In vivo experimentation: An introduction

Robert G.M. Hausmann

National Science Foundation
WHERE DISCOVERIES BEGIN

  July 8, 2008

Outline

- *In vivo* experimentation: **Next**
  - Motivation & definition
- 3 examples
  - Reflection on the 3 examples
- Distinguishing *in vivo* from other types of experiments
- Quiz & discussion
- IV track activities for rest of the week

  July 8, 2008

What is the problem?

- Need external validity
  - Address real instructional problems & content
  - Authentic students (e.g., backgrounds, pre-training)
  - Authentic context (e.g., motivations, social interactions, etc.)
- Need internal validity
  - Control of variables to avoid confounds
    - E.g., instructor effects

  July 8, 2008

Three approaches

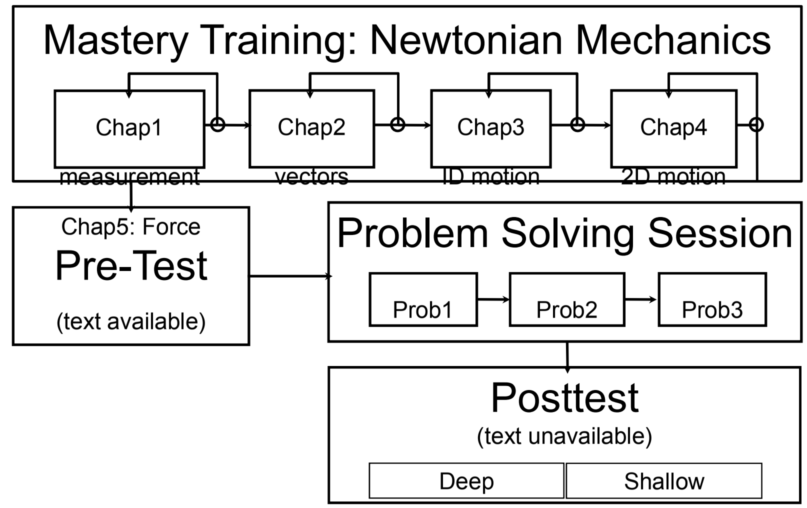
- Traditional
  - Laboratory experiments
  - Classroom experiments
- Novel
  - *In vivo* experimentation

  July 8, 2008

Lab experiments

- Students
  - Volunteers (recruited from classes?)
  - Motivated by money (or credit in psych course)
- Context
  - Instruction done in a lab (empty classroom?)
  - Experimenter or software does the instruction
  - Maximum of 2 hours per session
- Typical design
  - Pre-test, instruction, post-test(s)
  - Conditions differ in only 1 variable/factor
- High internal validity; low external validity

Chi, Roy, & Hausmann (2008)



Classroom experiments

- Participants & context
  - Students from real classes
  - Regular instructors (not experimenter) does teaching
- Design
  - Train instructors to vary their instruction
  - Observe classes to check that manipulation occurred
  - Assess via embedded pre- and post-test(s), or video
- High external validity; low internal validity
  - Weak control of variables

In vivo experimentation

- Students and context
  - In a real classroom with real students, teachers
  - Software controls part of instruction
    - In-class and/or homework exercises
    - Records all interactions (= log data)
- Design
  - Manipulation
    - Software's instruction differs slightly over a long period, or
    - More dramatic difference during one or two lessons
  - Assessment via regular class tests & log data

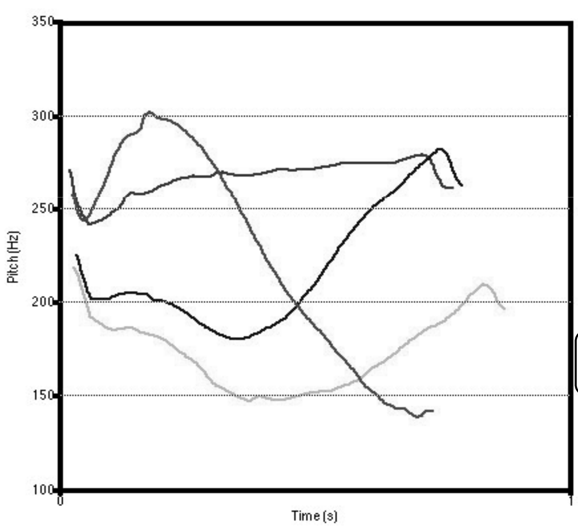
Outline

- In vivo experimentation: Motivation & definition
- **3 examples** **Next**
  - Reflection on the 3 examples
- Distinguishing in vivo from other experiments
- Quiz & discussion
- IV track activities for rest of the week

