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

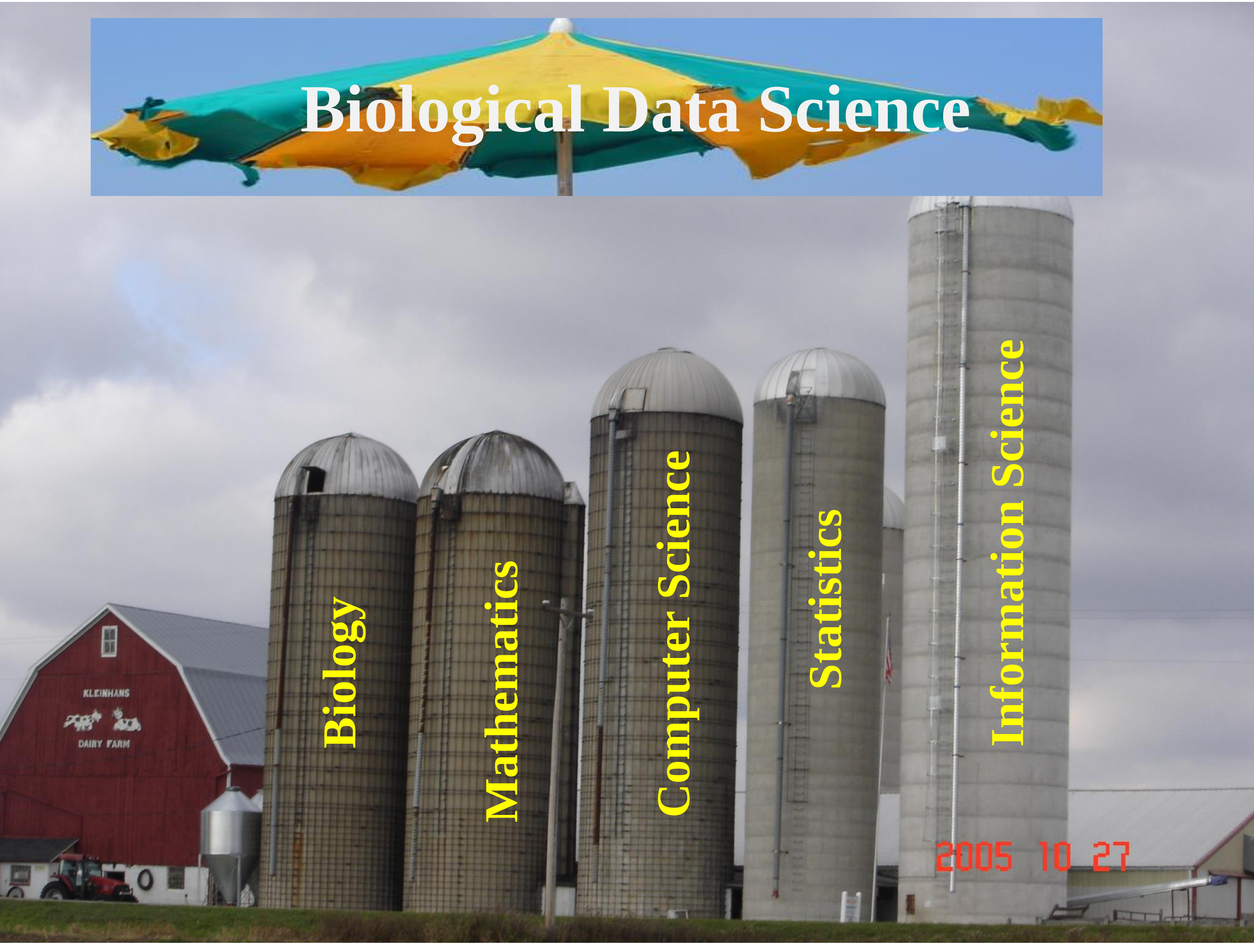
NIMBioS
National Institute for Mathematical and Biological Synthesis

