In vivo experimentation: An introduction

Kurt VanLehn

National Science Foundation
WHERE DISCOVERIES BEGIN

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

LearnLab
Pittsburgh Science of Learning Center

What is the problem?

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

LearnLab
Pittsburgh Science of Learning Center

Two most popular experimental methods

- Laboratory experiments
- Classroom experiments

LearnLab
Pittsburgh Science of Learning Center

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

LearnLab
Pittsburgh Science of Learning Center

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

LearnLab
Pittsburgh Science of Learning Center

In vivo experimentation

- Goals
  - Internal validity
  - External validity

LearnLab
Pittsburgh Science of Learning Center

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

LearnLab
Pittsburgh Science of Learning Center

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

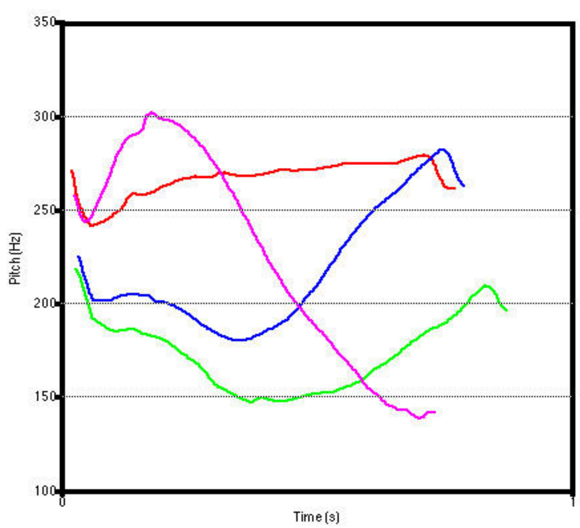
LearnLab Pittsburgh Science of Learning Center

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?

LearnLab Pittsburgh Science of Learning Center

Chinese tones



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

Pinyin

Tone number

LearnLab Pittsburgh Science of Learning Center

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-tests

LearnLab Pittsburgh Science of Learning Center

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

LearnLab Pittsburgh Science of Learning Center

Why is this an in vivo experiment?

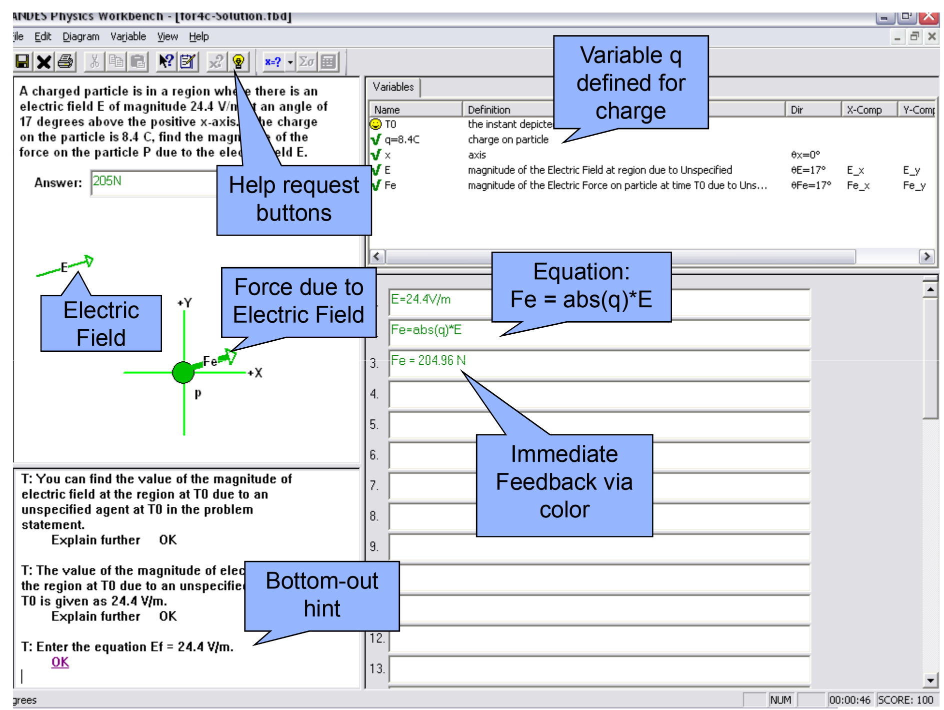
- External validity
  - Real class, student, teachers
  - Post-tests counted in students' grades
    - » Cramming?
- Internal validity
  - Varied only two factors: waveform, Pinyin
  - Collected log data throughout the semester
    - » Who actually did the exercises?
    - » Error rates, error types,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