1st example: Wang, Lui & Perfetti's Chinese tone learning experiment

- Context
  - CMU Chinese course
  - On-line exercises
    - Given spoken syllable, which tone (of 4) did you hear?
  - Very difficult to learn
- Hypothesis
  - Earlier work → subtle wave form differences exist
  - Does displaying them help?

Chinese tones



/ma/ 1: 'mother'
/ma/ 2: 'linen'
/ma/ 3: 'horse'
/ma/ 4: 'scold'

Tone number

Pinyin

Design

- Conditions
  - All conditions select tone from menu
  - All conditions given sound +…
    - Experiment: wave form & Pinyin
    - Control 1: number & Pinyin
    - Control 2: wave form
- Procedure
  - Pre-test
  - One exercise session per week for 8 weeks
  - Several post-test

Preliminary results

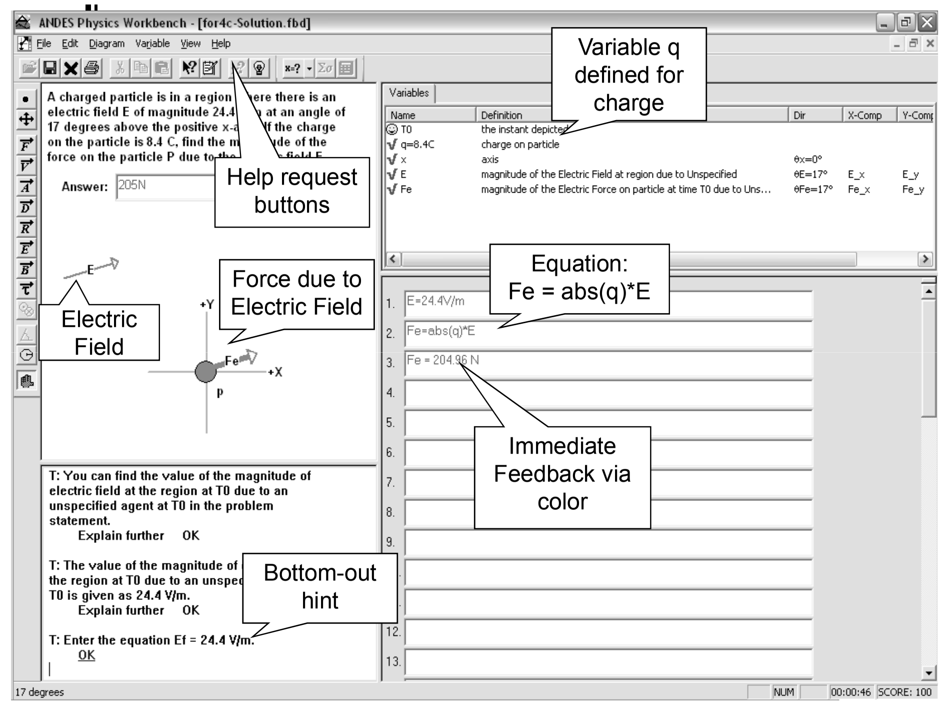
- Error rates during training
  - Experiments < Controls on lessons 2, 5, 6 & 7
- Pre/Post test gains
  - Experiments > Control 1 on some measures
  - Control 2 – too few participants
- Tentative conclusion
  - Displaying waveforms increases learning
  - Second semester data being analyzed
  - Other data being analyzed

Why is this an *in vivo* experiment?

