Closing the Achievement Gap: Evidence-Informed Teaching and Learning in Biology.

David Haak, Department of Biology
Indiana University
dhaak@indiana.edu
First: The goal (of higher education)

First: The goal (of higher education)

First: The goal (of higher education)

Imagination

rudderless

routine experts

Expertise

First: The goal (of higher education)

- Stay on the 45°
- rudderless
- routine experts

Imagination

Expertise

First: The goal (of higher education)

- Imagination
- Expertise

Stay on the 45°

rudderless

routine experts

Adaptive experts

Thank you: John Bransford (pers. comm. and Bransford et al. 2000. How People Learn (NAP: WashDC)
Today’s big question:

How can we lower failure rates—and help capable but underprepared students—in introductory biology courses?
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
Today’s big question:

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

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 and animal physiology
Today’s big question:

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

What do I mean by structure?
Today’s big question:

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

What do I mean by structure?

Modified Socratic + Minimal active learning

Think/pair share Case studies
Today’s big question:

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

What do I mean by structure?

<table>
<thead>
<tr>
<th>Modified Socratic + Minimal active learning</th>
<th>Modified Socratic + Active Learning + enforced work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Think/pair share</td>
<td>Peer grading</td>
</tr>
<tr>
<td>Case studies</td>
<td>Practice exams</td>
</tr>
<tr>
<td></td>
<td>Weekly quizzes</td>
</tr>
</tbody>
</table>
Today’s big question:

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

What do I mean by structure?

<table>
<thead>
<tr>
<th>Modified Socratic + Minimal active learning</th>
<th>Modified Socratic + Active Learning + enforced work</th>
<th>Modified Socratic + AL + enforced + peer work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Think/pair share</td>
<td>Peer grading</td>
<td>Daily quizzes</td>
</tr>
<tr>
<td>Case studies</td>
<td>Practice exams</td>
<td>Random call</td>
</tr>
<tr>
<td></td>
<td>Weekly quizzes</td>
<td>Peer to peer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Practice exams</td>
</tr>
</tbody>
</table>
Today’s big question:

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

What do I mean by structure?

<table>
<thead>
<tr>
<th>Low</th>
<th>Modified Socratic + Minimal active learning</th>
<th>Modified Socratic + Active Learning + enforced work</th>
<th>Modified Socratic + AL + enforced + peer work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Think/pair share</td>
<td>Peer grading</td>
<td>Practice exams</td>
<td>Daily quizzes</td>
</tr>
<tr>
<td>Case studies</td>
<td>Weekly quizzes</td>
<td>Practice exams</td>
<td>Random call</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Peer to peer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Practice exams</td>
</tr>
</tbody>
</table>
Today’s big question:
How can we lower failure rates—and help capable but underprepared students—in introductory biology courses? **STRUCTURE!**

What do I mean by structure?

<table>
<thead>
<tr>
<th>Low</th>
<th>Medium</th>
<th>Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified Socratic + Minimal active learning</td>
<td>Modified Socratic + Active Learning + enforced work</td>
<td>Modified Socratic + AL + enforced + peer work</td>
</tr>
<tr>
<td>Think/pair share Case studies</td>
<td>Peer grading Practice exams Weekly quizzes</td>
<td>Daily quizzes Random call Peer to peer Practice exams</td>
</tr>
</tbody>
</table>
Today’s big question:

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

What do I mean by structure?

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified Socratic + Minimal active learning</td>
<td>Modified Socratic + Active Learning + enforced work</td>
<td>Modified Socratic + AL + enforced + peer work</td>
<td></td>
</tr>
<tr>
<td>Think/pair share</td>
<td>Practice exams</td>
<td>Daily quizzes</td>
<td>Peer to peer Peer to peer Practice exams</td>
</tr>
<tr>
<td>Case studies</td>
<td>Weekly quizzes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Today’s big question:

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

Study design:
Today’s big question:

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

Study design:

6 Quarters (across 6 years)
Today’s big question:

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

Study design:

6 Quarters (across 6 years)

3 Structure levels (Low, Medium, High)
Today’s big question:

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

Study design:

6 Quarters (across 6 years)

3 Structure levels (Low, Medium, High)

Same instructor (Scott Freeman)
Today’s big question:

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

Study design:  

**6 Quarters (across 6 years)**

**3 Structure levels (Low, Medium, High)**

**Same instructor (Scott Freeman)**
Today’s big question:

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

Study design:

6 Quarters (across 6 years)

3 Structure levels (Low, Medium, High)

Same instructor (Scott Freeman)

Outcome measure:

Exam points (Common Qs)
Today’s big question:
How can we lower failure rates—and help capable but underprepared students—in introductory biology courses? STRUCTURE!

