Observation, Measurement, and Data Analysis in PER: Methodological Issues and Challenges

David E. Meltzer

Department of Physics and Astronomy

Iowa State University

Ames, Iowa

Supported in part by National Science Foundation grants DUE #9981140 and REC #0206683

Collaborators

Tom Greenbowe
(Department of Chemistry, ISU)

Mani K. Manivannan
(Southwest Missouri State University)

Graduate Students

Jack Dostal (ISU/Montana State)
Ngoc-Loan Nguyen
Tina Fanetti
Larry Engelhardt
Warren Christensen

Outline

- Observing Instruments
- Issues related to assessment of an individual student
  - context dependence of students' ideas
  - multidimensionality of student mental states ("models")
  - time dependence (rate of change) of students' thinking
- Measures of learning gain (*g*, *d*, etc.)
- Issues related to assessment of many students
  - "hidden variables" in students' pre-instruction state
  - sample-selection bias
- Dynamic Assessment (Time-dependent assessment)

Tools of Physics Education Research

- Conceptual surveys or "diagnostics": short-answer or multiple-choice questions emphasizing qualitative understanding, e.g., FCI, MBT, FMCE, CSEM, etc.
- Students' written explanations of their reasoning
- Interviews with students

  e.g. "individual demonstration interviews" (U. Wash.)
- Observations of student group interactions

Observations of Student Group Interactions

- Very time consuming
  - *real-time observation and/or recording*
- Identify more fruitful and less fruitful student group behaviors *e.g. R. Thornton, PERC 2001*
- Characterize student-technology interactions *e.g. V. Otero, PERC 2001; E. George, M. J. Broadstock, and J. Vasquez-Abad, PERC 2001*
- Identify productive instructor interventions *e.g. D. MacIsaac and K. Falconer, 2002*

Caution: Careful probing needed!

- It is ***very easy*** to overestimate students' level of understanding.
- Students ***frequently*** give correct responses based on incorrect reasoning.
- Students' written explanations of their reasoning, and interviews with students, are indispensable diagnostic tools.

Ignoring Students' Explanations Affects both Validity ***and*** Reliability

	Posttest Variant #1 *N = 435*			Posttest Variant #2 *N = 320*	
*[comparison of KE and **p**, two objects different mass, acted upon by same force]*	**Correct answer, *correct* explanation** (**Δt=const.**)	**Correct answer, explanation ignored** (**Δt=const.**)		**Correct answer, *correct* explanation** (**Δx, $\Delta t \neq$const.**)	**Correct answer, explanation ignored** (**Δx, $\Delta t \neq$const.**)
kinetic energy comparison	35%	65%		30%	45%
momentum comparison	50%	80%		45%	55%

T. O'Brien Pride, S. Vokos, and L. C. McDermott, Am. J. Phys. ***66****, 147 (1998)*

Consistent results when explanations taken into account

Context Dependence

- physical context
  - *minor variations in "surface features," e.g., soccer ball instead of golf ball*
- form of question
  - *e.g., free-response or multiple-choice*
- mode of representation
  - *verbal (words), graphs, diagrams, equations*
- physical system
  - *vary physical elements and/or form of interaction*
    e.g., car pushes truck vs. ice-skater collision

Context Dependence of Student Responses

- Changing physical context may significantly alter students' responses
  - E.g., FCI #13, forces on steel ball thrown straight up. When changed to "vertical pistol shot," many who originally included upward force in direction of motion changed to correct response (gravity only). *H. Schecker and J. Gerdes, Zeitschrift für Didaktik der Naturwissenschaften **5**, 75 (1999).*
- Changing form of question may significantly alter students' responses
  - E.g., free-response final-exam problems similar to several FCI posttest questions. In some cases, significant differences in percent correct responses among students who took both tests. *R. Steinberg and M. Sabella, Physics Teacher **35**, 150 (1997).*

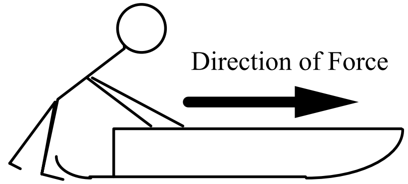
Different Results with Different Representations

Example: Elementary Physics Course at Southeastern Louisiana University.
(*DEM and K. Manivannan, 1998*)

Newton's second-law questions from FMCE.
(nearly identical questions posed in graphical, and "natural language" form.)

	Posttest (N = 18)
"force graph" questions:	56%
"natural language" questions*:	28%

*Force Sled Questions #1-4

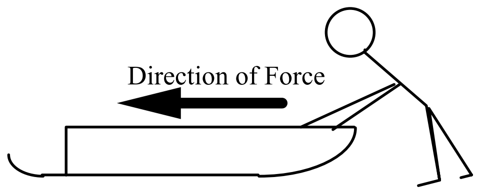


A. The force is toward the **right** and is **increasing** in strength (magnitude).

B. The force is toward the **right** and is of **constant** strength (magnitude).

C. The force is toward the **right** and is **decreasing** in strength (magnitude).

D. No applied force is needed



E. The force is toward the **left** and is **decreasing** in strength (magnitude).

F. The force is toward the **left** and is of **constant** strength (magnitude).

G. The force is toward the **left** and is **increasing** in strength (magnitude).

1. W

2. W

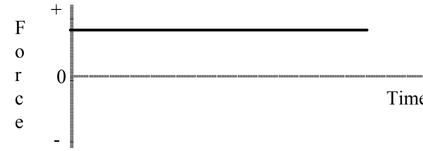
3. Th

4. W
rate (constant acceleration)?

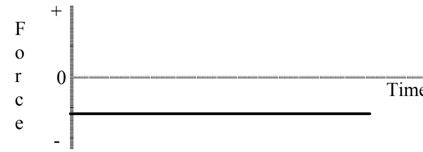
R. Thornton and D. Sokoloff, *Am. J. Phys.* ***66***, *38 (1998)*