- External validity
  - Real class, student, teachers
  - Post-tests counted in students' grades
    - Cramming?
- Internal validity
  - Varied only two factors: waveform, Pinyin
  - Collected log data throughout the semester
    - Who actually did the exercises?
    - Error rates, error types, latencies
  - Student profiles

Hausmann & VanLehn (2007)

- The **generation** hypothesis: self-explanation > instructional explanation
  - Quick—f___ > Quick—fast (Slameka & Graf, 1978)
  - The **fat** man read about the thin **ice**. (Bransford et al.)
  - How can a worm hide from a bird? (Brown & Kane)
- The **coverage** hypothesis: self-explanation = instructional explanation
  - Path-independence (Klahr & Nigam, 2004)
  - Multiple paths to mastery (Nokes & Ohlsson, 2005)
  - Variations on help (Anderson et al., 1995)



Terminology

- **Example** = problem + multi-entry solution
- **Complete example** = every entry is explained
  - "Because the force due to an electric field is always parallel to the field, we draw Fe at 17 degrees. It's in this direction because the charge is positive. If it had been negative, it would be in the opposite direction, namely 197 degrees."
- **Incomplete example** = no explanations of entries
  - "We draw Fe at 17 degrees."

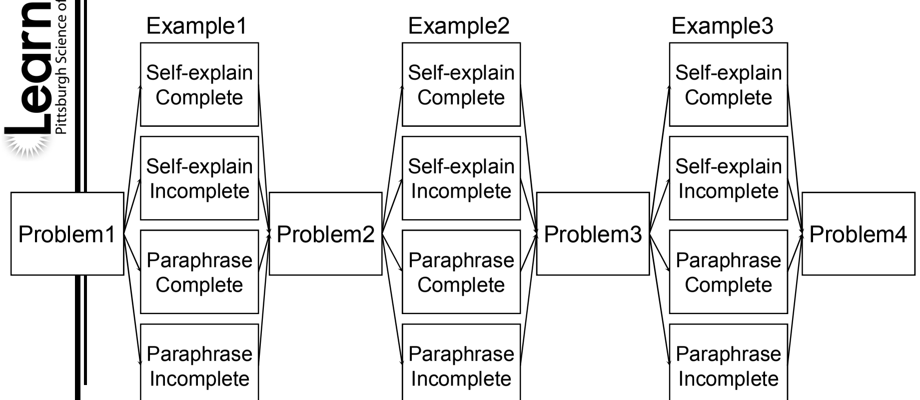
Study design

	Prompted to paraphrase	**Prompted to self-explain**
Incomplete Example (each entry presented without explanation)	No explanation → no learning	Self-explanation → learning
Complete Example (explains each entry)	Instructional explanation → **????**	Self-explanation → learning

Generation hypothesis: No learning

Coverage hypothesis: Learning

Procedure: Each problem serves as a pre-, mid- or post-test

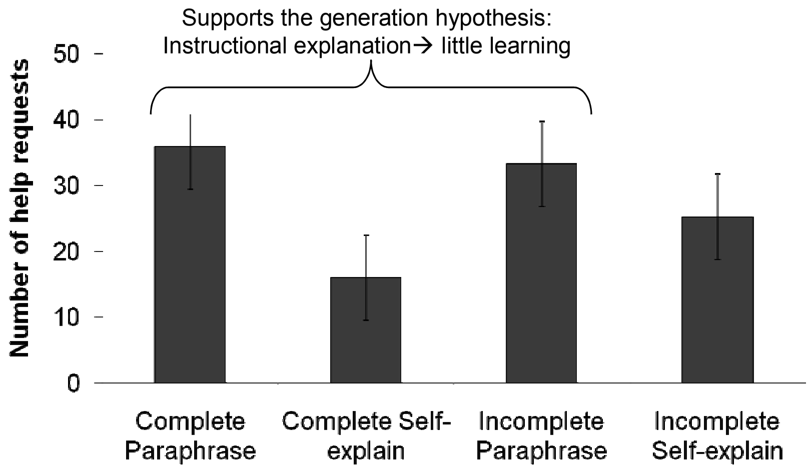


Dependent variables (DVs)

