Evidence-based teaching in introductory biology

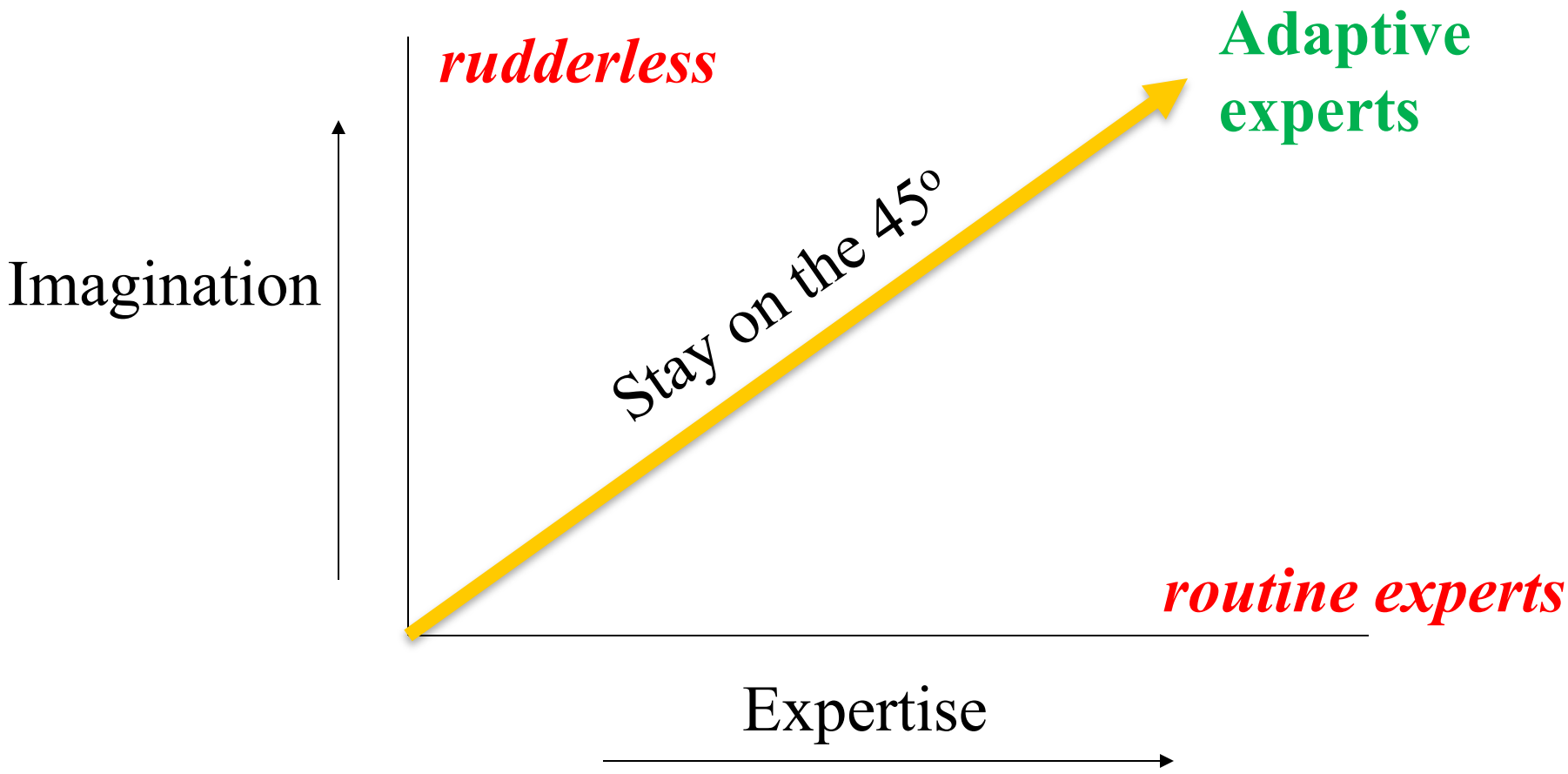
Scott Freeman, Department of Biology

University of Washington

srf991@u.washington.edu

Why are we still lecturing?

But first: The goal (of higher education)



Thank you: John Bransford (pers. comm. and Bransford et al. 2000. How People Learn (NAP: WashDC)
Hatano, G. & K. Inagaki. 1986. Child Development and Education in Japan (W.H. Freeman, New York)
Schwartz et al. in Mestre, ed. Transfer of Learning from a Modern Interdisciplinary Perspective.

Today's big question:

How can we lower failure rates—and help capable but underprepared students—in introductory biology courses?

Research on the introductory sequence required for biology-related majors at the University of Washington:

- Bio180: evolution, Mendelian genetics, ecology
- Bio200: molecular genetics, cell biology, development
- Bio220: plant and animal physiology

Bio180 background:

	2000-2007	2008	2009-
Students/qtr	340	390	700
Students/year	1,200	1,350	2,100

5,650 students in 2011 freshman class … ~40% of all undergrads at UW are taking Bio180

10% of UW freshmen are first in their families to attend college; >50% receive financial aid; 1/3rd eligible for Pell grants; 25% pay no tuition.

