Undergraduate data science: Biological connections and assessing impacts

Louis J. Gross
National Institute for Mathematical and Biological Synthesis
National Institute for STEM Evaluation and Research
Departments of Ecology and Evolutionary Biology and Mathematics
University of Tennessee, Knoxville

Funded by NSF awards DUE-1544375 and DBI-1300426
Project Personnel

Pamela Bishop (PI) - Director of NISER, NIMBioS Associate Director for STEM Evaluation

Suzanne Lenhart - NIMBioS Associate Director for Education and Outreach, Professor Mathematics

Kelly Sturner - former NIMBioS Education and Outreach Coordinator

Robin Taylor - NIMBioS Evaluation Postdoctoral Fellow
National Academies Committee on Envisioning the Data Science Discipline: The Undergraduate Perspective

http://sites.nationalacademies.org/CSTB/CurrentProjects/CSTB_175246

Look for a report soon
Curricular Organization as a Data-science Problem

- Prioritization of concepts/skills
- Use of education research to guide pedagogy
- Borrowing/adapting existing materials (downscaling, not dumbing-down)
- Sequencing coverage of concepts/skills
- Enhancing high priority concepts/skills through repetition
- Assessing success
Training Fearless Biologists: Quantitative Concepts for all our Students

1. Rate of change
2. Modeling
3. Equilibria and stability
4. Structure
5. Interactions
6. Data and measurement
7. Stochasticity
8. Visualizing
9. Algorithms

Slide presented at Bio2010 public release - Sept. 10, 2002
Listing arose from Workshops at UTK in 1992 and 1994.
See *Bio2010: Transforming Undergraduate Education for Future Research Biologists* (NRC, 2003)
Rule of Five- different learning styles to meet needs of diverse students: Symbolically, Graphically, Numerically, Verbally, Data-driven

We use this approach throughout the text which includes descriptive statistics (regression, semi-log, log-log), matrix algebra (eigenvalues, eigenvectors), discrete probability, discrete dynamical systems, basic calculus, differential equations, emphasizing data and hypothesis formulation (math and biological) using Matlab and R.
What evidence is there that indeed introducing and/or motivating quantitative concepts through inclusion of data and models from the life sciences actually enhances learning of these concepts by our students?
Instructional design choices. Different choices along different instructional dimensions can be combined to produce a vast set of instructional options. The path with thicker arrows illustrates one set of choices within a space of trillions of such options.

Evidence for the effectiveness of learning quantitative concepts through concrete examples and real data over abstract methods is mostly anecdotal. Very few studies investigate learning gains in mathematics arising from the use of scientific examples.

A first step towards evaluating the potential impact of biological examples on mathematics comprehension is to develop a robust assessment instrument designed for college-level math concepts.

Quantitative Biology Concept Inventory (QBCI)
Quantitative Biology Concept Inventory Construction

Objectives/Planning

Content Specification

Item Development

Test administration

Validity Analysis
Objectives/Planning

Literature review of concept inventories
Content Specification

• Calculus
  – Rates of change
  – Sums and Integration
  – Modeling
  – Interpreting Graphs and Data
2. Slow motion videos were used to collect time-series data on the vertical position of a cat’s tongue as it drank water. The graph below shows the vertical position of the cat’s tongue above a bowl of water during one lap of the tongue.

Over what interval is the vertical position increasing?

(a) $[-150, 30]$
(b) $[-70, 0]$
(c) $[-50, 30]$
(d) $[-30, 10]$
(e) $[-100, -60]$
Test Review and Administration

Expert review

- Student focus group completion
- Fall 2016 Administration
  - Calculus I, Calculus II, Math for Life Sciences
  - Pre-post testing

- Spring 2017 administration (n ~ 200)
  - Calculus I, Calculus II, Math for Life Sciences
  - Pre-post testing
Validity Analysis

- Item Response Theory (IRT): family of latent trait models used to establish psychometric properties of items and scales
- Rasch Modeling: Demonstrate relationship between item difficulty and person ability
  - Probabilistic unidimensional model
    - Easier test question, higher likelihood of a correct response
    - More capable the student, higher likelihood of getting questions correct versus less capable (or able) student
  - Assumes probability a student correctly answers a question is a logistic function of the difference between the student’s ability and the difficulty of a question
- PROVIDES A LOT OF INFORMATION ABOUT TEST QUESTIONS TO HELP EXAMINE THE QUALITY OF TEST CONSTRUCTION
Example distribution of QBCI responses

Correct responses are in blue

Pre-administration of the QBCI

Post-administration of the QBCI
Item discrimination by item difficulty for QBCI
Item difficulty by groups of overall test performance

- **Low performers** (<=7 correct)
- **Mid-Low Performers** (8 or 9 correct)
- **Mid-High performers** (10 and 11 correct)
- **High performers** (>= 12 correct)
There is some, but not much, evidence that quantitative education of life science students can be enhanced by incorporating biological examples and data in quantitative courses. *Many (likely fundable) projects on this remain to be done by education researchers.*

There is some, but not much, evidence that quantitative education of life science students can be enhanced by incorporating quantitative ideas in biology courses. *Many (likely fundable) projects on this remain to be done by education researchers.*

STEMeval.org
References and Further Reading

• Quantitative Biology

• Concept Inventories

• Test development

• Rasch analysis