- Log data from problem solving
  - Before, during and after the manipulation
  - Errors
  - Help requests
  - Bottom-out hints
  - Learning curves
- Audio recordings of student's explanations
- Midterm exam

LearnLab
Pittsburgh Science of Learning Center

DV: Help requests



  July 8, 2008

LearnLab
Pittsburgh Science of Learning Center

Butcher & Aleven (2007)

- Scientific Problem
  - Can coordination between and integration of visual and verbal information improve robust learning?
  - Can this integration be supported by scaffolds during tutored practice?
- Hypothesis
  - Interacting and self-explaining with geometry diagrams will:
    - Decrease use shallow problem-solving strategies
    - Support integration of verbal and visual knowledge

  July 8, 2008

LearnLab
Pittsburgh Science of Learning Center

Study design

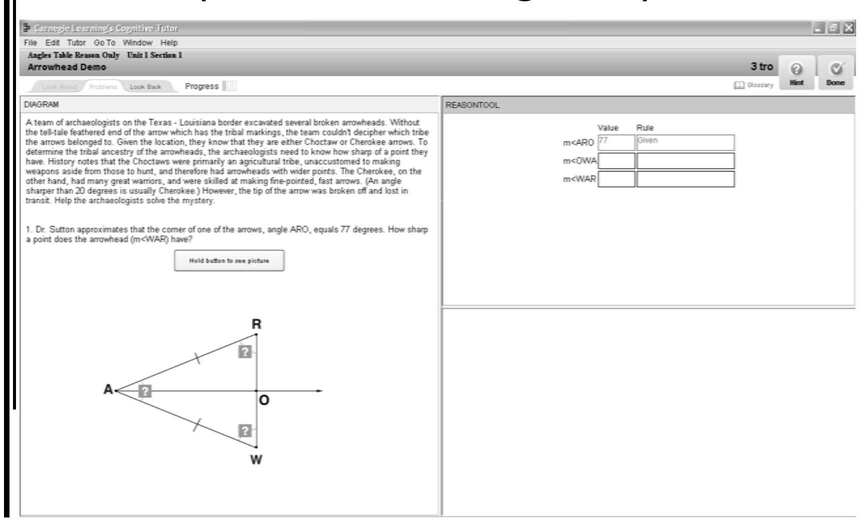
Site of Interaction During Problem Solving

Type of Explanation		
	TABLE (Non-contiguous) GEOMETRY RULE (Verbal Explanation)	DIAGRAM (Contiguous) GEOMETRY RULE (Verbal Explanation)
	TABLE (Non-contiguous) GEOMETRY RULE + APPLICATION (Verbal + Visual Expl.)	DIAGRAM (Contiguous) GEOMETRY RULE + APPLICATION (Verbal + Visual Expl.)

  July 8, 2008

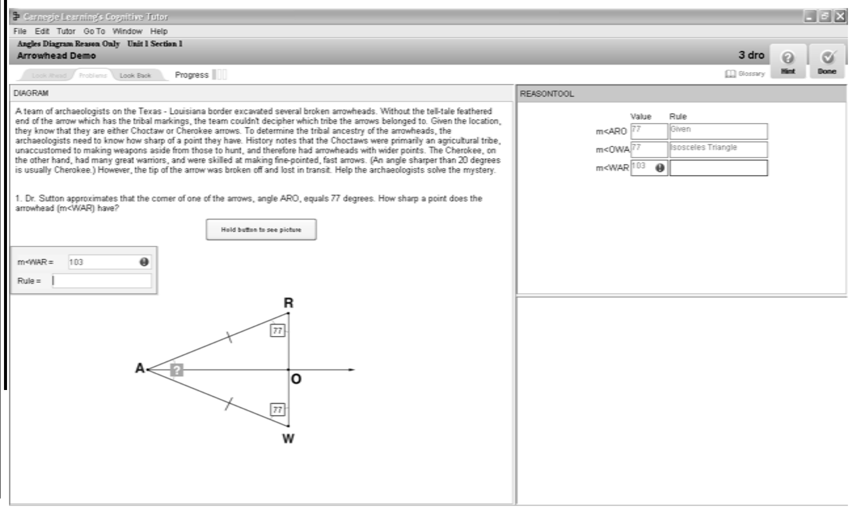
LearnLab
Pittsburgh Science of Learning Center