Bio 180 demographics:

Most students are sophomores (Chem prereq)

Gender & ethnicity: 61% female; 39% male

- 44.6% white
- 45.3% Asian-American and International
- 8.4% underrepresented minorities

~30% ESL

90% pre-grad/professional school

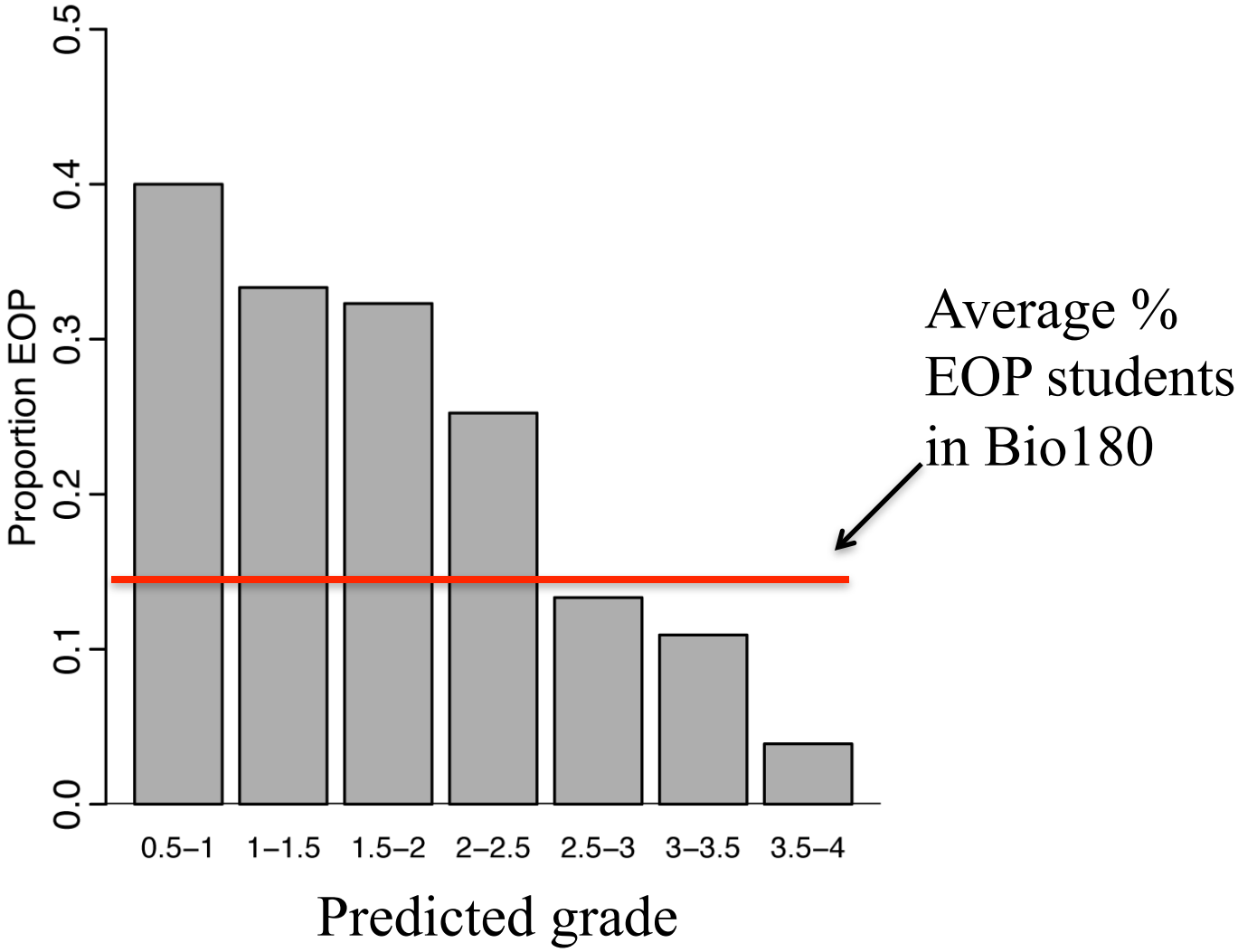
Bio180 performance thresholds

Advance to Bio200: minimum 1.5 (4.0 scale)

Declare major: minimum 2.5 (OR, need to average 2.0 over the series)

For the College, the department, and the students, these are the relevant criteria for failure.

Why be concerned about the failure rate?



Two timelines:

(U.S. data)

1920: 4%

2010: 55%

1860s: first land grant colleges

1900: first community colleges

1944: GI bill

1962: James Meredith integrates the University of Mississippi

1963: Vivian Malone and James Hood integrate the University of Alabama

2010: 57% of U.S. undergrads are women

Spring 2002 Course design

Modified Socratic style

Student performance (does not include drops):

	Spr ‘02					
< 1.5	18.2%					
< 2.5	44.8%					

Spring 2003 Course Design:

Modified Socratic + 3-5 daily, active-learning exercises in class

- think/pair/share: state a hypothesis, make a prediction, interpret a graph
- exam-style questions: work, give answer, discuss
- minute papers (handed in but not graded): muddiest point, write an exam question
- case studies on tough topics: informal groups
- in-class demonstrations with student participation

Spring 2003 Course Design Results

Student performance:

	Spr ‘02	Spr ‘03				
< 1.5	18.2%	15.8%				
< 2.5	44.8%	42.3%				

Who is failing, and why?