Questions 14-21 refer to a toy car which

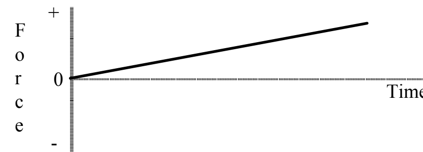
A



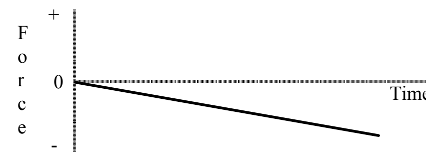
B



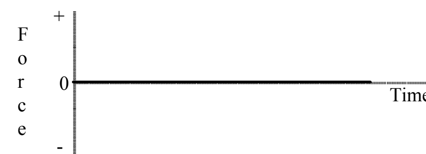
C



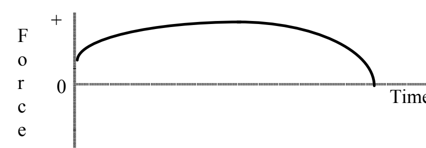
D



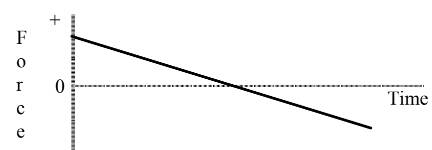
E



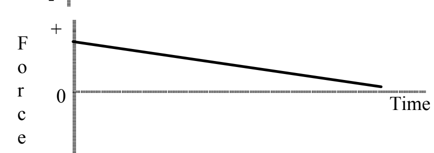
F



G



H



J None of these graphs is correct.

DEM: I need you to explain #3 [Force Sled Question #3].

["The sled is moving to the right. Which force would slow it down at a steady rate (constant acceleration)?"]

Student: [reads answer she chose] ***"The force is toward the left and is decreasing in strength."*** *. . . I was picturing the sled, and I was thinking that it would take less force once it started slowing down . . . I don't know . . .*

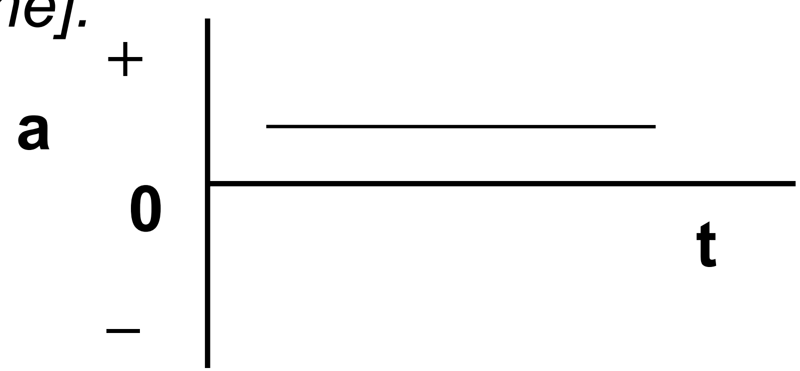
You want it to slow down at a steady rate. So since it's moving towards me and I want it to slow down, I'm actually going to have to go with it . . . and I guess I would ***increase*** *my force to slow it down, not decrease it. I don't know . . .*

DEM: Does the fact that it says "constant acceleration," does that help you to figure this out?

STUDENT: Only in so far as if the acceleration is constant, then the slope is zero . . .

DEM: The slope of what?

STUDENT: The slope of the acceleration, and so the slope of the force is going to be zero: they mirror each other. The force is going to be constant. [Draws graph to explain her reasoning.] *When I think of constant acceleration, I think of this [horizontal line].*



DEM: Now, on this one we've gone all the way around. At first you said less force was needed once it started slowing down, then you said maybe you have to increase the force. And now you're saying, "constant force."

STUDENT: Well, according to what I know, or what I think I know about graphs, I would say that the force had to remain constant because the acceleration is constant.

According to the visual image I have in my head, if a skater was coming towards me and I wanted to slow her down at a steady rate, I don't think that my force would be constant. I don't know ***why*** *I don't think that, I just think it would take less force towards the end.*

Warning

Just because you saw it once does not necessarily mean you'll see it the next time

- Class averages on *full sets* of test items tend to be very stable, one measurement to the next (e.g., different year)
- Measurements on individual test items fluctuate significantly

Example: Algebra-based physics, male students at ISU, FCI #29

Original: forces acting on office chair at rest on floor [no graphic]

Variant ("Gender FCI"; L. McCullough): forces acting on diary at rest on nightstand [drawing of system is shown]

FCI #29	**original version** ***% correct***	variant ***% correct***	significance
Spring 2001	30% (n = 69)	60% (n = 65)	$p = 0.0005$
Fall 2001	40% (n = 55)	37% (n = 46)	*n.s.*

➡ *Replication is important, **especially** for surprising results*

Superposition of Mental States

- Students tend to be inconsistent in applying same concept in different situations, implying existence of "*mixed-model*" mental state. ***E.g., use "impetus" model in one case, Newtonian model on another.*** *I. Halloun and D. Hestenes, Am. J. Phys.* ***53,*** *1058 (1985).*

- Time-dependent changes in *degree of consistency* of students' mental states may correlate with distinct learning patterns with different physical concepts. ***E.g., students learn to recognize presence of normal force, but still believe in "force in direction of motion."*** *L. Bao and E. F. Redish, PERS of AJP* ***69****, S45 (2001).*