Variable 1: Site of Interaction (Table v. Diagram)

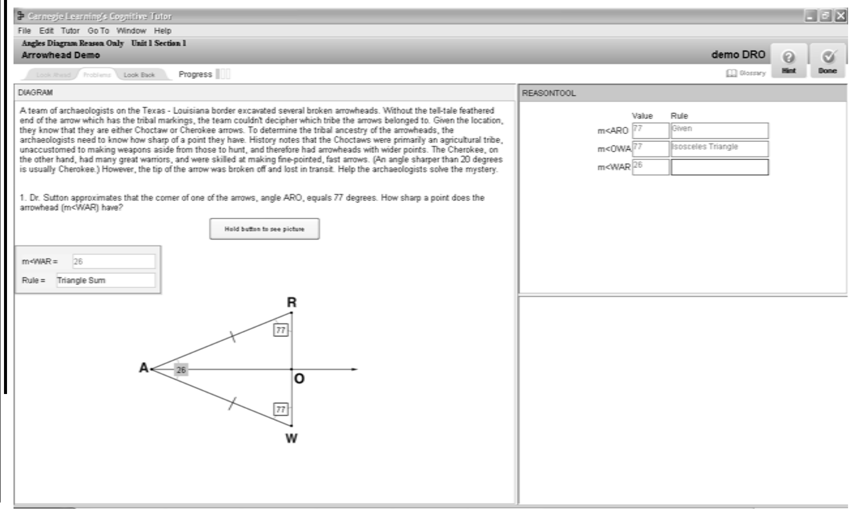


  July 8, 2008

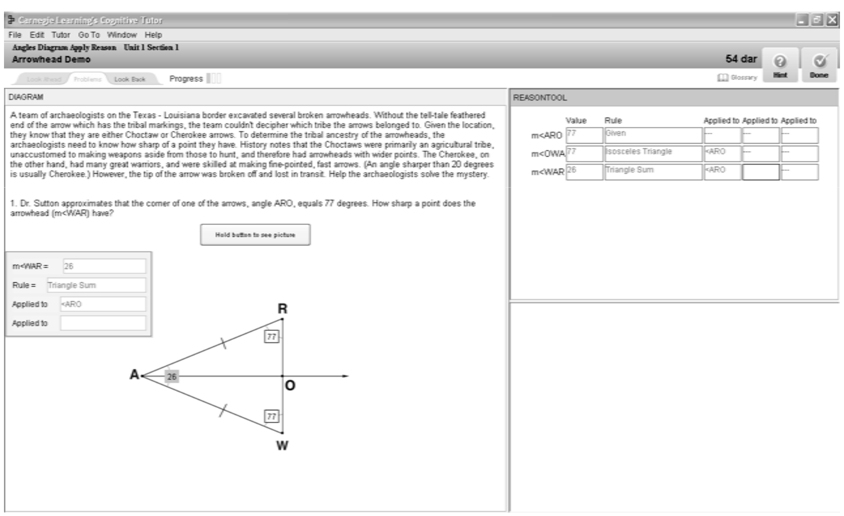
Variable 1: Site of Interaction (Table v. Diagram)



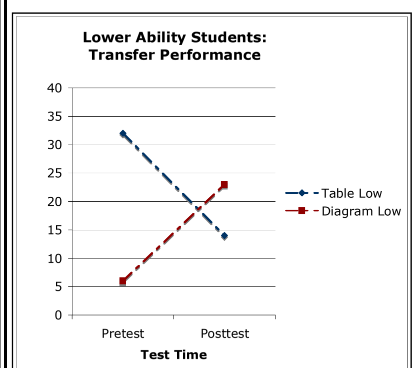
Variable 2: Type of Explanation (Rule vs. Rule + Application)