LearnLab Pittsburgh Science of Learning Center

2nd example: Bob Hausmann's first experiment

- 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)

LearnLab Pittsburgh Science of Learning Center



Terminology

- Example = problem + multi-entry solution
- Complete example = explains every entry
  - "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."

4 conditions

	Prompted to paraphrase	Prompted to self-explain
Incomplete Example (each entry presented without explanation)		
Complete Example (explains each entry)		

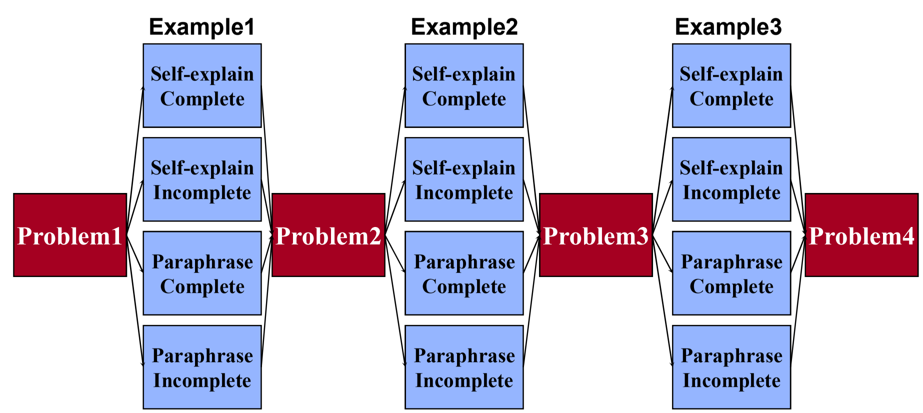
Predictions

	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



In the Physics LearnLab: Spring semester 2006 at the USNA

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

Dependent measures

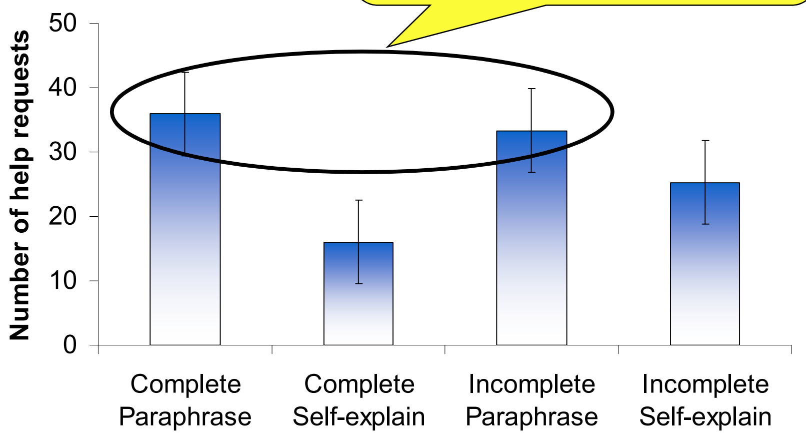
- Log data from problem solving
  - Before, during and after the manipulation
  - Errors
  - Help requests
  - Bottom-out hints
  - Ditto, but main principle only
  - Learning curves
- Audio recordings of student's explanations
- Midterm exam