Study design:
6 Quarters (across 6 years)
3 Structure levels (Low, Medium, High)
Same instructor (Scott Freeman)

Outcome measure:
Exam points
(Common Qs)
Course grade
(Corrected)
Study System
Bio 180 background:

<table>
<thead>
<tr>
<th></th>
<th>2000-2007</th>
<th>2008</th>
<th>2009-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students/qtr</td>
<td>340</td>
<td>390</td>
<td>700</td>
</tr>
<tr>
<td>Students/year</td>
<td>1200</td>
<td>1350</td>
<td>2,100</td>
</tr>
</tbody>
</table>

5,338 students in 2009 freshman class ... > 40% of all undergrads at UW are taking Bio 180
Bio 180 demographics:

Most students are sophomores (Intro. Chem. prereq)
Bio 180 demographics:

Most students are sophomores (Intro. Chem. prereq)

- Female: 39%
- Male: 61%
Bio 180 demographics:

Most students are sophomores (Intro. Chem. prereq)
Bio 180 demographics:

Most students are sophomores (Intro. Chem. prereq)

~30% ESL
Bio 180 demographics:

Most students are sophomores (Intro. Chem. prereq)

- Female: 39%
- Male: 61%

- White: 41%
- Asian-American: 51%
- Underrepresented groups: 9%

~30% ESL

90% pre-grad/professional school
Bio180 performance thresholds:
Bio180 performance thresholds:

Advance to Bio200: minimum 1.5 (4.0 scale)
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)
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.
Spring 2002 Course design

Modified Socratic style
Spring 2002 Results

Modified Socratic style

Student performance:

Percent below threshold

Spr ‘02

<1.5

<2.5
Spring 2003 Course design

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

Student performance:

<table>
<thead>
<tr>
<th>Percent below threshold</th>
<th>Spr '02</th>
<th>Spr '03</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1.5</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>&lt;2.5</td>
<td></td>
<td>37.5</td>
</tr>
</tbody>
</table>
Spring 2005 Course design

Modified Socratic + 3-5 ENFORCED daily questions + weekly, peer-graded practice exam
Spring 2005 Course design

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

Section A:
  Cards

Section B:
  Clickers
Spring 2005 Results

Student performance:

- Spr '02
- Spr '03
- Spr '05

Percent below threshold

<1.5
<2.5
Spring 2005 Results

Student performance:

- Total exam points increased by an average of 14

- Percent below threshold:
  - Spr ’02
  - Spr ’03
  - Spr ’05

- <1.5
- <2.5
• Total exam points increased by an average of 14

• Median on identical midterm (spring ’03) increased by 7 points
Fall 2005 Course design

Modified Socratic + 3-5 ENFORCED daily questions + weekly, peer-graded practice exam + Clickers
Fall 2005 Course design

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

Question: How should we grade clicker points?
Fall 2005 Course design

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

Question: How should we grade clicker points?

Section A: Clicker points for right/wrong answers

Section B: Clicker points for participation
Fall 2005 Results

Student performance:

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?
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?
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?
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
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
Fall 2007 Results

Student performance:

- Spr ‘02
- Spr ‘03
- Spr ‘05
- Fall ‘05
- Fall ‘07

- <1.5
- <2.5
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:

Percent below threshold

<1.5  <2.5

Spr '02  Spr '03  Spr '05  Fall '05  Fall '07  Fall '09
Fall 2009 Results

Student performance:

Percent below threshold

Spr '02  Spr '03  Spr '05  Fall '05  Fall '07  Fall '09
Low structure  Medium structure  High structure
Student equivalency-predicting performance

Analyze 3,338 students in Bio180/200/220, 2001-2005
Student equivalency-predicting performance

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

<table>
<thead>
<tr>
<th>Gender</th>
<th>H.S. GPA</th>
<th>UW ChemGPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>SATverb</td>
<td>TOEFL score</td>
</tr>
<tr>
<td>Class rank</td>
<td>SATquant</td>
<td>EOP standing</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>UW GPA</td>
<td>Math placement</td>
</tr>
</tbody>
</table>

We use a regression model to predict student grades in Bio180.
Student equivalency-predicting performance

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

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

SATverb

UW GPA

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

![Graph showing predicted vs. actual grades with a linear relationship]

Actual grade (GPA Points) vs. Predicted grade
Students are not equivalent across quarters

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>327</td>
<td>338</td>
<td>334</td>
<td>328</td>
<td>339</td>
<td>691</td>
</tr>
</tbody>
</table>
Are we reducing the failure rate?
Are we reducing the failure rate?