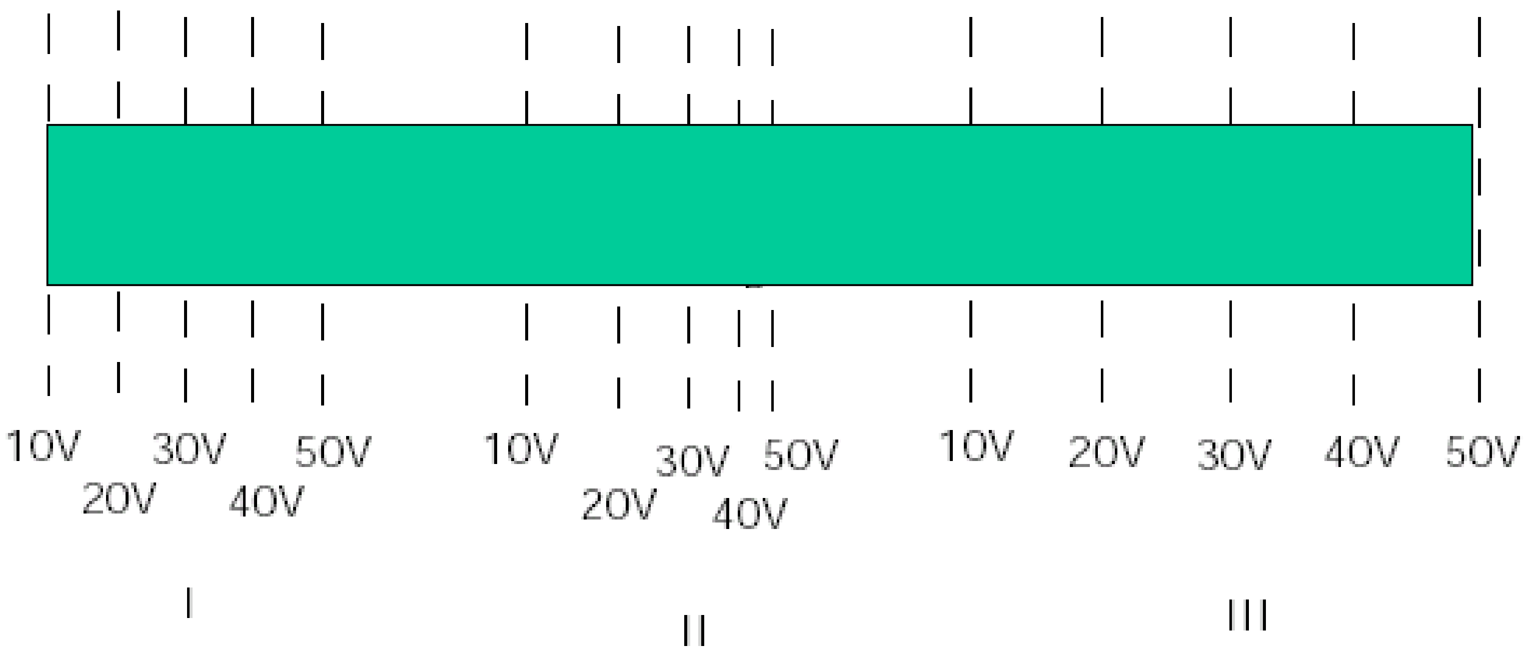
Issues Related to Multiple-Choice Exams

Cf. N. S. Rebello and D. A. Zollman, PERS of AJP (in press)

- Even well-validated multiple-choice questions may miss significant categories of responses.
- Selection of distracters made available to students can significantly affect proportion of correct responses.
- As a result of instruction, new misconceptions may arise that are not matched to original set of distracters.

Deciphering Students' Mental Models from their Exam Responses

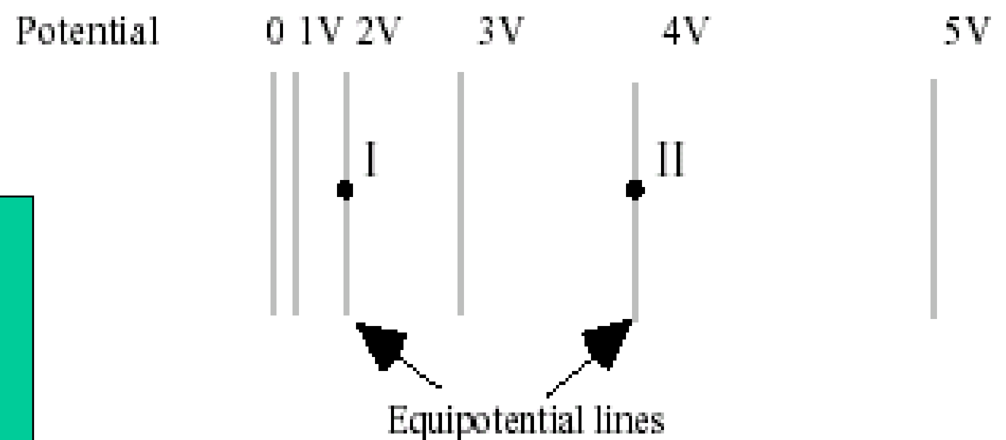
- Distinct patterns of incorrect responses may correlate to transitional mental states. *R. Thornton, ICUPE Proceedings (1997)*
- Varying the selection of answer options can alter the models ascribed to students' thinking. *R. Dufresne, W. Leonard, and W. Gerace, Physics Teacher **40**, 174 (2002).*
- Students' justifications for incorrect responses may change as a result of instruction. *J. Adams and T. Slater (1997)*
- Precision design of questions and answer options necessary for accurate targeting of students' mental models. *Bao and Redish, PERS of AJP **69**, S45 (2001); Bao, Hogg, and Zollman, AJP, **70**, 772 (2002).*



1. How does the magnitude of the electric field at B compare for these three cases?
   (a) I > III > II
   (b) I > II > III
   (c) III > I > II
   (d) II > I > III
   (e) I = II = III

*D. Maloney, T. O'Kuma, C. Hieggelke, and A. Van Heuvelen, PERS of AJP **69**, S12 (2001).*

2. A p [redacted] II in a
reg [redacted] set of
ar [redacted] ectric force
ex [redacted]



	Force at I	Force at II
(a)	[redacted]	[redacted]
(b)	[redacted]	[redacted]
(c)	[redacted]	[redacted]
(d)	[redacted]	[redacted]
(e)	0	0

Hypothetical Student Models on Relation Between Electric Field and Equipotential Lines

- **Model 1 [correct]:** field stronger where lines closer together. *Responses:* #1: D #2: B or D

- **Model 2:** field stronger where lines farther apart

  Responses: #1: C #2: A or C

- **Model 3:** field stronger where potential is higher

  Responses: #1: E #2: A or C

- **Model 4:** Mixed models, all other responses

Evolution of Student Models

Algebra-based physics at ISU (1998-2001)

n = 299	**Pre-test**	**Post-test**	
Model #1	20%	63%	
Model #2	14%	2%	*[disappears]*
Model #3	9%	8%	*[remains]*
Model #4	57%	27%	

Caution: Models much less firm than they may appear

Spring 2002: 116 Students in same course gave answers pre-instruction ***with explanations*** to the two questions.

	n	**explanation consistent with model**
Model #1	15	5 *(33%)*
Model #2	19	2 *(11%)*
Model #3	21	7 *(33%)*
Model #4	61	—

Patterns of student thinking that seemed to be present on pretest may actually have been largely random.

