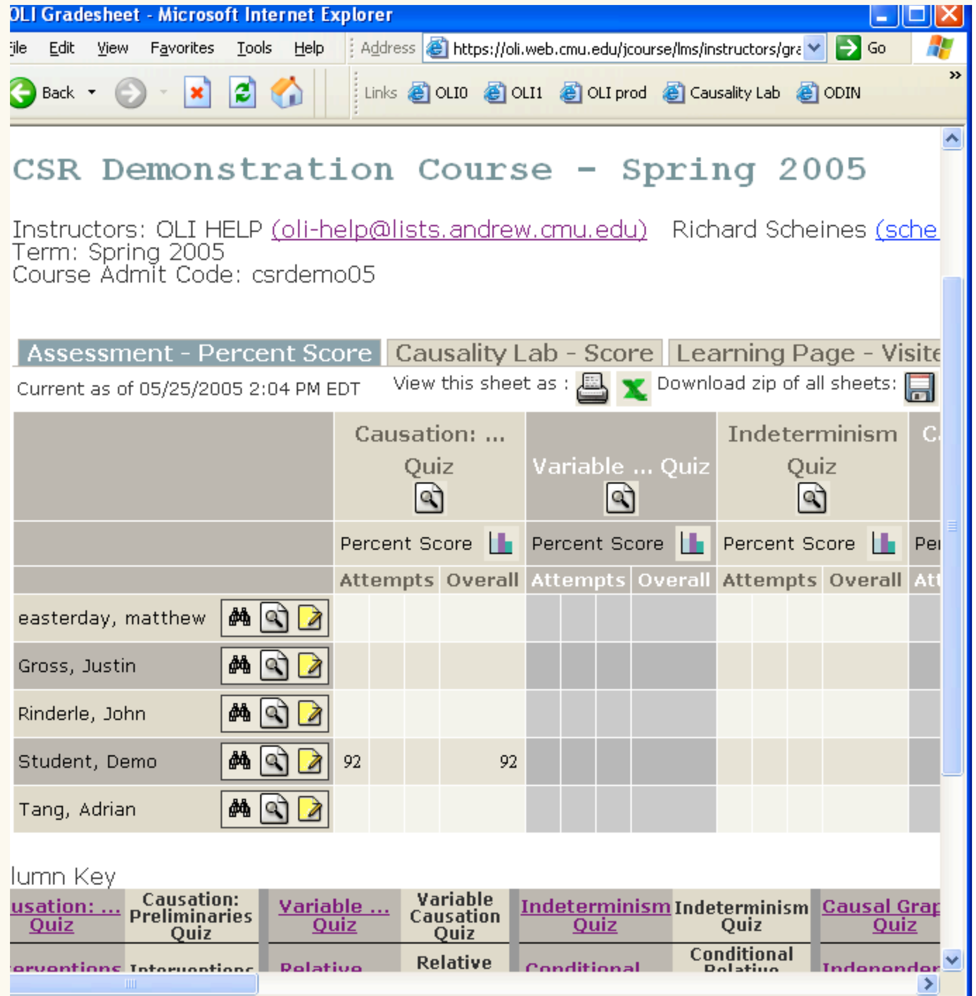
The Online Course





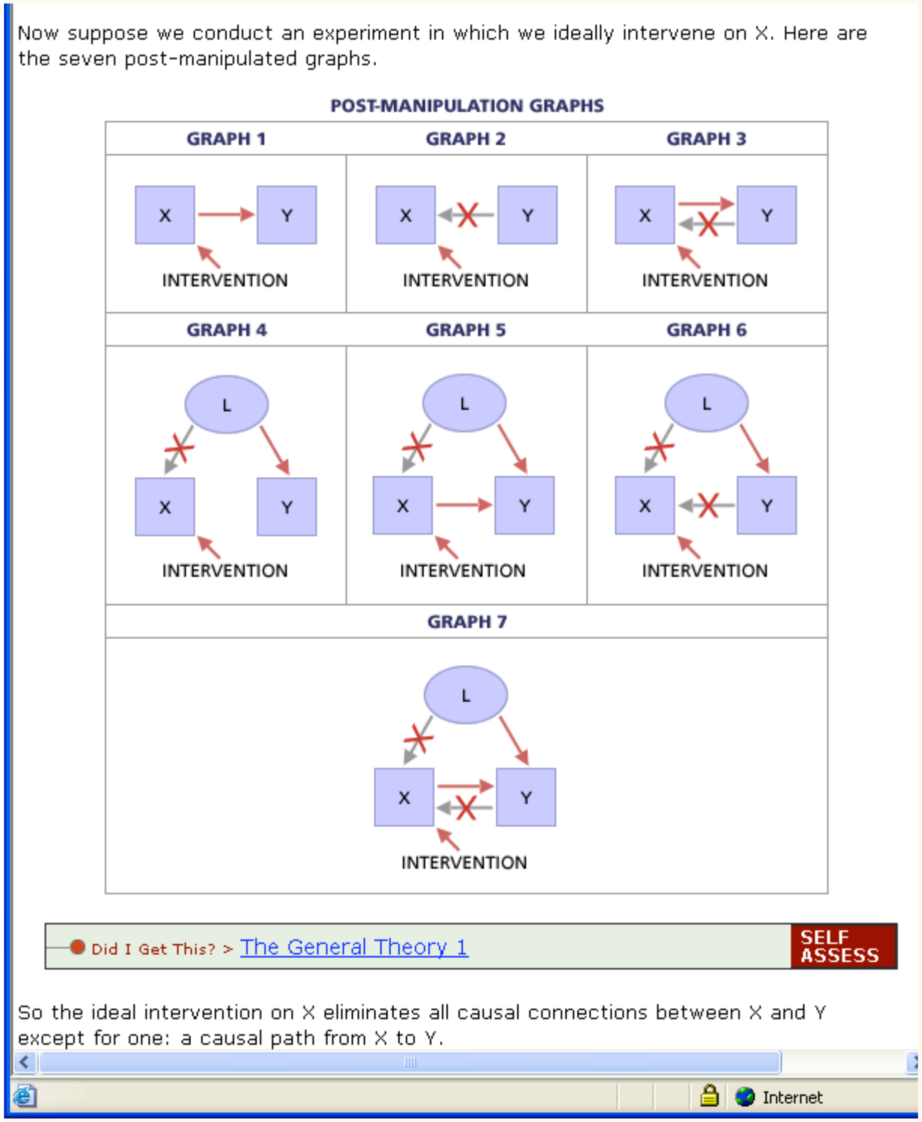
The Online Course:
Roster and Gradebook



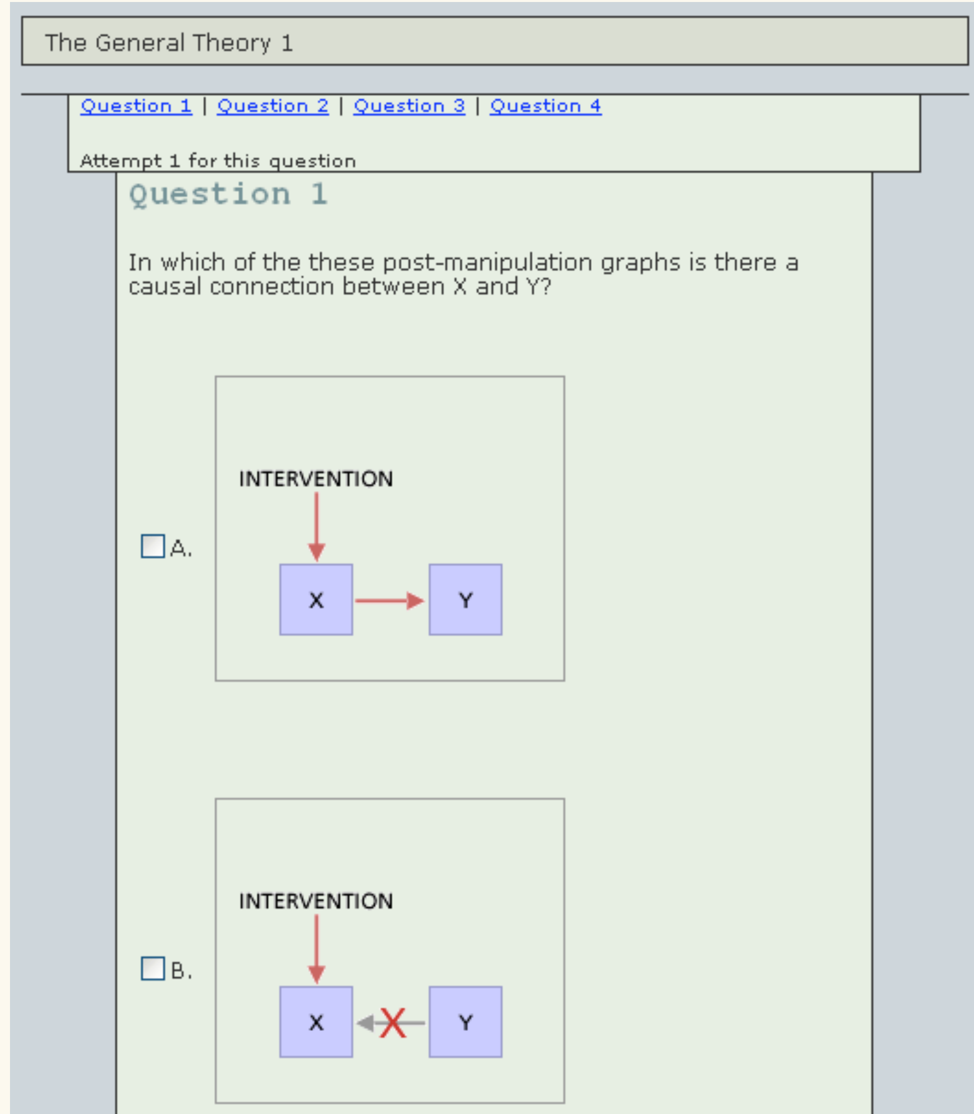


The Online Course: Content Modules

Presenting Concepts



Comprehension Checks



The Online Course: Case Studies

Case Studies

All Case Studies: Alphabetically

- **THE DEADLY AFTER-EFFECT OF ABORTION: BREAST CANCER**
- DOES ACNE DRUG ACCUTANE CAUSE DEPRESSION, SUICIDAL BEHAVIOR?
- After 20 Years, Debate Over Drug Persists
- Acupuncture may stop cocaine cravings
- Acupuncture 'cures dental gagging'
- Possible link between Agent Orange, leukemia
- The Race Card: DOES AN HIV VACCINE WORK DIFFERENTLY IN VARIOUS RACES?
- Alcohol Linked to Breast Cancer
- Alzheimer's first blow
- Aspirin 'could fight prostate cancer'
- Damp homes 'increase asthma risk'
- Asthma chances 'linked to family size'
- Asthma 'linked to obesity'
- Study 'proves' asthma cause
- Prenatal Device Found to Offer Little Help On Early Labor
- Babies' taste 'established in womb'
- Bald Is Bad for Discipline, Says Former Referee
- Baldness pill 'passing early tests'
- Indian State to Target Barbers in AIDS Fight
- Safety: Guarding Basketball Players' Teeth
- Hey, Gorgeous, Here's a Raise! As for you fatties, we're cutting your salaries.
- Higher beer prices 'cut gonorrhoea rates'
- Heart risk link to big families
- T-shirts and shorts a possible cause of global warming
- 'Brain training' link to hunger
- Brazil nut mineral cancer claim
- Study Suggests Breast Cancer Is Linked to Use of Antibiotics

Asthma 'linked to obesity'

Asthma 'linked to obesity'