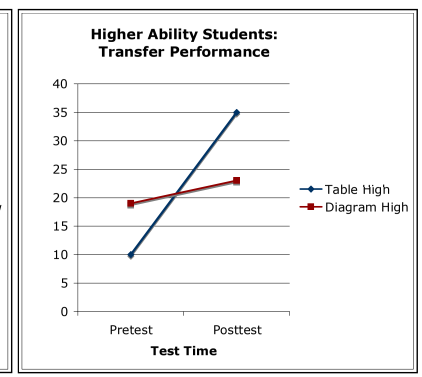


Variable 2: Type of Explanation (Rule vs. Rule + Application)



Results





3-way interaction: Test Time * Condition * Ability: F (1, 38) = 4.3, p < .05

Outline

- In vivo experimentation: Motivation & definition
- **3 examples** — **Next**
  - Reflection on the 3 examples
- Distinguishing in vivo from other experiments
- Quiz & discussion
- IV track activities for rest of the week

Reflection

- Domain
  - Chinese
  - Physics
  - Geometry
- Context
  - Homework
  - Lab work
- Duration
  - Hours
  - Days
  - Weeks

Outline

- In vivo experimentation: Motivation & definition
- 3 examples
  - Reflection on the 3 examples
- Distinguishing in vivo from other experiments — **Next**
- Quiz & discussion
- IV track activities for rest of the week

How does *in vivo* experimentation differ from course development?

- Research problem to be solved
  - Primary: "An open question in the literature on learning is …"
  - Secondary: "One of the hardest things for students to learn in <class> is …"
- Scaling up not necessary
  - One unit of curriculum may suffice
- Sustainability not necessary
  - OK to use experimenter instead of technology

LearnLab Pittsburgh Science of Learning Center

How does *in vivo* experimentation differ from lab experimentation?

- Instructional objectives and content
  - Already taught in course, or
  - Negotiated with instructor
- Control group must receive good instruction
- Logistics
  - Timing – only one opportunity per semester/year
  - Place
- Participation not guaranteed
  - Count toward the student's grade?

LearnLab Pittsburgh Science of Learning Center

How does *in vivo* experimentation differ from classroom experimentation?

- Superficial differences
  - Treatment implemented by training teachers
    - And observing whether they teach as trained
    - Or better!
  - Can only do between-section, not between-student
  - Control groups are often absent or weak
- Underlying difference
  - Granularity of the hypotheses and manipulations
  - See next few slides

LearnLab Pittsburgh Science of Learning Center

An example of a large-grained classroom experiment: PUMP/PAT

- Early version of CL Algebra (Koedinger et al.)
  - Tutoring system (PAT)
  - Curriculum (PUMP) including some teacher training
  - Whole year
- Hypothesis
  - PUMP/PAT is more effective than conventional instruction

LearnLab Pittsburgh Science of Learning Center

A 2nd example of large grained classroom experiments: CECILE

- CECILE (Scardamalia, Bereiter et al.)
  - Networked collaborative learning software
  - Long, complex math activities done in small groups
  - Developed and published on the web
  - Whole year
- Hypothesis
  - CECILE community of learning increases gains

A 3rd example of large grained classroom experiments: *Jasper*

- Anchored instruction (Bransford et al.)
  - "Jasper" video provide a vivid, shared anchor
  - Long, complex math activities tied to anchor
  - Whole year
- Hypothesis:
  - Anchored instruction prevents inert knowledge

Outline

- In vivo experimentation: Motivation & definition
- 3 examples
  - Reflection on the 3 examples
- Distinguishing in vivo from other experiments
- Quiz & discussion **Next**
- IV track activities for rest of the week

How would you classify this classroom experiment?

- Reciprocal teaching (Palinscar & Brown)
  - Small, teacher-led groups
  - Students trained two switch roles with teacher & each other
  - Multiple weeks
- Hypothesis: Reciprocal teaching is more effective than normal small group learning

How would you classify this classroom experiment?