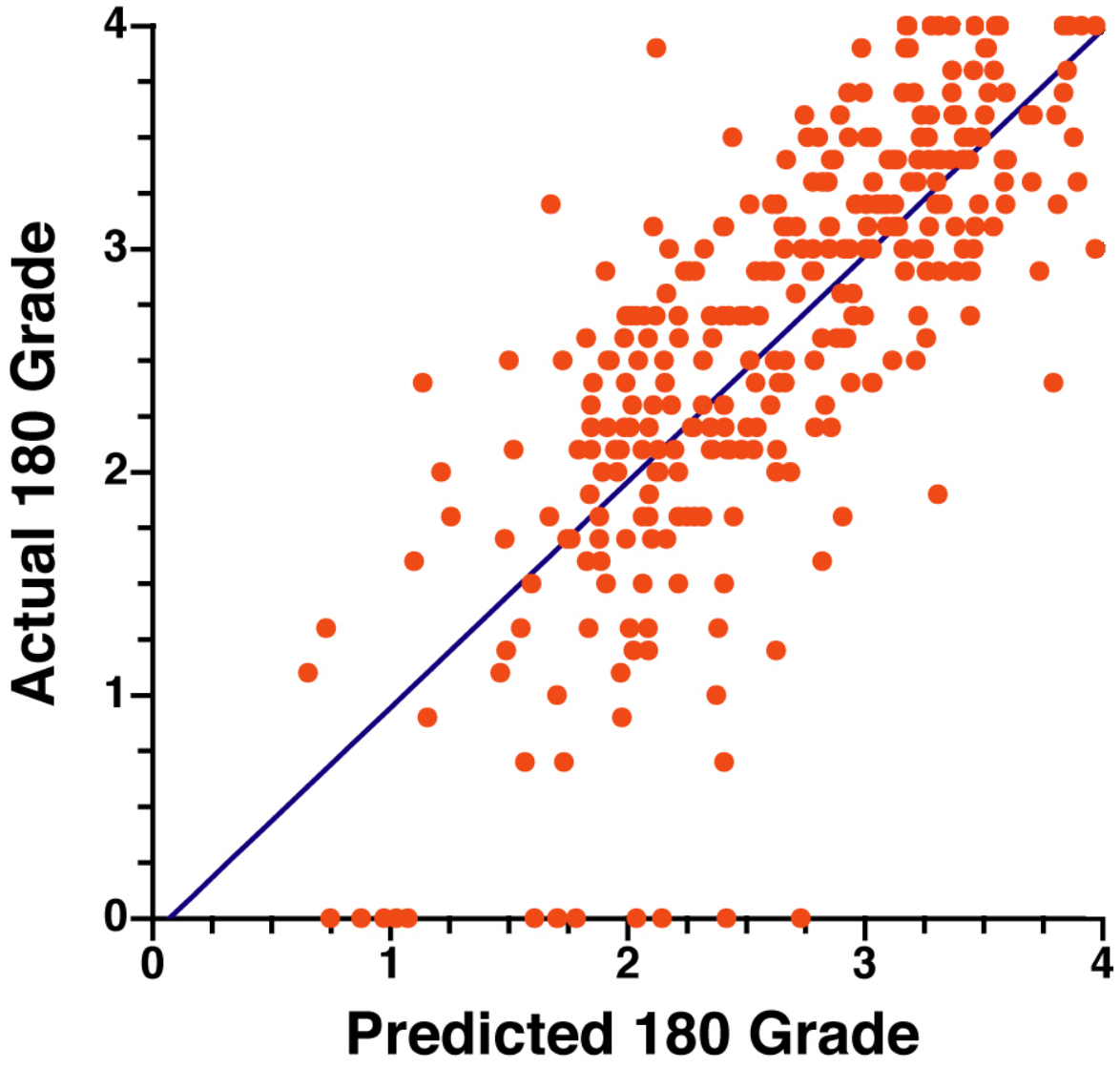
Analyze 3,338 students in Bio180/200/220, 2001-2005

Gender	H.S. GPA	UW ChemGPA
Age	**SATverbal**	TOEFL score
Classrank	SATquant	EOP standing
Ethnicity	**UW GPA**	Math placement

We use a regression model to predict student grades in Bio180.

Michael Griego

Deb McGhee


Actual 180 Grade
Predicted 180 Grade
0
1
2
3
4
0
1
2
3
4

Spring 2005 Course design

Modified Socratic + 3-5 ENFORCED daily questions + weekly, peer-graded practice exam

Section A:

Cards + practice exam done individually

Cards + practice exam done in a group

(*Structured* groups: 1 low-risk, 2 medium-risk, 1 high-risk)

Section B:

Clickers + practice exam done individually

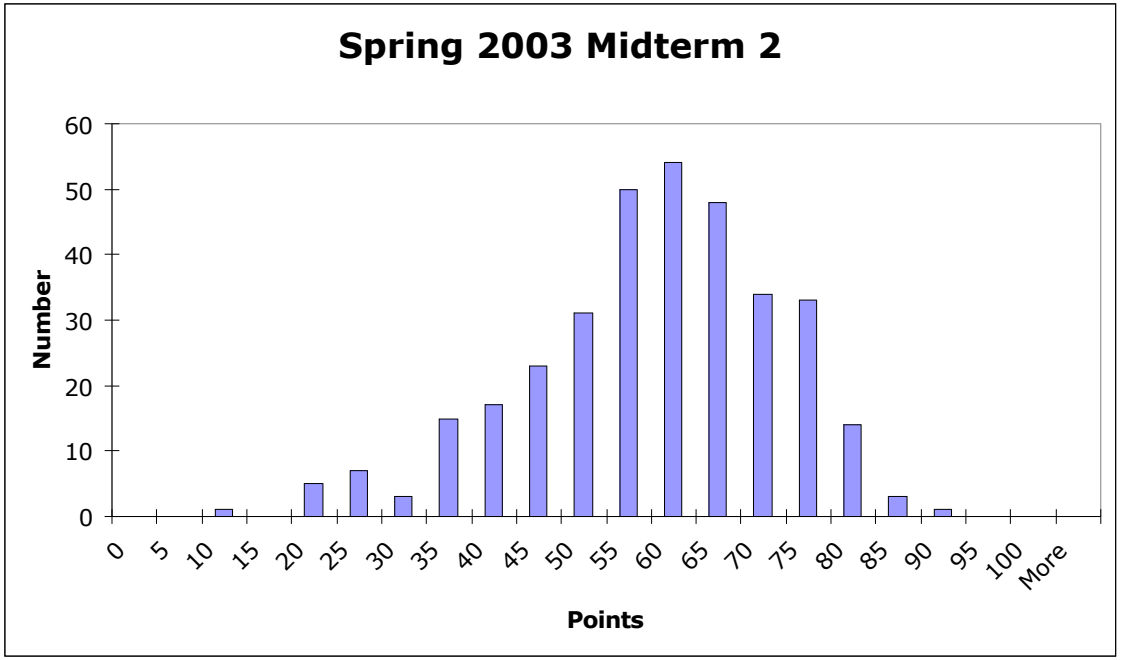
Clickers + practice exam done in a group