NISER
NATIONAL INSTITUTE FOR STEM EVALUATION AND RESEARCH

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


Biological Data Science
Biology
Mathematics
Computer Science
Statistics
Information Science
KLEINHANS
DAIRY FARM
2005 10 27

Dynamical systems

Optimization

Networks

Graph theory

Calculus

Modeling

Linear algebra

Stochastics

Inference

Distributions

Dynamical systems

Optimization

Networks

Likelihood

Graph theory

Calculus

Hypothesis testing

Regression

Modeling

ANOVA

Linear algebra

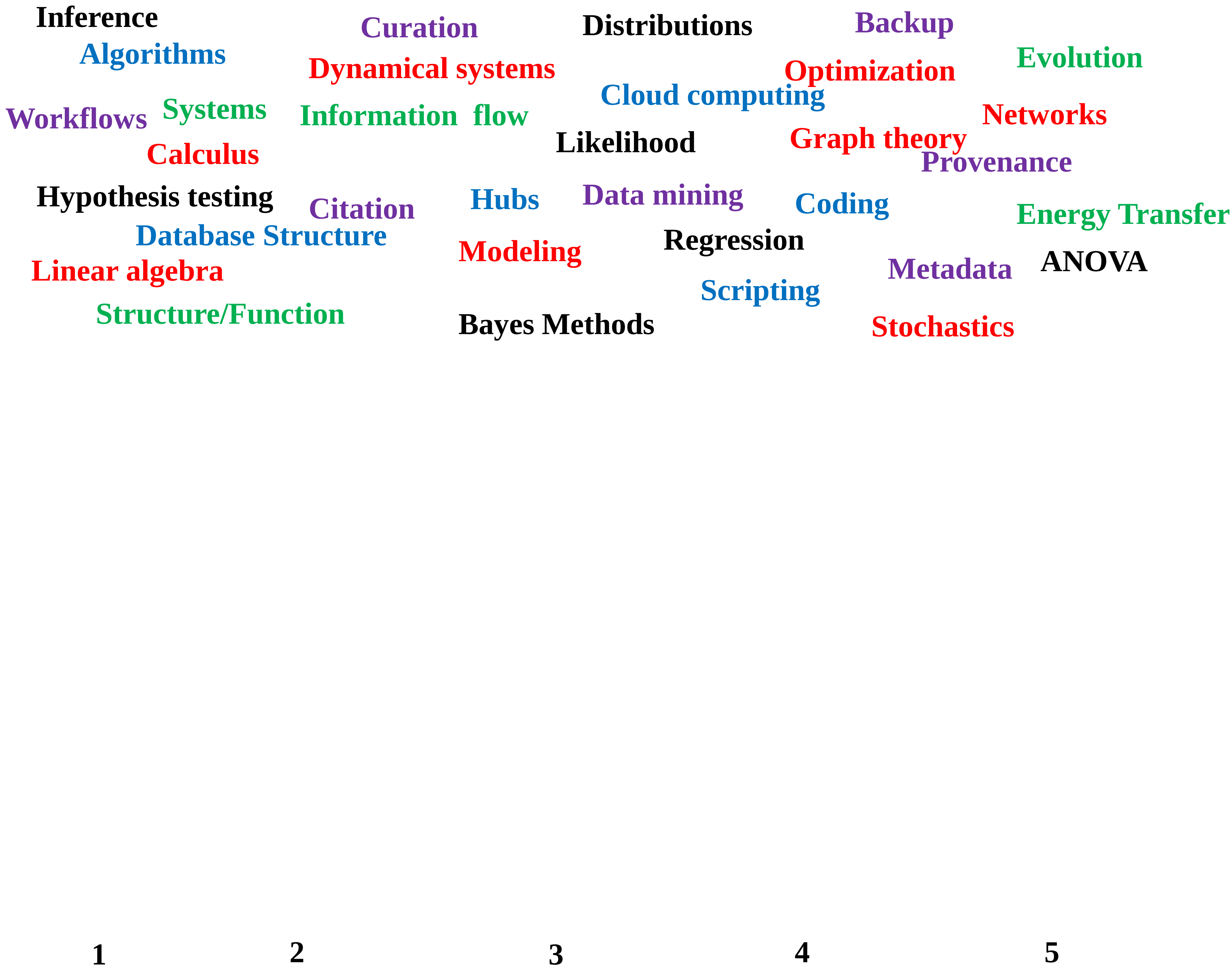
Bayes Methods

Stochastics

Inference
Distributions
Algorithms
Dynamical systems
Optimization
Cloud Computing
Networks
Likelihood
Graph theory
Calculus
Hypothesis testing
Hubs
Coding
Database Structure
Regression
Modeling
ANOVA
Linear algebra
Scripting
Bayes Methods
Stochastics

Inference
Curation
Distributions
Backup
Algorithms
Dynamical systems
Optimization
Cloud computing
Workflows
Networks
Likelihood
Graph theory
Calculus
Provenance
Hypothesis testing
Data mining
Hubs
Citation
Coding
Database Structure
Regression
Modeling
ANOVA
Linear algebra
Metadata
Scripting
Bayes Methods