Interview Evidence of Students' Mental State-Function

- Initially, students may offer largely formulaic responses *e.g., equations, verbatim repetition of phrases, etc.*
- Later responses may contradict earlier ones; sometimes resolvable by student, sometimes not. *Sometimes they have **no** well-defined concept.*
- Even with minimum-intensity probing, students may in time succeed in solving problem that was initially intractable.

➢ *If student "learns" during interview, have we measured "knowledge" or "learning ability"?*

Time Dependence of Student Learning

- Multi-dimensionality of student mental states (i.e., diversity of individual model states) suggests possible correlations with diverse learning trajectories and learning rates.
- Can initial learning rate be correlated with final learning gains? Ambiguous results so far. *(DEM, 1997)*
- To date there has been little focus on assessing physics students' learning rates.

Measures of Learning Gain

- Single exam measures only instantaneous knowledge state, but instructors are interested in improving *learning*, i.e., transitions between states.
- Need a measure of learning gain that has maximum dependence on *instruction*, and minimum dependence on students' *pre-instruction* state.

$\Longrightarrow$ *search for measure that is correlated with instructional activities, but has minimum correlation with pretest scores.*

Normalized Learning Gain "*g*"

*R. R. Hake, Am. J. Phys. **66**, 64 (1998)*

$$g \equiv \frac{posttest\ score - pretest\ score}{maximum\ score - pretest\ score} = \frac{gain}{maximum\ possible\ gain}$$

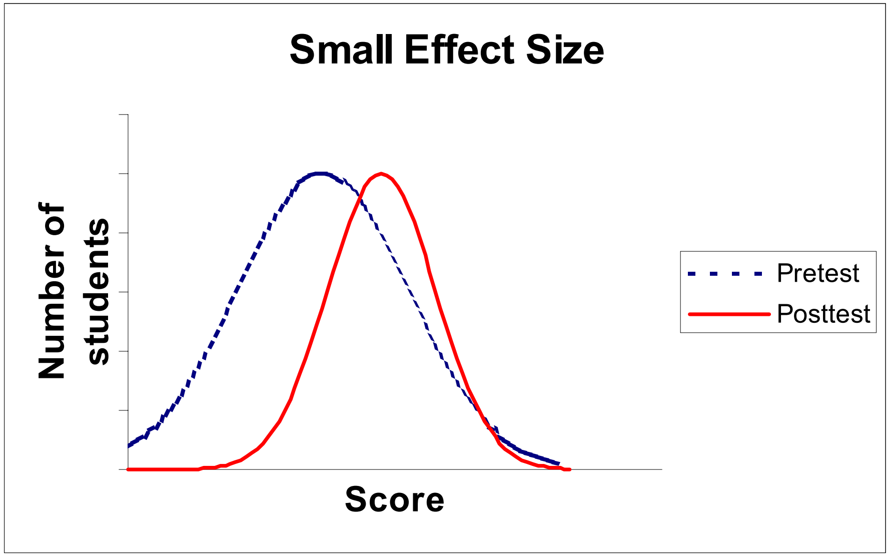
In a study of 62 mechanics courses enrolling over 6500 students, Hake found that mean normalized gain <*g*> on the FCI is:

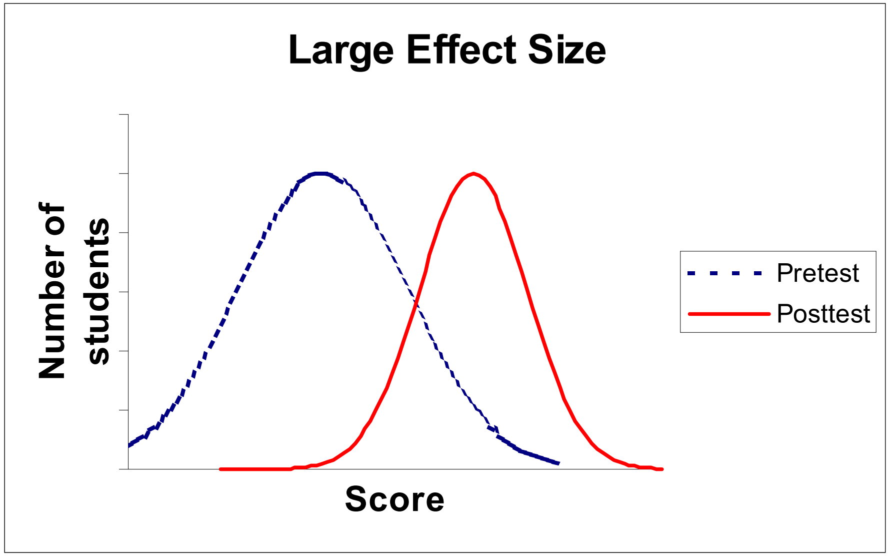
- virtually independent of class mean pretest score (r = +0.02);
- = 0.23±0.04(s.d.) for traditional instruction, nearly independent of instructor;
- =0.48±0.14(s.d.) for courses employing "interactive engagement" active-learning instruction.

➡ *These findings have been largely confirmed in hundreds of physics courses worldwide*

Effect Size *d*: Measure of Non-Overlap

$$d = \frac{|mean\ \ posttest\ \ score - mean\ \ pretest\ \ score|}{|mean\ \ square\ \ standard\ \ deviation|}$$