25 students all talking into headset mikes

LearnLab Pittsburgh Science of Learning Center

One measure: Help requests

Supports the generation hypothesis: Instructional explanation→ little learning



LearnLab Pittsburgh Science of Learning Center

3rd example: Butcher, Aleven et al. geometry study

- Hypothesis
  - Splitting visual attention harms learning.
- Geometry Cognitive Tutor: 2 conditions
  - Entries in the diagram: Keeps attention in diagram
  - Entries in a table: Splits attention

LearnLab Pittsburgh Science of Learning Center

Table Condition splits attention

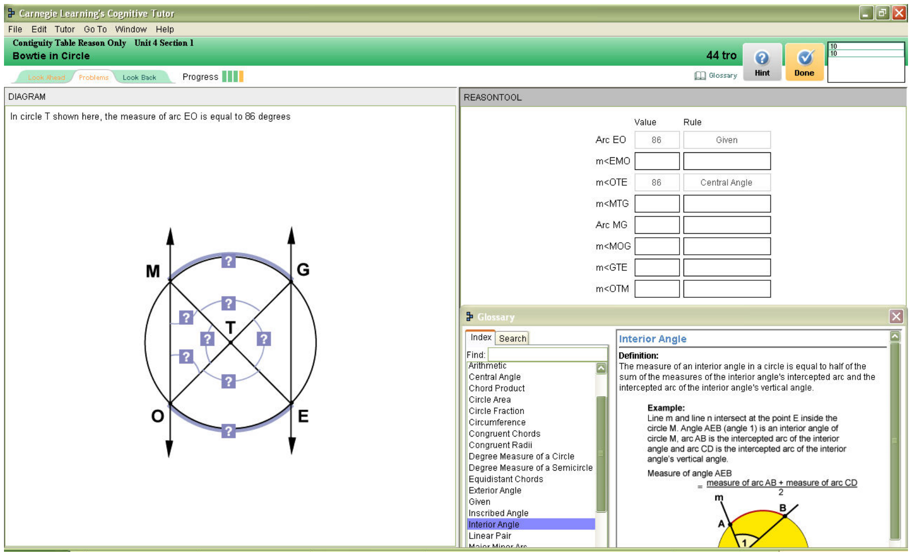
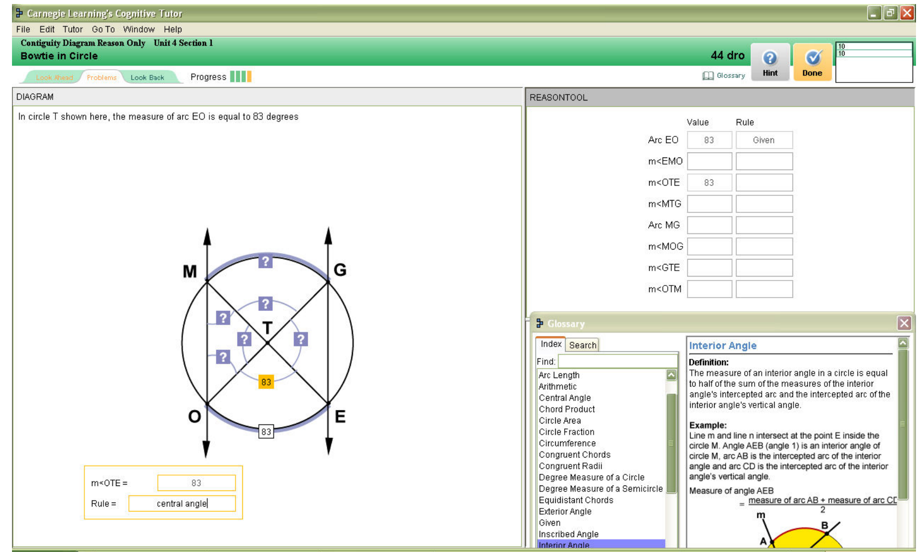
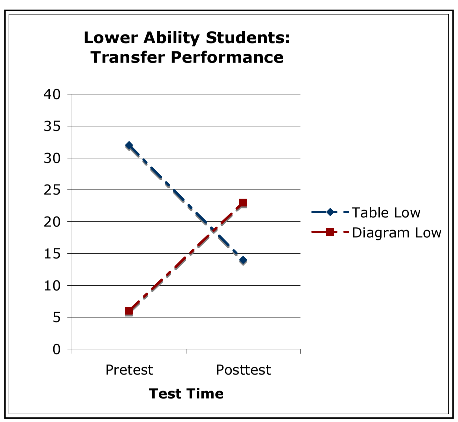
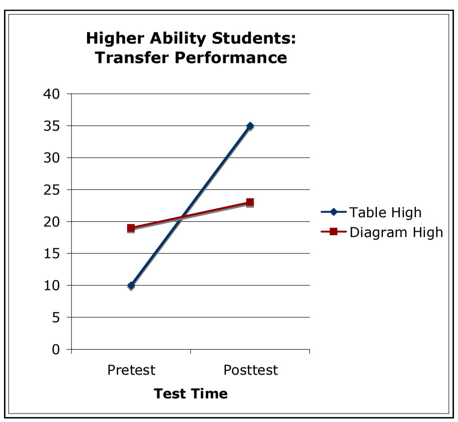