Date: April 27, 1999

Source : http://news6.thdo.bbc.co.uk/hi/english/health/default.stm

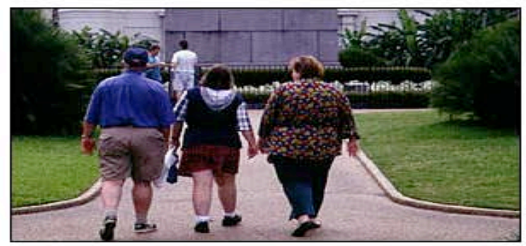
Copyright: 1999 BBC

Concepts

- variables
- causal graphs
- confounders

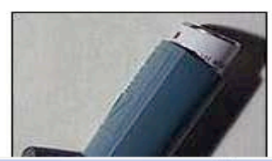
Keywords

- asthma
- obesity

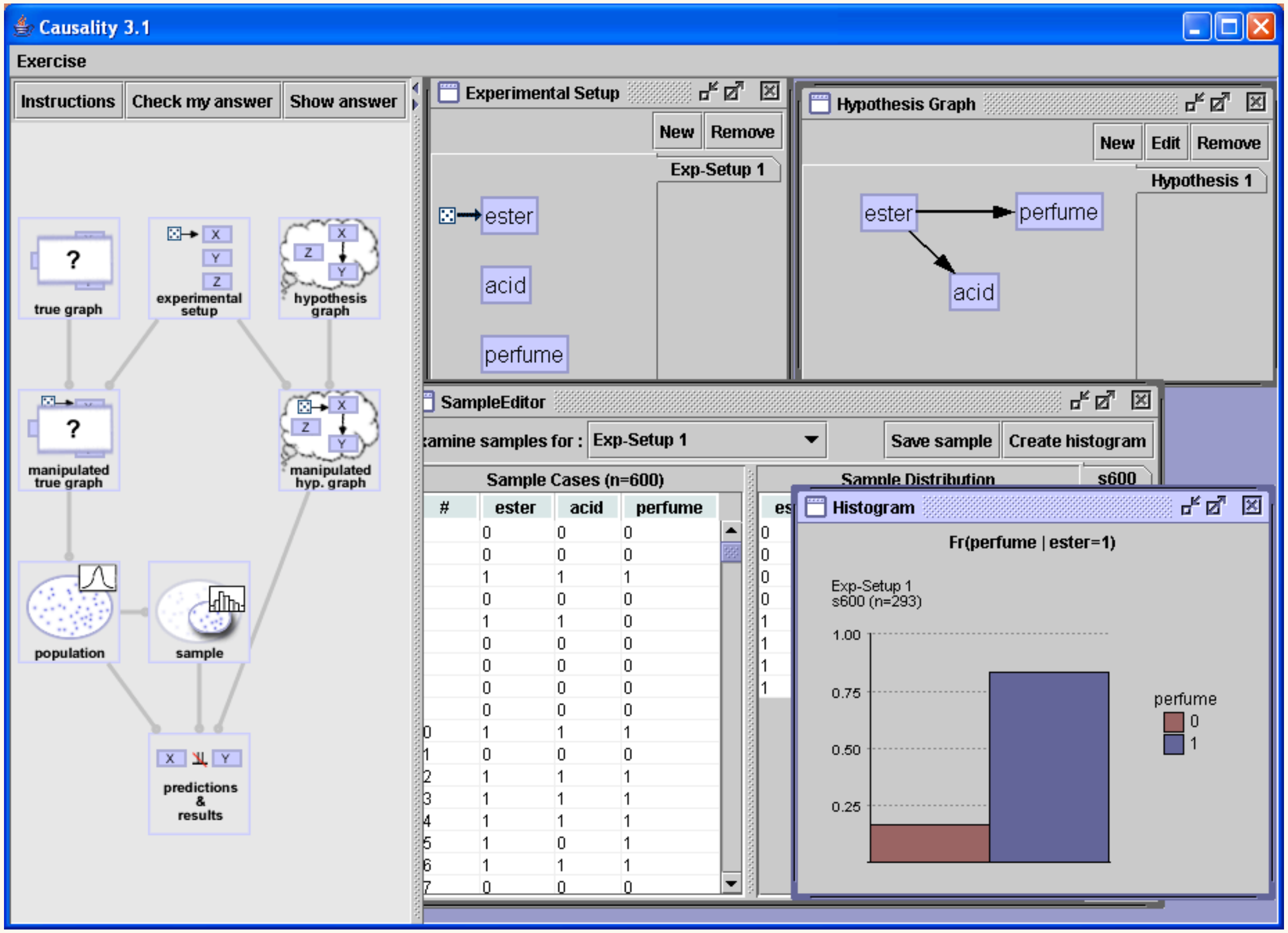


Researchers assessed data from the 1970 British Cohort Study, an on-going study of almost 9,000 people born between 5 and 11 April in 1970 whose health and behaviour have so far been followed up at the ages of 5, 10, 16 and 26 years.

They found that the fatter the adult, the greater the likelihood of asthma.



The Causality Lab



www.phil.cmu.edu/projects/causality-lab

Support Materials

Recitation Lessons, Tests

CSR Usage

- ~ 2,800 total students
- ~ 75 different courses