pretest

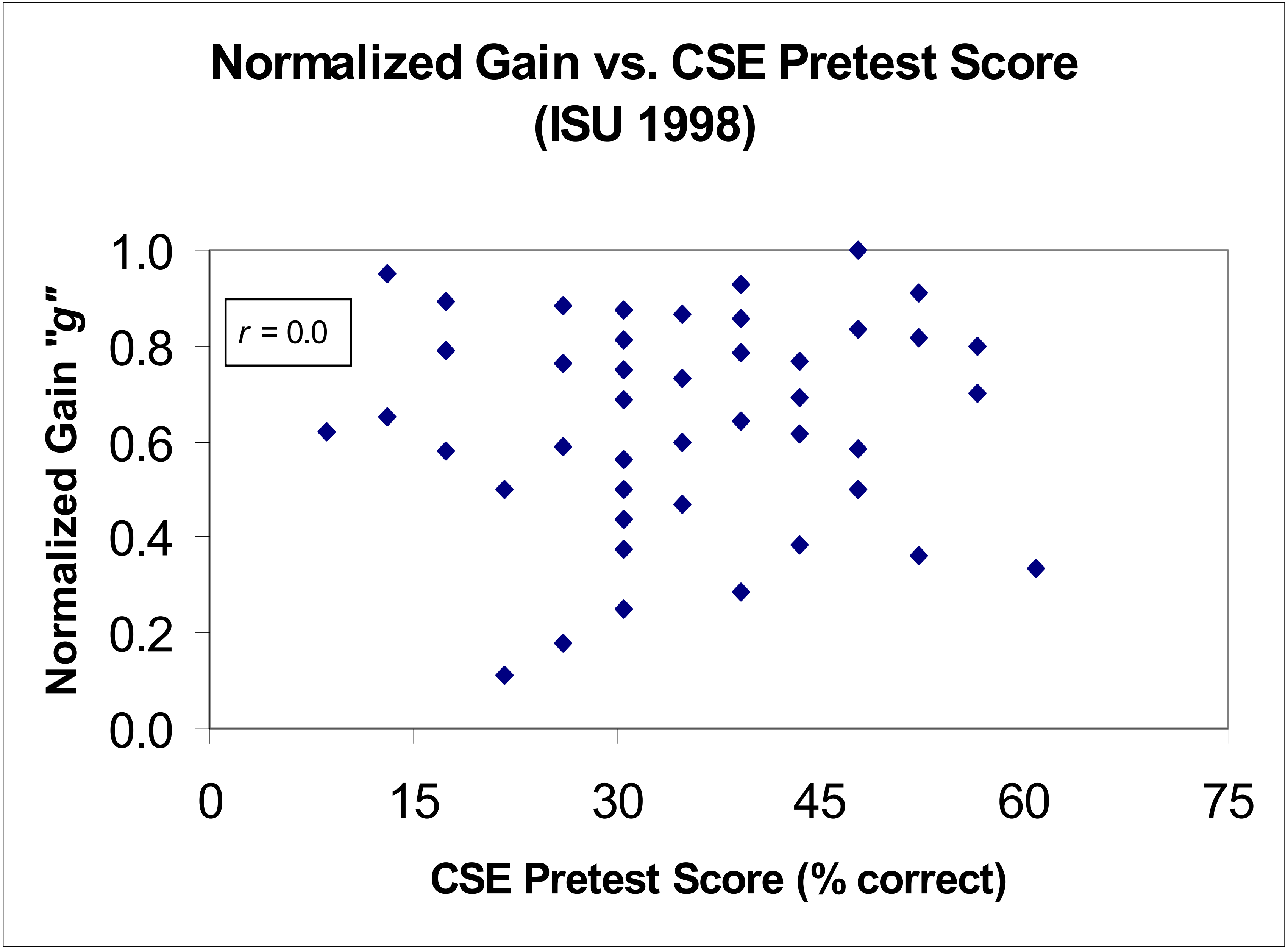
posttest

*Large effect size does **not** necessarily imply "significant" gain!*

But is *g* really independent of pre-instruction state?

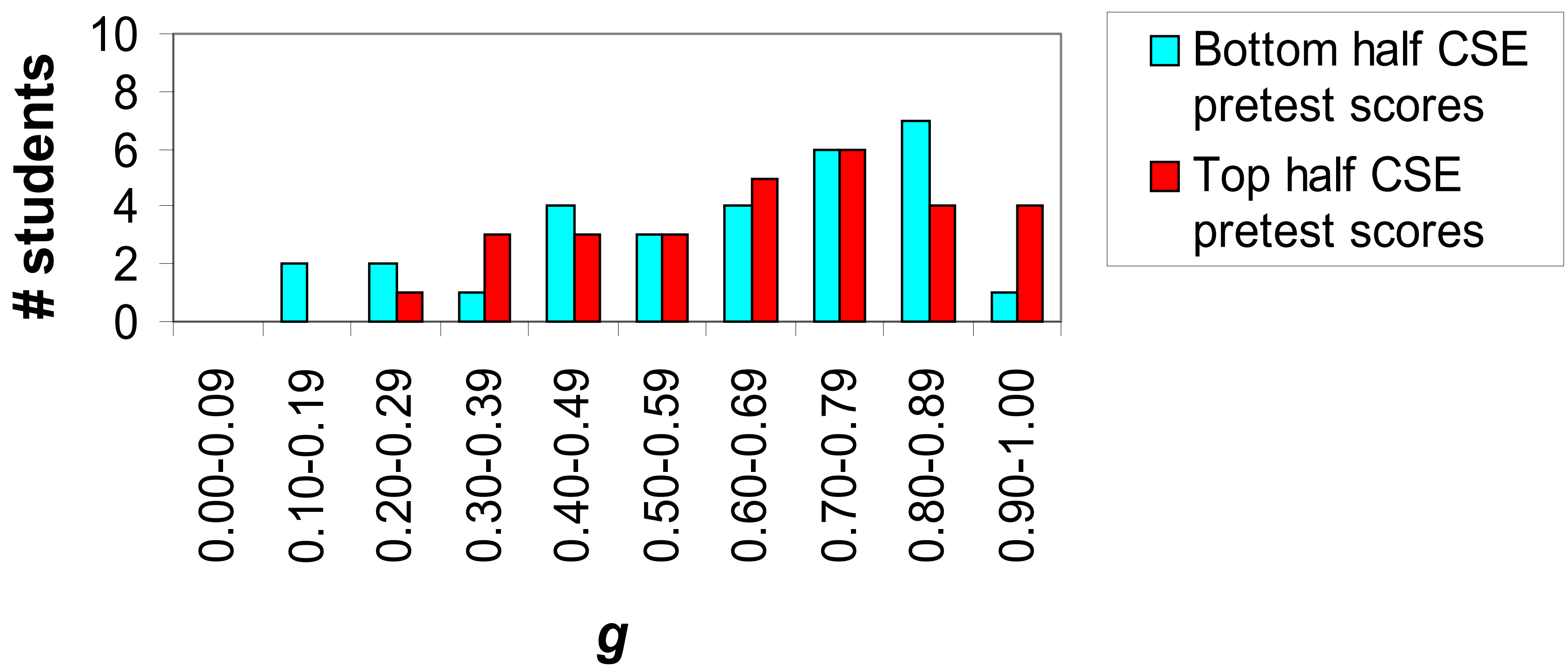
Possible "hidden variables" in students' pre-instruction mental state:

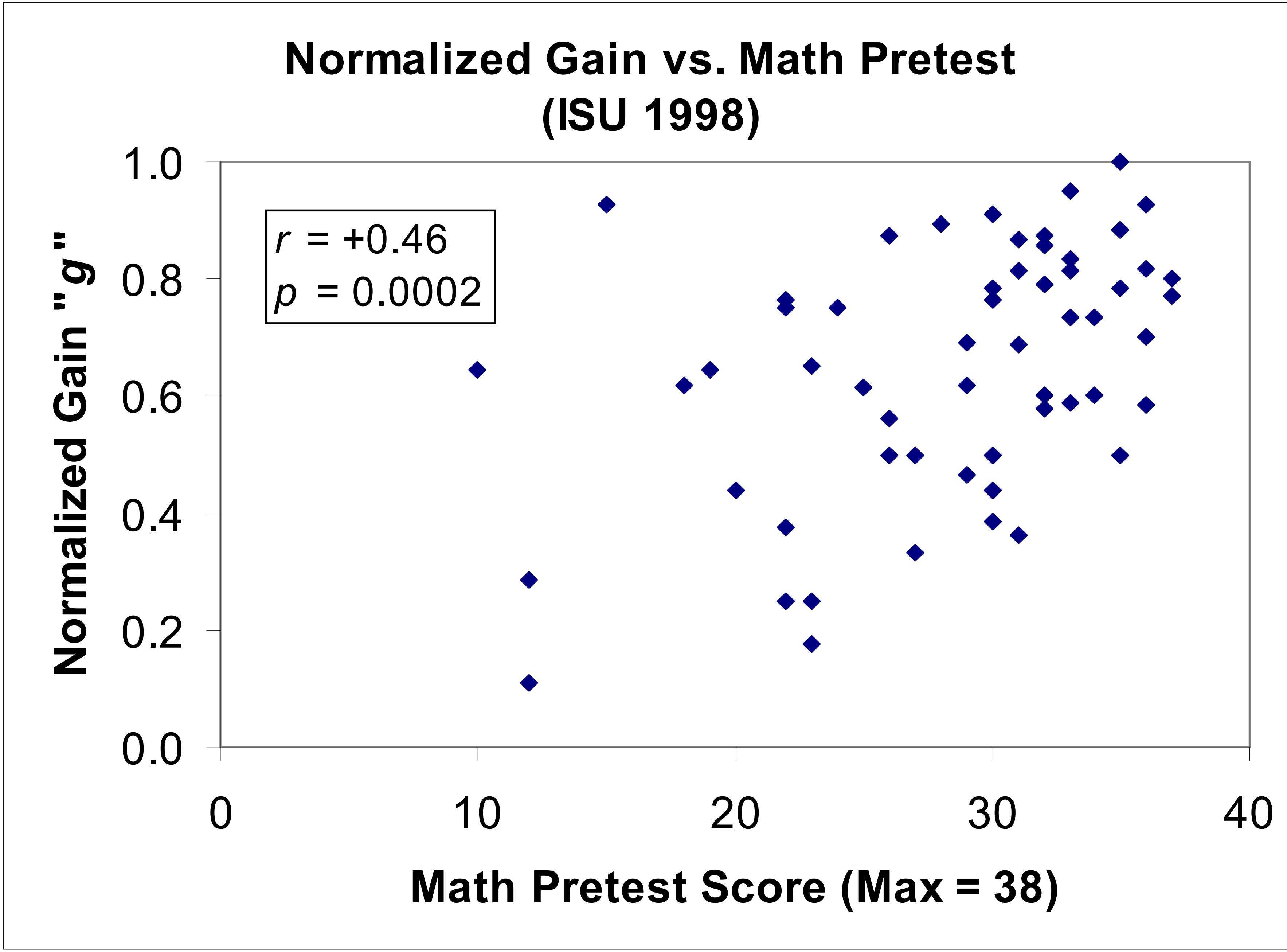
- mathematical skill *R. Hake et al., 1994; M. Thoresen and C. Gross, 2000; D. Meltzer, PERS of AJP (in press)*
- spatial visualization ability *R. Hake 2002*
- gender *L. McCullough 2000; R. Hake 2002*
- reasoning ability *J. M. Clement, 2002*
- *and even pretest score!?* *C. Henderson, K. Heller, and P. Heller, 1999*


Normalized Gain vs. CSE Pretest Score
(ISU 1998)
r = 0.0
Normalized Gain "g"
1.0
0.8
0.6
0.4
0.2
0.0
0
15
30
45
60
75
CSE Pretest Score (% correct)