Diagram Condition keeps attention in diagram



Preliminary Results: Transfer





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

LearnLab Pittsburgh Science of Learning Center

Outline

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

LearnLab Pittsburgh Science of Learning Center

Methodological variation: Duration of training

- Wang: Whole semester
- Hausmann: 2 hour lab session
- Butcher: 3 week unit on circles

LearnLab
Pittsburgh Science of Learning Center

Methodological variation: Condition assignment

- Wang: Between sections
  - Different sections get different treatments
  - All students in a section assigned to same treatment
- Hausmann & Butcher: Between subjects
  - Different students assigned to different treatments
  - All sections have all conditions
- Others: Within subjects
  - Same student gets different treatments at different times
  - All students are in all conditions

LearnLab
Pittsburgh Science of Learning Center

Relationship of experimenter's software to course's tutoring system

- Wang's software
  - replaced course's tone-drill software
- Hausmann
  - Did not develop software
  - Used 4 different video tapes, one per condition
  - Experimental activities replaced a physics lab activity
- Butcher's software
  - Variation of Carnegie Learning's tutoring system
  - Designed by Butcher et al.
  - Implemented mostly by Carnegie Learning
  - Replaces course's normal software

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

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 differ from other 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

LearnLab
Pittsburgh Science of Learning Center

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

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** ← Next
- IV track activities for rest of the week

LearnLab
Pittsburgh Science of Learning Center

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

LearnLab
Pittsburgh Science of Learning Center

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

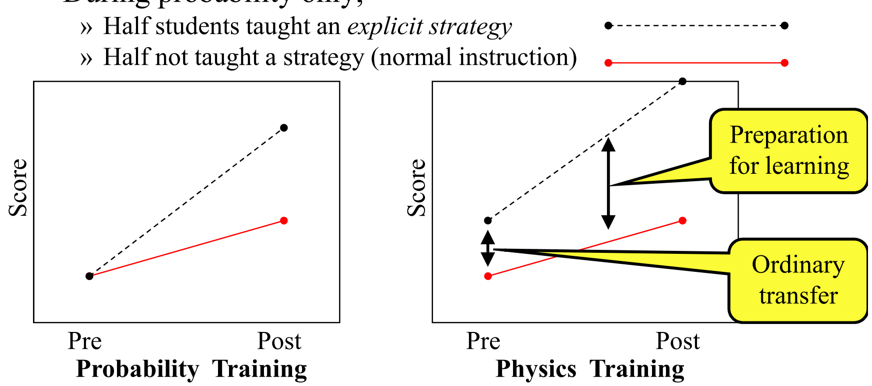
How would you classify this experiment? (Lui, Perfetti, Mitchell et al.)

- 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)



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