- ~ 45 Institutions

- Disciplines:
  - Philosophy
  - Statistics
  - Psychology
  - Political Science
  - Math
  - Management
  - Nursing
  - Speech
  - Economics
  - Marketing

CSR Evaluation

- How do students fare with online vs. lecture delivery of identical material?
- What factors affect the pedagogical outcome?

  e.g., face-to-face attendance, time online, exercises attempted, etc.
- What does it cost?

Experiments

2000 : Online vs. Lecture, UCSD

- Winter (N = 180)
- Spring (N = 120)

2001: Online vs. Lecture, Pitt & UCSD

- UCSD - winter (N = 190)
- Pitt (N = 80)
- UCSD - spring (N = 110)

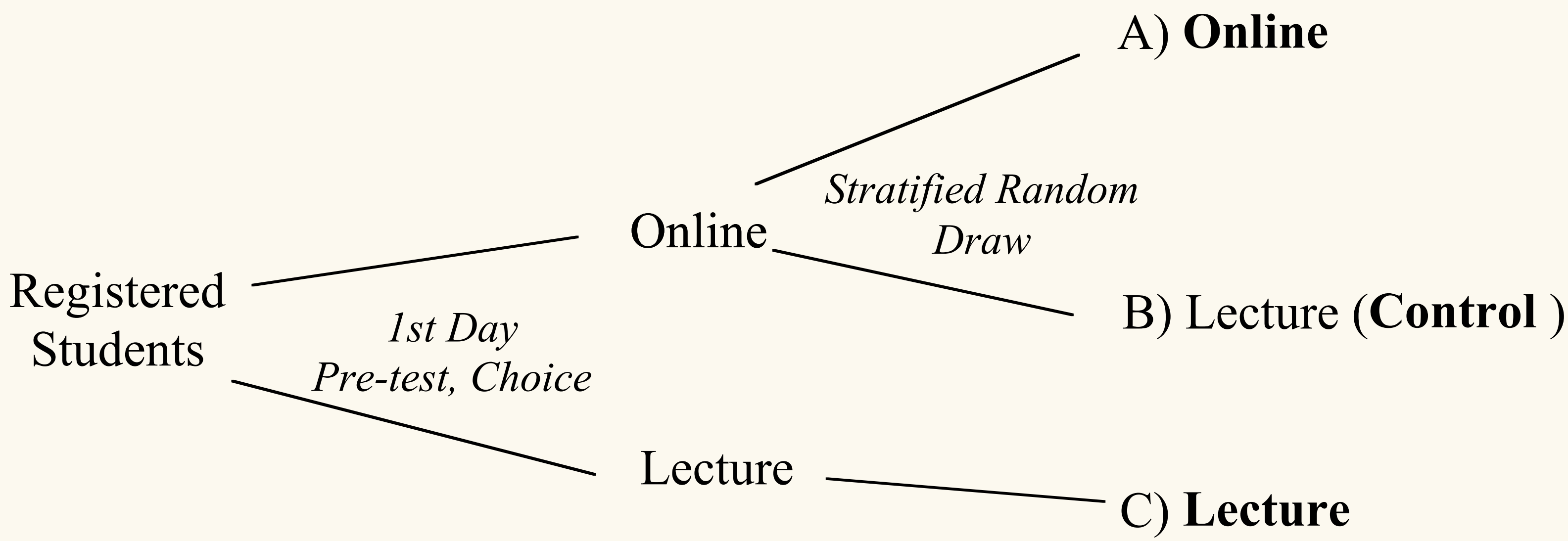
Online vs. Lecture Delivery

- Online:
  - *No* lecture / one recitation per week
  - Required to finish approximately 2 online modules / week