Distribution of Gains: ISU 1998

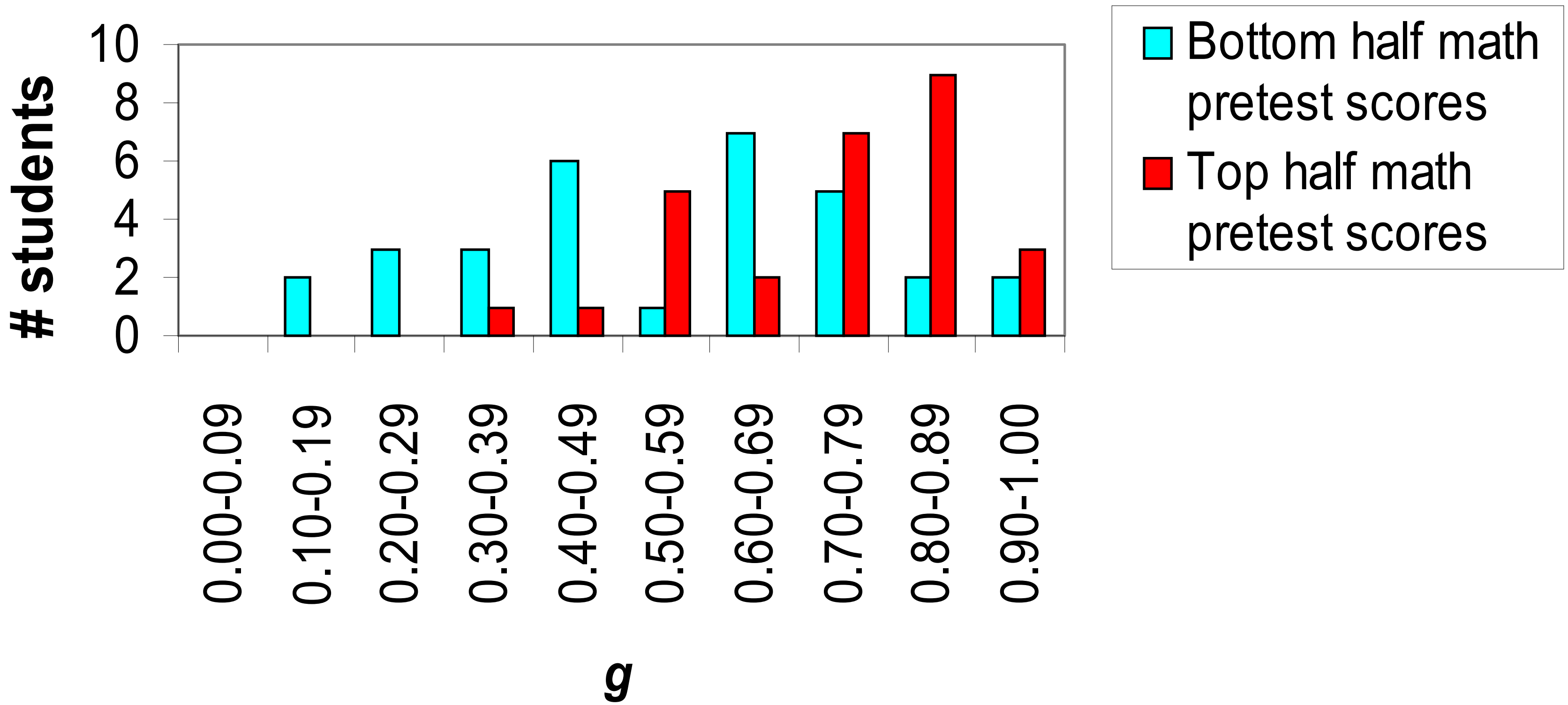
(high and low CSE pretest scores)




Normalized Gain vs. Math Pretest
(ISU 1998)
r = +0.46
p = 0.0002
Normalized Gain "g"
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
Math Pretest Score (Max = 38)

Distribution of Gains: ISU 1998

(high and low math pretest scores)

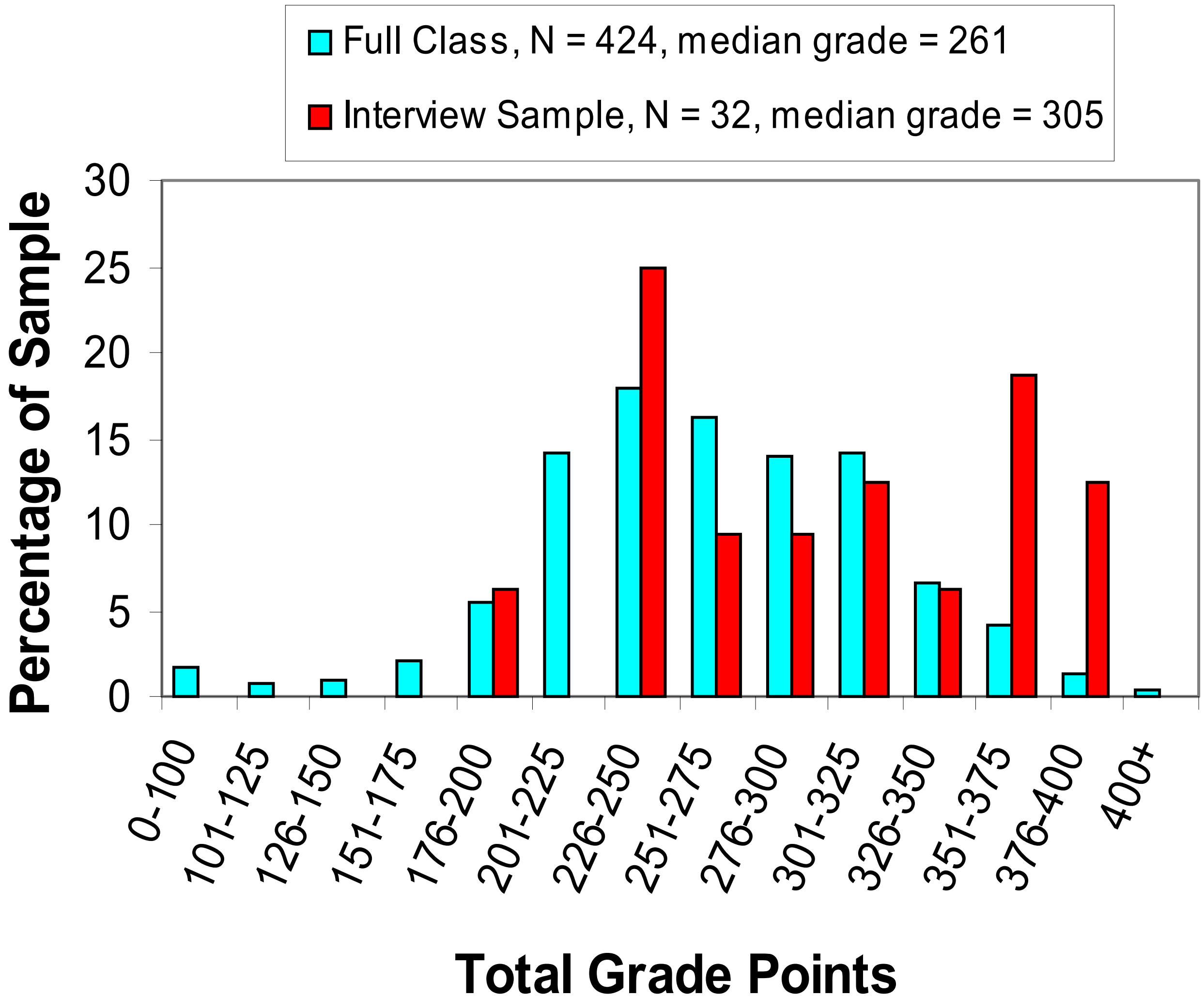


Sample-Selection Bias

- self-selection factor in interview samples
  - interviewees tend to be above-average students
- biasing due to student availability
  - students attending recitations may have above-average grades
- spring semester/fall semester differences
  - possible tendency for off-sequence courses to attract better-prepared students