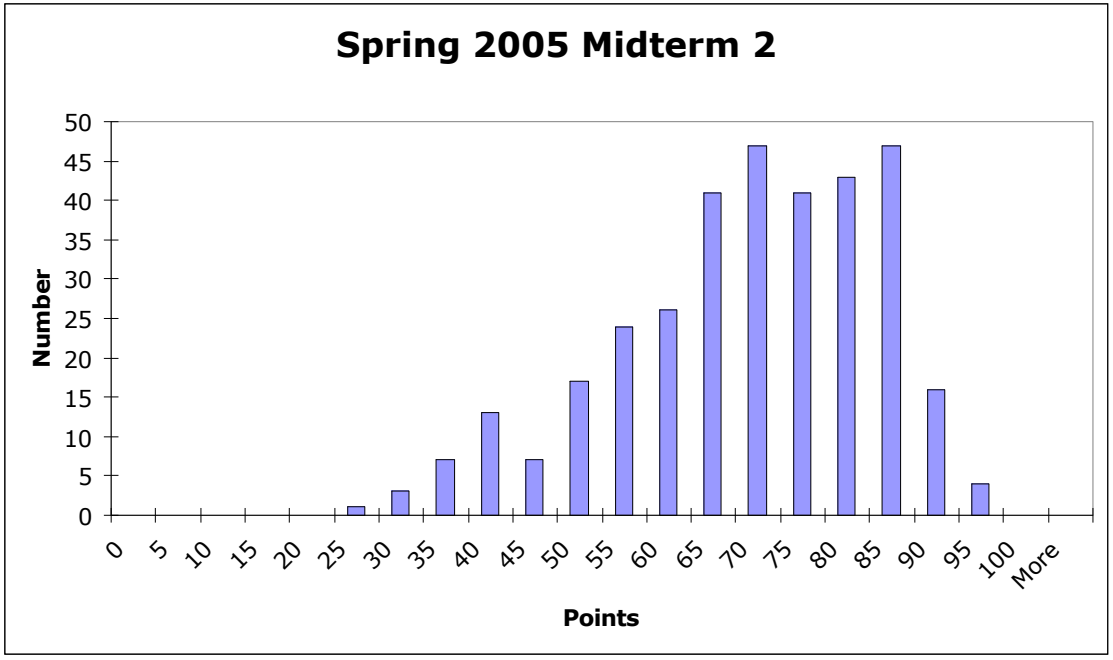
Spring 2005 Results

Student performance:

	Spr ‘02	Spr ‘03	Spr ‘05			
< 1.5	18.2%	15.8%	10.9%			
< 2.5	44.8%	42.3%	37.9%			

- Total exam points increased by an average of 14
- Median on identical midterm (spring ’03) increased by 7 points


Spring 2003 Midterm 2
Number
60
50
40
30
20
10
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 More
Points


Spring 2005 Midterm 2
Number
50
45
40
35
30
25
20
15
10
5
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 More
Points

Fall 2005 Course design

Question: How should we grade clicker points?

Modified Socratic + 3-5 daily clicker questions + weekly practice exam

Section A: Clicker points for right/wrong answers

Section B: Clicker points for participation

Fall 2005 Results

Student performance:

	Spr ‘02	Spr ‘03	Spr ‘05	Fall ‘05		
< 1.5	18.2%	15.8%	10.9%	11.7%		
< 2.5	44.8%	42.3%	37.9%	39.3%		

Total exam points increased by an average of 12 over Spr ’02, Spr ’03

Fall 2007 Course design

Questions:

1. Was failure rate lower because the class was half the size?
2. Will even more structure help high-risk students?
3. Do EOP/URM students benefit most from group or individual practice?

"No lecturing" + ~4 daily clicker questions + weekly practice exam + daily reading quiz + weekly notes check + some random call during class