- Andes tutoring system (VanLehn et al.)
  - Homework exercises done on Andes vs. paper
  - Same exercises, textbook, labs, exams, rubrics
  - Whole semester
- Hypothesis:
  - Doing homework problems on Andes is more effective than doing them on paper

LearnLab Pittsburgh Science of Learning Center

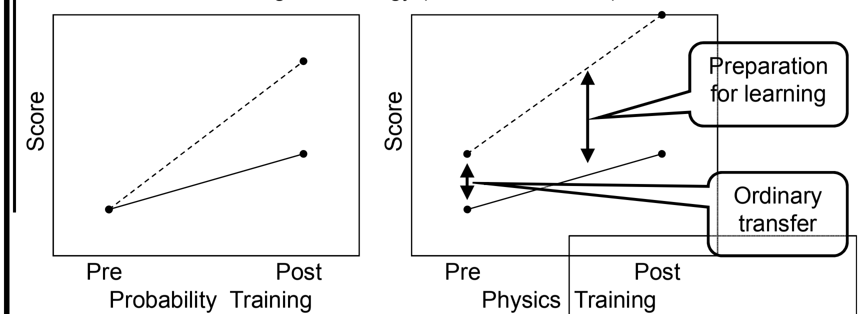
How would you classify this experiment? (Lui, Perfetti, Mitchell et al.)

- Normal drill (used as pretraining)
  - Present Chinese character (visual) and pronunciation (sound)
  - Select English translation. Get applauded or corrected
- Manipulation
  - Select English translation. No feedback.
  - Present character, pronunciation, both or neither
- Co-training hypothesis
  - Drill with both character and pronunciation
    > drill with either character or pronunciation (not both)
    > no extra drill at all
- Pull out

LearnLab Pittsburgh Science of Learning Center

Should this experiment be *redone in vivo*? (Min Chi & VanLehn)

- Design
  - Training on probability then physics
  - During probability only,
    - Half students taught an explicit strategy
    - Half not taught a strategy (normal instruction)



LearnLab Pittsburgh Science of Learning Center

Critique of *in vivo* experimentation

1. Normal instruction for several weeks
   - Including use of Andes for homework
2. Hausmann's study during a 2-hour physics lab period
3. Normal instruction for several more weeks
4. Craig's study, also during a 2-hour lab period
5. Normal instruction for several more weeks

LearnLab Pittsburgh Science of Learning Center

Outline

- In vivo experimentation: Motivation & definition
- 3 examples
  - Reflection on the 3 examples
- Distinguishing in vivo from other experiments
- Quiz & discussion
- IV track activities for rest of the week

Next

LearnLab
Pittsburgh Science of Learning Center

Your job: Simultaneously design 3 elements of an *in vivo* experiment

- A hypothesis
  - Fits into literature on learning
  - High information value (in Shannon's sense)
- A context
  - unit of the curriculum & instructional objective
  - training content and assessments
- A manipulation
  - Tests the hypothesis
  - Fits well in the context

LearnLab
Pittsburgh Science of Learning Center

Schedule

- Tuesday
  - AM: Become familiar with course & tutoring system
  - Early PM: Become familiar with theory
  - Late PM: Start writing Letter of Intent (2 pgs)
    - State background lit, hypothesis, context, manipulation
- Wednesday AM
  - Letter of Intent (LOI) due 10:45 am
  - Feedback from course committee representatives
- Wednesday PM & Thursday
  - Revise design, add details, write proposal & slides
- Friday
  - Presentation

LearnLab
Pittsburgh Science of Learning Center

Contact Information

Robert G.M. Hausmann
University of Pittsburgh
706 Learning Research and Development Center
3939 O' Hara Street
Pittsburgh, PA, 15260-5179

Web page: http://www.pitt.edu/~bobhaus
Email: bobhaus@pitt.edu
Phone: 412.624.7536
Fax: 412.624.9149