Stochastics

Inference
Curation
Distributions
Backup
Algorithms
Dynamical systems
Optimization
Evolution
Cloud computing
Workflows
Systems
Information flow
Networks
Likelihood
Graph theory
Calculus
Provenance
Hypothesis testing
Citation
Hubs
Data mining
Coding
Energy transfer
Database Structure
Modeling
Regression
Linear algebra
Metadata
ANOVA
Scripting
Structure/Function
Bayes Methods
Stochastics



Inference
Curation
Distributions
Backup
Algorithms
Dynamical systems
Optimization
Evolution
Cloud computing
Workflows
Systems
Information flow
Networks
Likelihood
Graph theory
Calculus
Provenance
Hypothesis testing
Citation
Hubs
Data mining
Coding
Energy Transfer
Database Structure
Regression
Modeling
ANOVA
Linear algebra
Metadata
Scripting
Structure/Function
Bayes Methods
Stochastics
1
2
3
4
5

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)

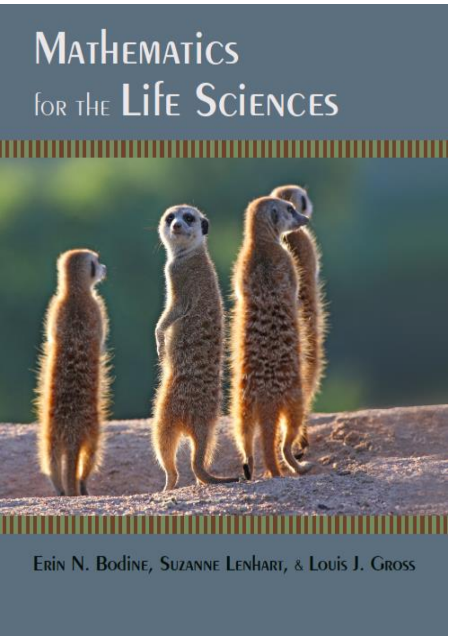
NIMBioS
National Institute for Mathematical and Biological Synthesis

NISER
NATIONAL INSTITUTE FOR STEM EVALUATION AND RESEARCH

Mathematics for the Life Sciences – Princeton U. Press

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.