Half the students did the weekly practice exam online

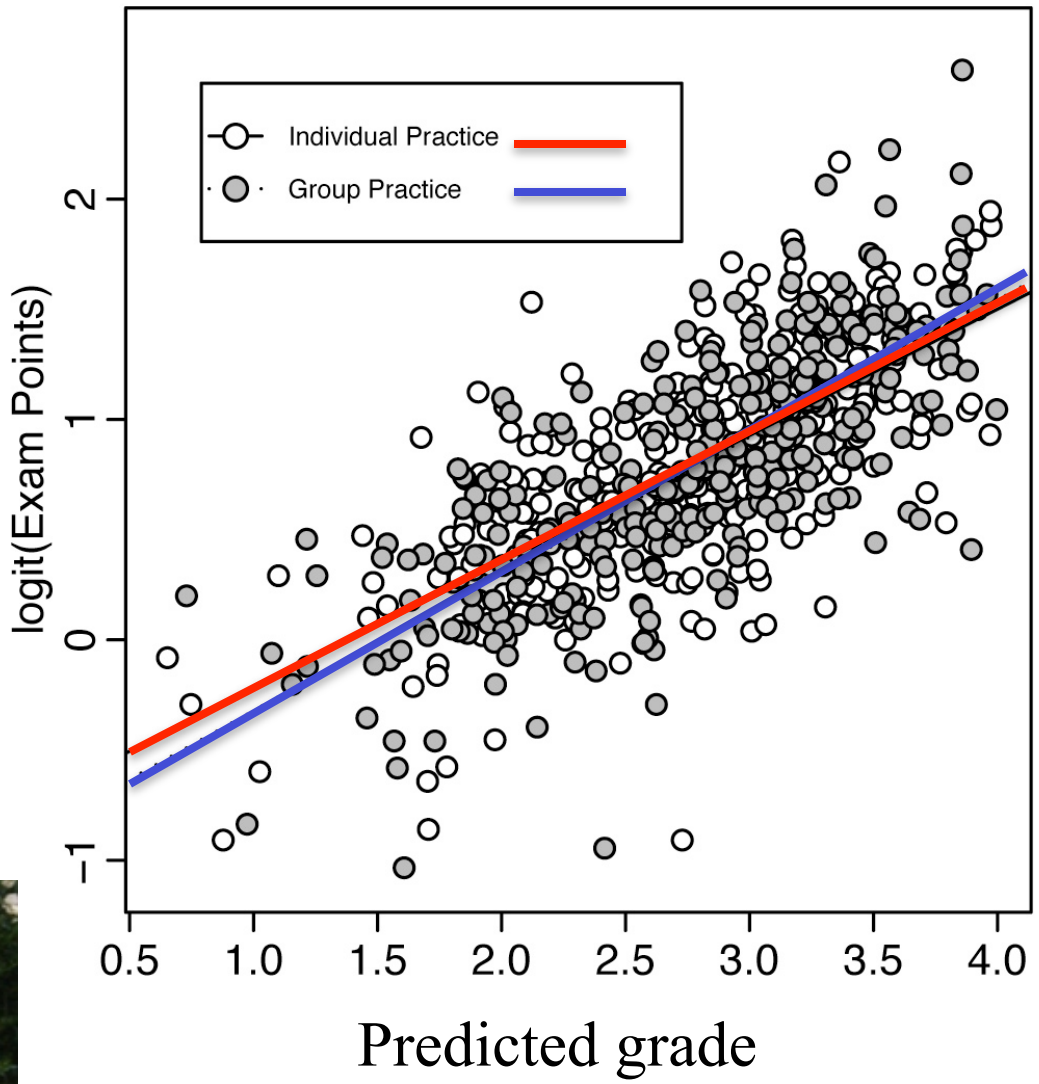
Half the students did the weekly practice exam in structured groups

Fall 2007 Results

Student performance:

	Spr ‘02	Spr ‘03	Spr ‘05	Fall ‘05	Fall ‘07	
< 1.5	18.2%	15.8%	10.9%	11.7%	7.4%	
< 2.5	44.8%	42.3%	37.9%	39.3%	33.9%	

Does group work benefit high-risk students?



Fall 2009 Course design

Questions:

1. Can we implement a highly structured course design in an EXTREMELY large-enrollment course? (700 students)

2. And live to tell the tale?

No lecturing (at all) + ~4 daily clicker questions + weekly practice exam + daily reading quiz + ~15 random call exercises in class

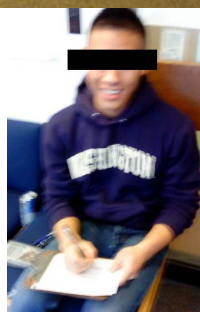
Fall 2009 Results

Student performance:

	Spr ‘02	Spr ‘03	Spr ‘05	Fall ‘05	Fall ‘07	Fall ‘09
< 1.5	18.2%	15.8%	10.9%	11.7%	7.4%	6.3%
< 2.5	44.8%	42.3%	37.9%	39.3%	33.9%	28.3%

Low structure Medium structure High structure

Why put a course point on everything? Why “enforce”?



Are exams equivalent across quarters?

Approach #1: Predicted exam score

Recruit 3 experienced graders to predict average number of points per question. Evaluate ALL exam questions, 6 quarters.

- Questions in identical format, random order
- Graders blind to hypothesis and date of exam
- Norming sessions; report average of 3 raters

	Spr ‘02	Spr ‘03	Spr ‘05	Fall ‘05	Fall ‘07	Fall ‘09
Course Average PES (100pt exam)	70.6	70.2	70.9	70.5	68.0	67.5

Are exams equivalent across quarters?

Approach #2: "Blooming" the exams

Analyze:
Can I recognize underlying patterns and structure?

Synthesize:
Can I put ideas and information together to create something new?

Evaluate:
Can I make judgments on the relative value of ideas and information?

Higher order thinking

Apply: Can I use these ideas in a new situation?

Understand: Can I explain these ideas to someone else?

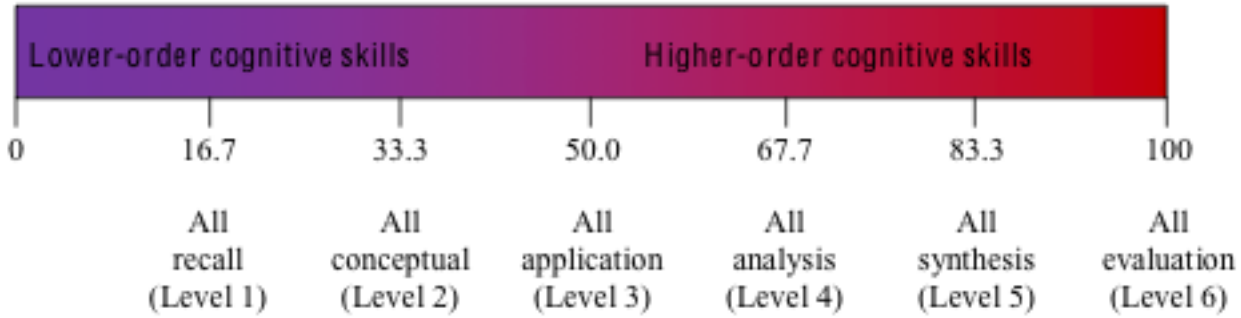
Remember: Can I recall key terms and ideas?