- Lecture:
  - 2 Lectures / one recitation per week
  - Printed out modules as reading – extra assignments

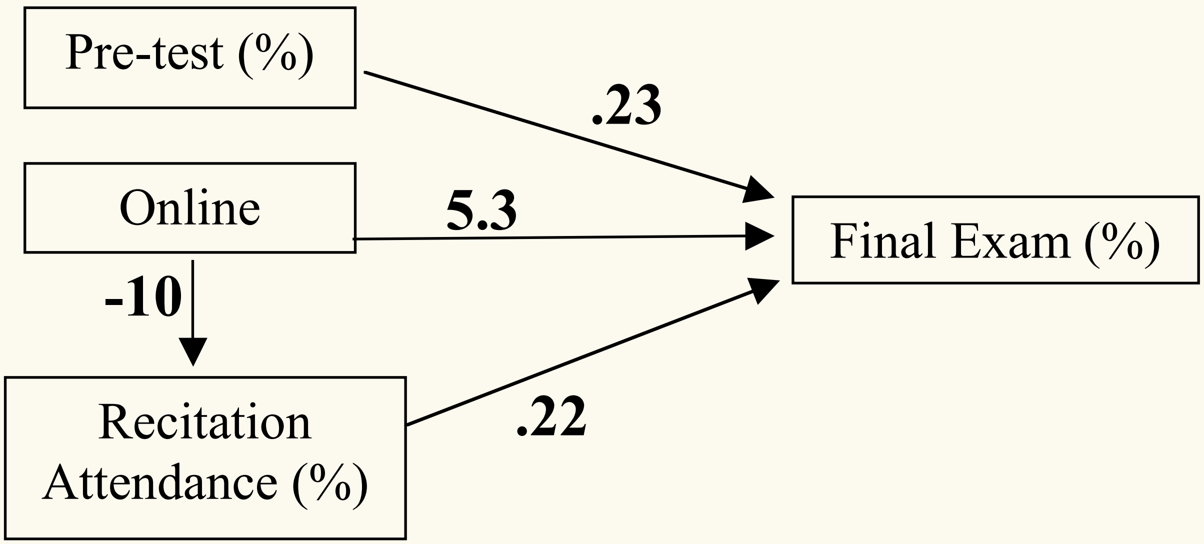
- Same Material, same Exams:
  - 2 Paper and Pencil Midterms
  - 1 Paper and Pencil Final Exam