GLMM, N=2267, *p=0.06, **p=0.0004
Why be concerned about the failure rate?
Why be concerned about the failure rate?

Average % EOP students in Bio180
Active learning EOP/URM

Questions:

Do EOP/URM students benefit most from group or individual practice?
Experimental design

Questions:
Do EOP/URM students benefit most from group or individual practice?

Binned by risk of failure

Random assignment to group or individual work
Experimental design

Questions:

Do EOP/URM students benefit most from group or individual practice?

Binned by risk of failure

   Random assignment to group or individual work

Structured groups using predicted performance

   1-High risk, 2-Medium risk, 1-Low risk
Experimental design

Questions:
Do EOP/URM students benefit most from group or individual practice?

Binned by risk of failure
Random assignment to group or individual work

Structured groups using predicted performance
1-High risk, 2-Medium risk, 1-Low risk

Implemented across 2 Quarters (Sp05, Fa07)
Does group work benefit high-risk students?
Does group work benefit high-risk students?

Predicted grade

Exam points (Logit Transformed)

-1.0  -0.5  0.0  0.5  1.0  1.5  2.0  2.5  3.0  3.5  4.0

1.0  1.5  2.0  2.5  3.0  3.5  4.0

Predicted grade

Individual

Group
Does group work benefit high-risk students?

Predicted grade

Exam points (Logit Transformed)

Predicted grade
Last question:

Did we reduce the achievement gap?
Last question:

Did we reduce the achievement gap?

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

Did we reduce the achievement gap?

… without spending a lot more money? or maybe even less money?
Is there an interaction between degree of course structure and EOP status? (many instructors)

Across quarters we see improvement in performance.

GLMM: LRT, $p = 0.0027$
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 a drop in the achievement gap with increased structure.

Error bars indicate 95% confidence intervals.
What could cause a *disproportionate* increase in performance by underprepared students?
What could cause a *disproportionate* increase in performance by underprepared students?

The Carnegie Hall hypothesis:

How do you get to Carnegie Hall?
What could cause a disproportionate increase in performance by underprepared students?

The Carnegie Hall hypothesis:

How do you get to Carnegie Hall? PRACTICE!
What could cause a *disproportionate* increase in performance by underprepared 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);
What could cause a disproportionate increase in performance by underprepared 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) and individual practice;
What could cause a *disproportionate* increase in performance by underprepared 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) and individual practice;
3. daily/weekly quizzes (enforcement)
Questions for the future:

• Can we eliminate achievement gaps?

• How to assess quality of teaching?

• Curriculum assessment: Are our courses and programs meeting stated learning objectives?
Acknowledgements

UW-BERG: Biology Education Research Group

NSF
HHMI HOWARD HUGHES MEDICAL INSTITUTE
My all-time favorite line from a course evaluation:

“Keep pushing us—we can do it!”
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

<table>
<thead>
<tr>
<th>Course Average PES (100pt exam)</th>
<th>Spr ‘02</th>
<th>Spr ‘03</th>
<th>Spr ‘05</th>
<th>Fall ‘05</th>
<th>Fall ‘07</th>
<th>Fall ‘09</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>70.6</td>
<td>70.2</td>
<td>70.9</td>
<td>70.5</td>
<td>68.0</td>
<td>67.5</td>
</tr>
</tbody>
</table>
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?

**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?
Computing a Weighted Bloom’s Index

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

Weighted Bloom’s Index = \frac{\sum_{i}^{n} P \times B}{T \times 6} \times 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)

<table>
<thead>
<tr>
<th>Course Average (weighted Bloom’s index)</th>
<th>Spr ‘02</th>
<th>Spr ‘03</th>
<th>Spr ‘05</th>
<th>Fall ‘05</th>
<th>Fall ‘07</th>
<th>Fall ‘09</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>45.8</td>
<td>52.1</td>
<td>46.9</td>
<td>52.2</td>
<td>52.1</td>
<td>53.5</td>
</tr>
</tbody>
</table>