Lower order thinking

Computing a Weighted Bloom's Index

Recruit 3 experienced TAs to rank all exam questions on Bloom's taxonomy of learning.

$$\text{Weighted Bloom's Index} = \frac{\sum_i^n P \text{ x } B}{T \text{ x } 6} \text{ x } 100$$

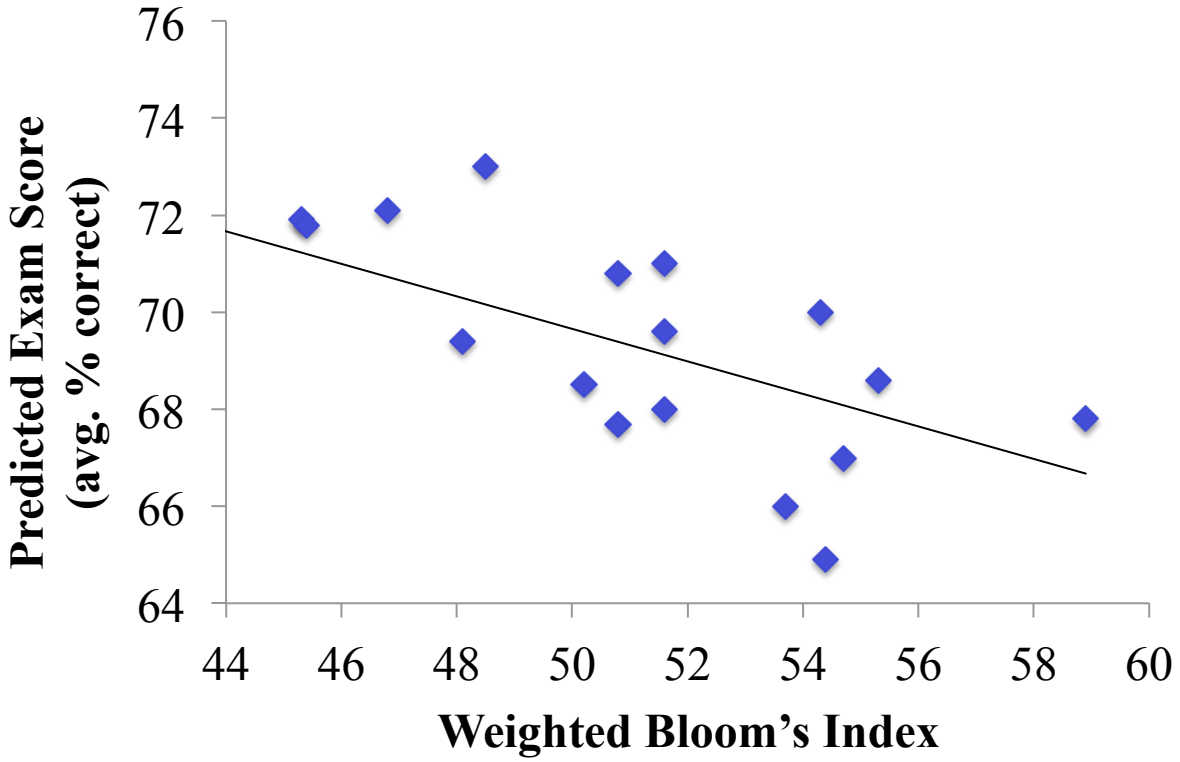


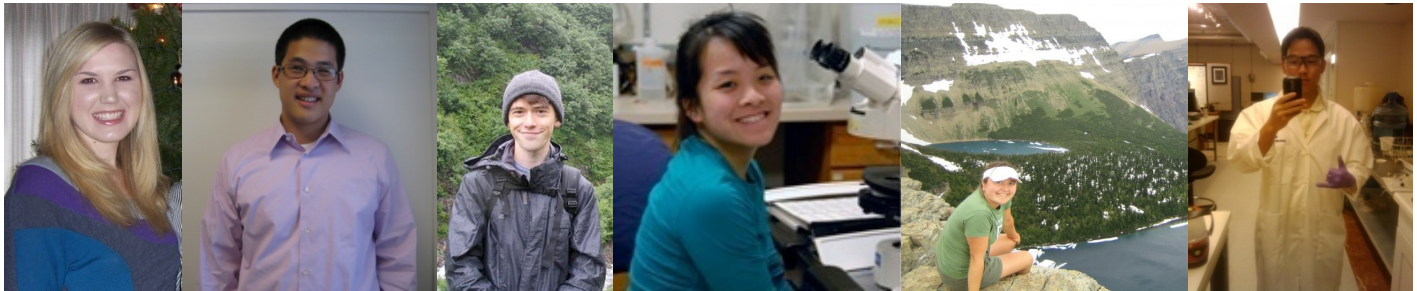
Are exams equivalent across quarters?

For Weighted Bloom's Index:

- Questions in identical format
- Graders blind to hypothesis and date of exam
- Norming sessions, then "decision rules" (following Zheng et al. 2008)

	Spr '02	Spr '03	Spr '05	Fall '05	Fall '07	Fall '09
Course Average **(weighted Bloom's index)**	45.8	52.1	46.9	52.2	52.1	53.5