Experimental Design



- A vs. B -- Main effect
- B vs. C -- Selection Bias

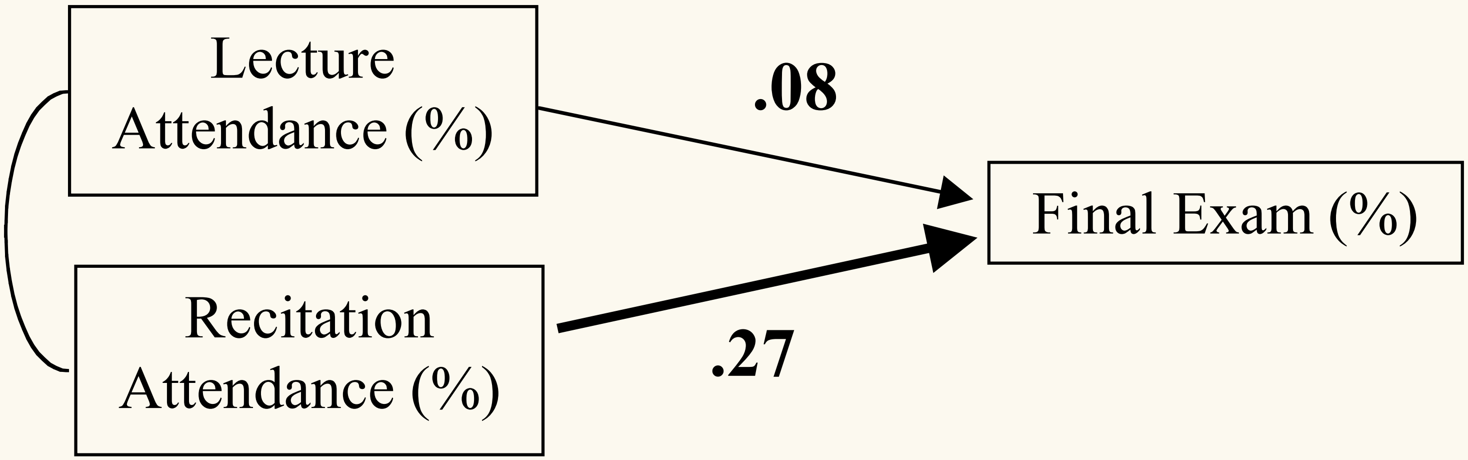
Online vs Lecture



df = 2
$\chi^2 = 0.08$
p-value = .96

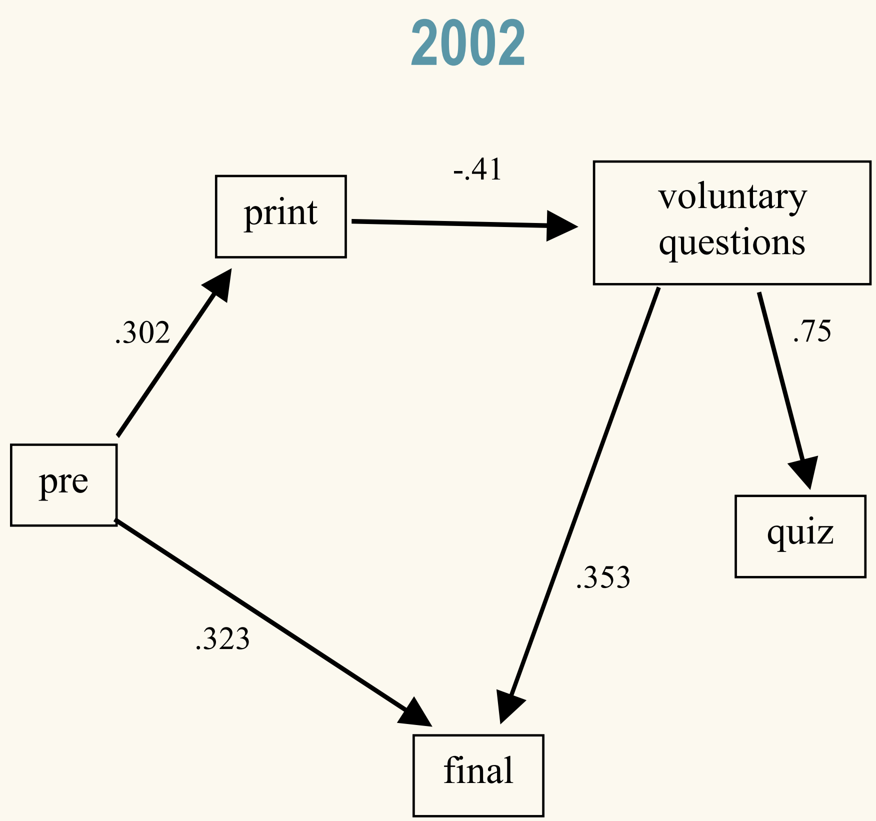
- Online students averaged 1/2 a Stdev better than lecture students (p = .059)
- Factors affecting performance: *Practice Questions Attempted*
- Cost: Online costs 1/3 *less* per student