Grade Distributions: Interview Sample vs. Full Class

(DEM, 2002)



***Interview Sample*:**

34% above 91st percentile; 50% above 81st percentile

Grade Comparison: Students attending recitation vs. All students

J. Dostal and DEM, 2000

	N	Scores on Exam #1 [*before* using worksheets]
Students using special worksheets (ALL attended recitation session)	129	69% (std.dev. = 20%)
Students *not* using special worksheets (MOST attended recitation session)	325	65% (std. dev. = 18%)

Difference of 4% is statistically significant **(p < 0.05)**

(*Same difference* found on final exam on "non-worksheet" questions)

Score Comparison on Vector Concept Quiz

fall-semester courses vs. spring-semester courses

N. L. Nguyen and DEM, PERS of AJP (in press)

Algebra-based physics: **A-I (**mechanics); **A-II** (E&M)
Calculus-based physics: **C-I** (mechanics); **C-II** (E&M, thermo, optics)

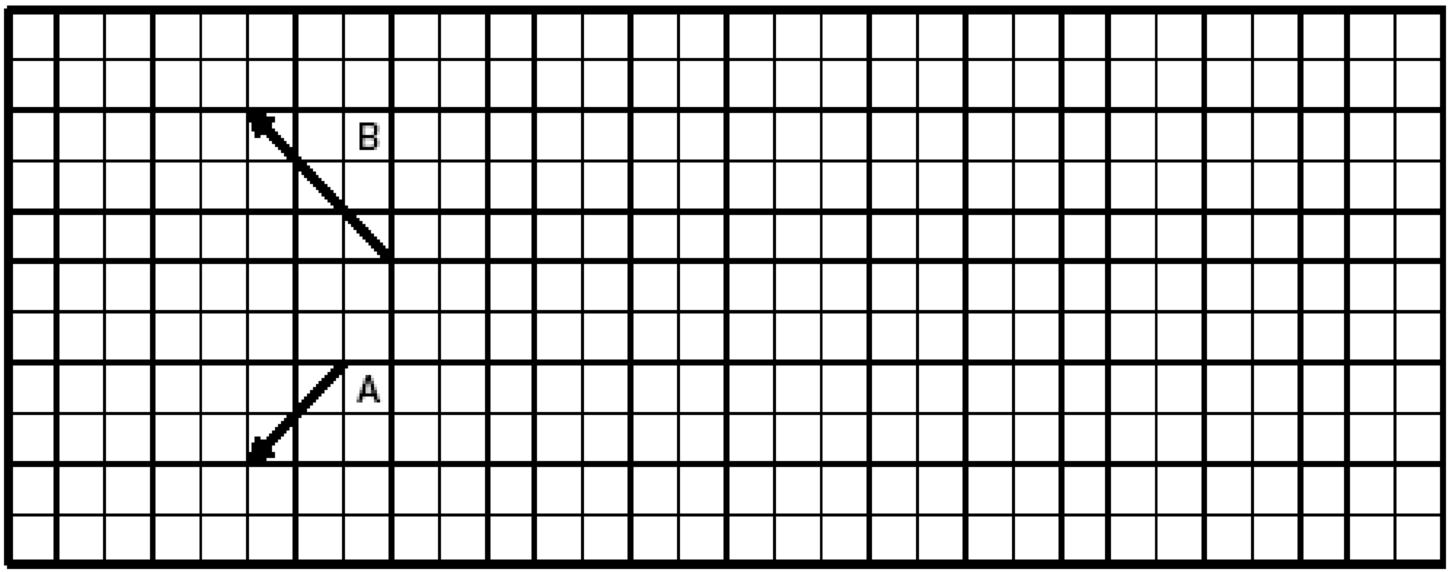
(Quiz given during first week of class)

	Fall	Spring	
A-I	44% (n = 287)	51%* (n = 233)	$p < 0.001$
C-I	65% (n = 192)	74%* (n = 416)	$p = 0.0003$
A-II	63%* (n = 83)	61% (n = 118)	*not significant*
C-II	83%* (n = 313)	78% (n = 389)	$p < 0.01$

* *"off-sequence" course*

Vector Concept Quiz

5. In the figure below there are two vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$. Draw a vector $\vec{\mathbf{R}}$ that is the sum of the two, (i.e. $\vec{\mathbf{R}} = \vec{\mathbf{A}} + \vec{\mathbf{B}}$). Clearly label the resultant vector as $\vec{\mathbf{R}}$.



Fundamental Quandary: Assessment of "Knowledge" or "Learning"?

(To analyze motion of particle, initial position ***and*** *momentum required. And to analyze student understanding? . . .)*

- To assess the impact of the teaching environment, we examine students "before" and "after." How do we measure magnitude of learning effect?
- Two students at same instantaneous knowledge "point" may be following very different trajectories. How can they be distinguished?

(Imagine ensemble of points representing individual students' mental state-functions. The trajectory of the ensemble is influenced by the teaching "force field," but also depends on initial momentum distribution.)

“Dynamic” Assessment?

Cf. *C. S. Lidz,* Dynamic Assessment *(Guilford, New York, 1987)*

- Even within the time period of a single interview, a student’s mental state may vary significantly.
  - random fluctuation, or secular change?
- Full description of mental state function requires dynamical information, i.e., rates of change, reaction to instructional “perturbation,” etc. (and remember student state-function is multi-dimensional!)
- Full analysis of teaching/learning environment will require broad array of interaction parameters.
- Simplification a practical necessity (just as in all other physics research!), but can’t lose sight of underlying reality.

Conclusion

- “Detector” design for data collection in PER has just begun to scratch the surface.
- We need to improve identification and control of variables.
- Dynamic, time-dependent assessment is likely to increase in importance.