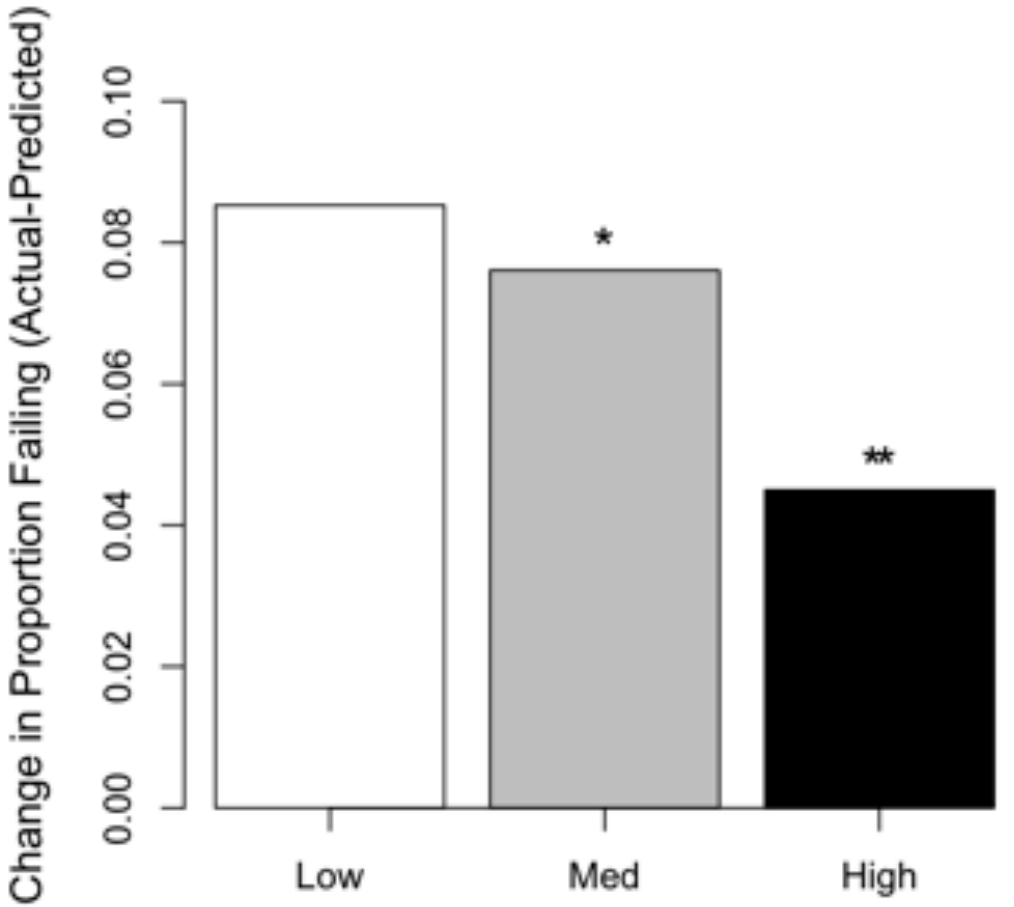

Predicted Exam Score
(avg. % correct)
76
74
72
70
68
66
64
44
46
48
50
52
54
56
58
60
Weighted Bloom's Index



Are students equivalent across quarters?

	Spring 2002	Spring 2003	Spring 2005	Autumn 2005	Autumn 2007	Autumn 2009
Predicted grade (mean)	2.46	2.57	2.64	2.67	2.85	2.70
n	327	338	334	328	339	691

Create a general linear model to explain actual grade, based on predicted grade and degree of structure in course.



2002, 03 2005 2007,09

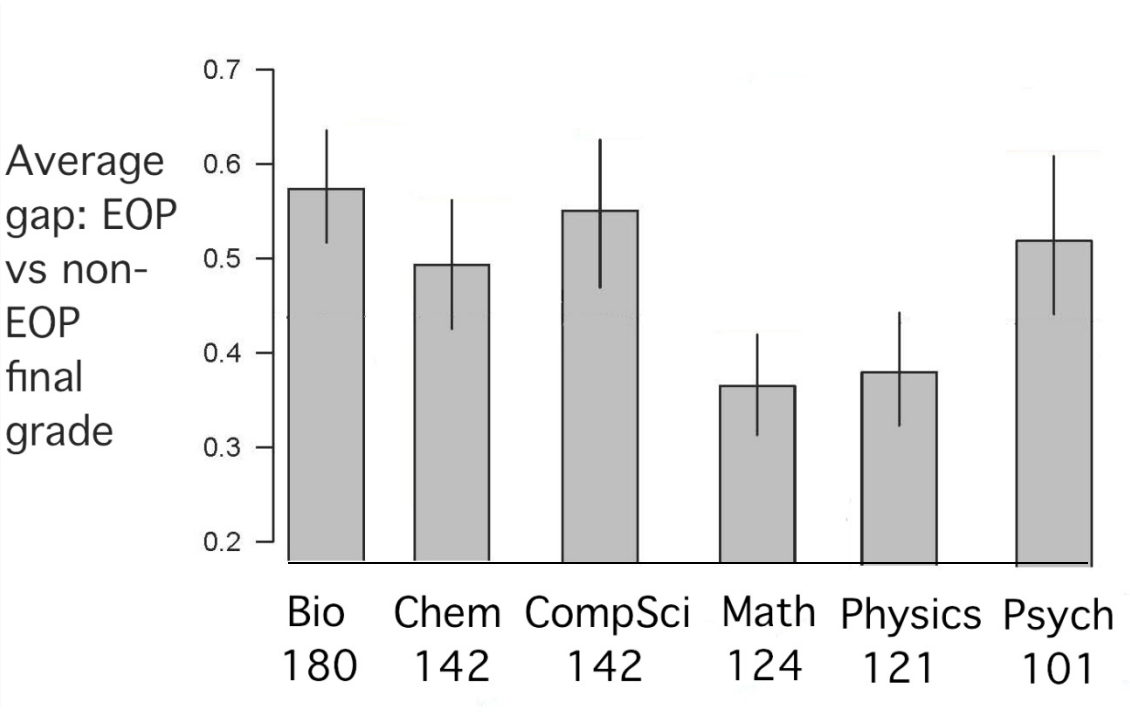
Course structure

Last question:

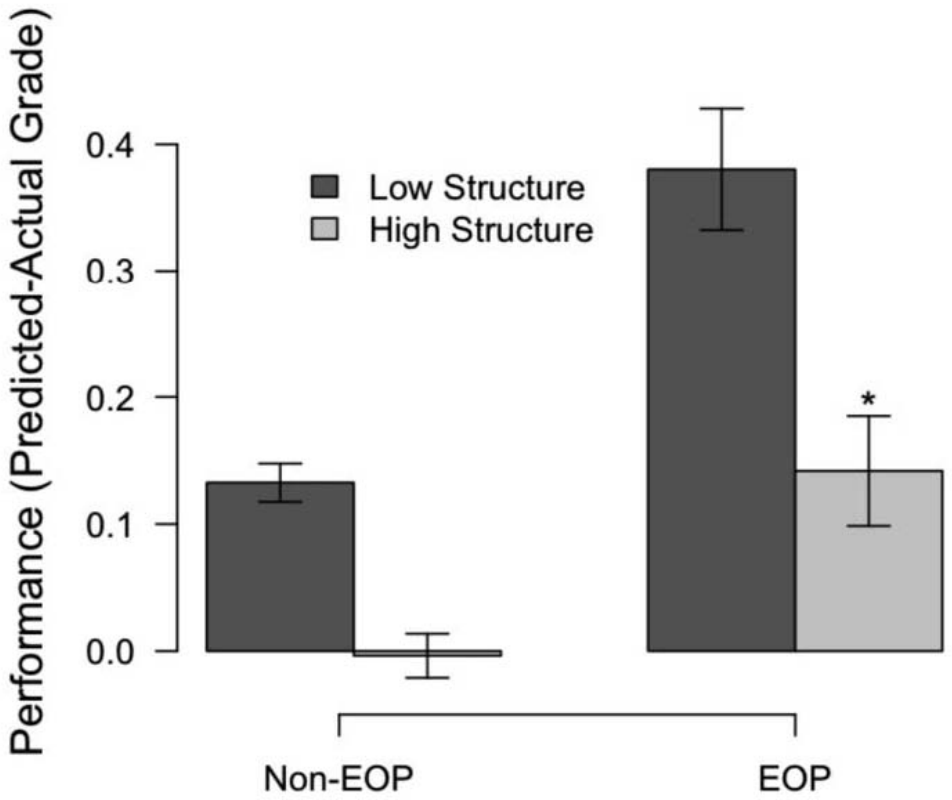
Did we reduce the achievement gap?

… without spending a lot more money? or maybe even less money?

2003-2008 (Aut/Win/Spr) averages: EOP v non-EOP final grade differences in UW gateway STEM courses

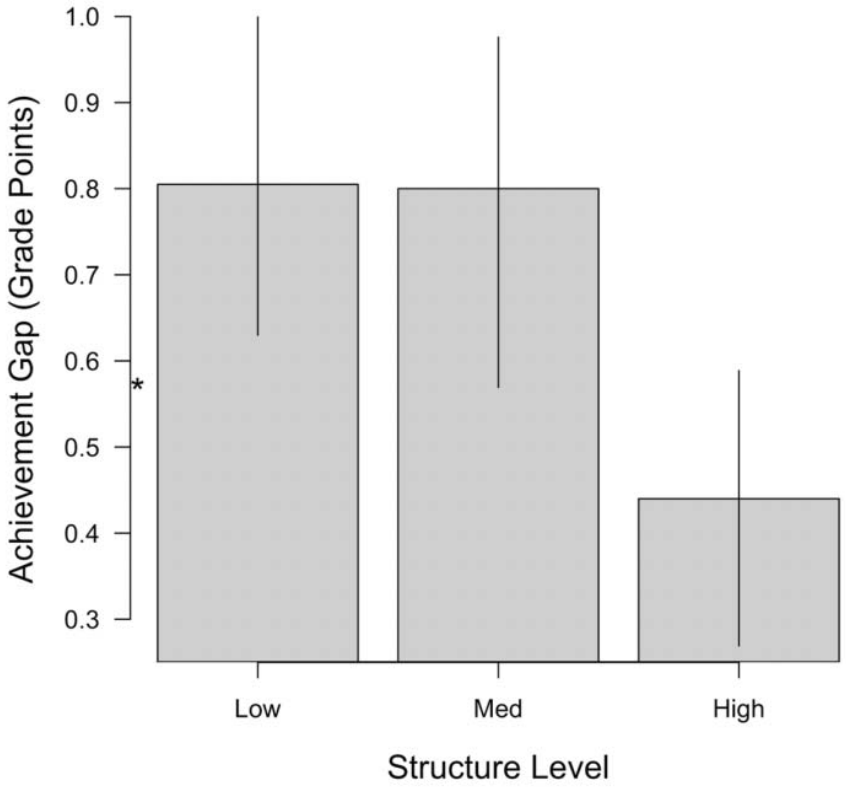


Is there an interaction between degree of course structure and EOP status? (many instructors)



General linear mixed-effects modeling and MMI:
Best models include EOP as a fixed effect; likelihood-ratio test, p = 0.0027).

Changes in the EOP vs. non-EOP achievement gap, by quarter (same instructor)



Controlling for changes in student ability/preparation (average predicted grade), there is also a drop in the achievement gap with medium structure.

What could cause a *disproportionate* increase in performance by disadvantaged students?

The Carnegie Hall hypothesis:

How do you get to Carnegie Hall? **PRACTICE!**

… and how you practice matters:
1)high-level questions (new contexts/applications);
2)group work (teach others/explain yourself, challenge and be challenged);
3)daily/weekly basis

Current questions

- Curriculum/program assessment: Are students achieving mastery of stated learning objectives?

- Does high structure work elsewhere? Does active learning work across the STEM disciplines?

- Can we promote change from the bottom up?

- Faculty development (including future faculty): Moving from evidence to action.

A clicker question from Autumn 2011: Why aren't more professors using evidence-based teaching?

	1st	2nd
	18.8	11.2
1.The data are too new—there hasn't been time to change.	9.9	5.3
2.They don't get rewarded for good teaching.	23.3	20.8
3.They haven't received training in these approaches.	6.4	5.3
4.Students don't demand it.	41.5	57.5
5.They don't have access to the curriculum, needed, and don't have time to create it themselves.		

My all-time favorite line from a course evaluation:

"Keep pushing us—we can do it!"