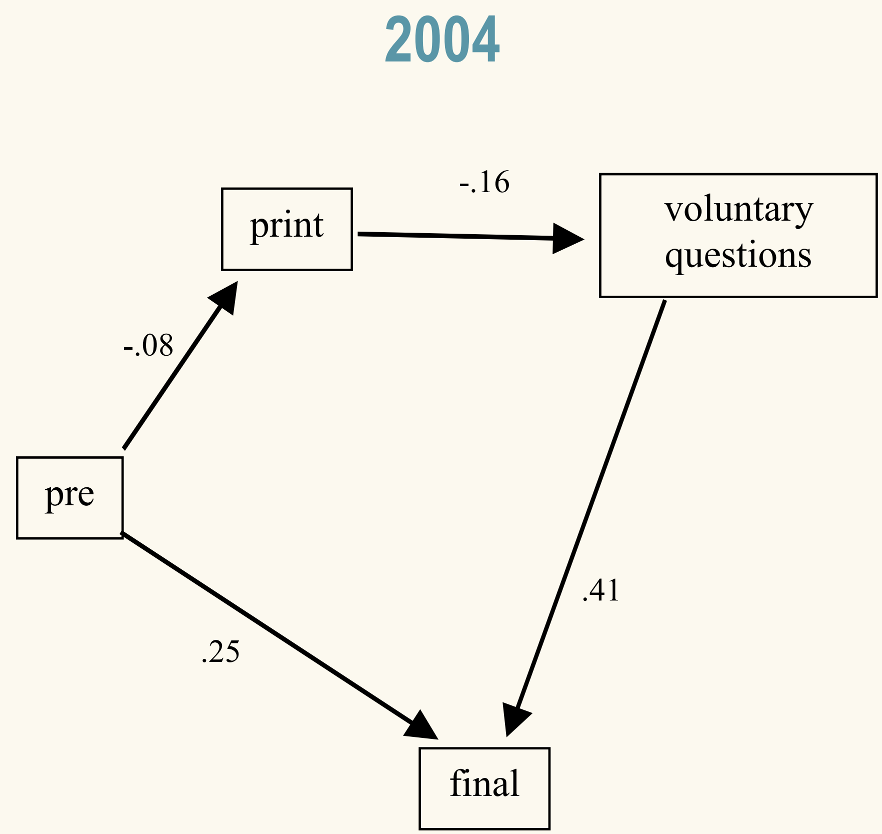
Pitt 2001: Lecture Students Only



Recitation attendance is more important than lecture attendance

Printing and Voluntary Comprehension Checks: 2002 --> 2004





Student Behavior Patterns Relevant to Learning

Time on task:

- time *engaged*
- pattern of work (deadline proximate work only, etc.)

Activities:

- Reading
- Comprehension checks
- Simulations
- Lab exercises

Work Patterns, Help seeking, etc.

Time Spent on Learning Pages

Fall 04:

- 2 classes, 40 Students,
- > 8,000 learning page visits

Filtered (to 6,418):

- 5 seconds < page visits < 600 seconds
- Only pages hit at least 5 times
- Only sessions including at least 5 page visits
- Only students with at least 5 sessions

Variables:

- *Page_demand(i)* : mean time spent over all visits to learning page I
- *Session(j)* : mean time spent on learning pages during session j
- *Student(k):* mean time spent on learning pages by student k

Visit_length on page i during session j by student k =

f(*Page_demand(i), Session(j), Student(k)*, ε)

Time Spent on Page

Visit_length on page i during session j by student k =

$f(\textit{Page\_demand}(i), \textit{Session}(j), \textit{Student}(k), \varepsilon)$

Linear Regression:

Visit_length = .838 *Session* + .837 *Page_demand* + .141 *Student*

R-square = .315

R-square (w/o *Student*) = .314

References

Causal and Statistical Reasoning Online
www.phil.cmu.edu/projects/csr

Open Learning Initiative
http://oli.web.cmu.edu

Spirtes, Glymour, and Scheines, (2000), Causation, Prediction, and Search, 2nd Edition, MIT Press

Scheines, R., Leinhardt, G., Smith, J., and Cho, K. (2005) "Replacing Lecture with Web-Based Course Materials, *Journal of Educational Computing Research*, 32, 1.