Erin Bodine

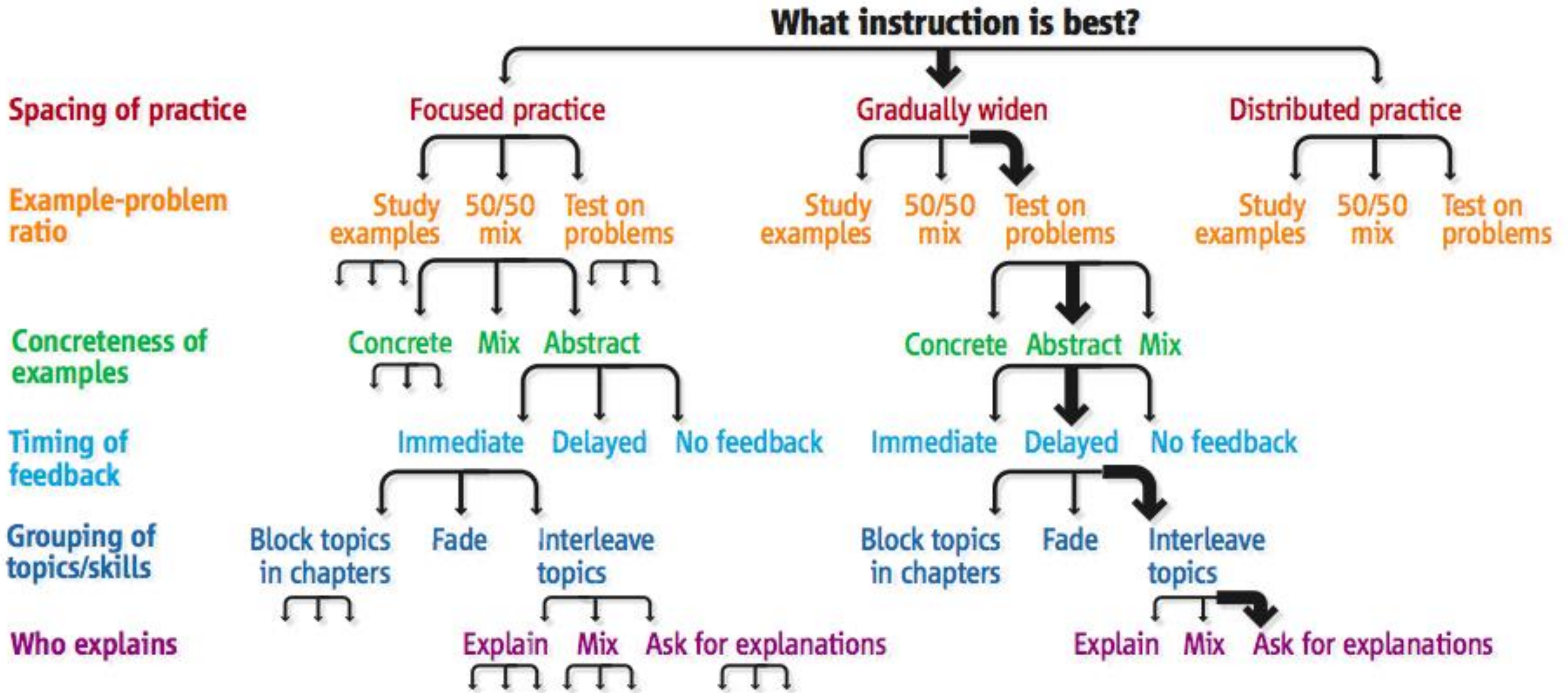
Suzanne Lenhart

NIMBioS
National Institute for Mathematical and Biological Synthesis

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?

NIMBioS
National Institute for Mathematical and Biological Synthesis

NISER
NATIONAL INSTITUTE FOR STEM EVALUATION AND RESEARCH



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.

Koedinger, K. R., J. L. Booth, and D. Klahr1 Instructional Complexity and the Science to Constrain It. *Science* 342:935-7 (2013)

NIMBioS National Institute for Mathematical and Biological Synthesis

NISER National Institute for STEM Evaluation and Research

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)

NIMBioS
National Institute for Mathematical and Biological Synthesis

NISER
NATIONAL INSTITUTE FOR STEM EVALUATION AND RESEARCH

Quantitative Biology Concept Inventory Construction

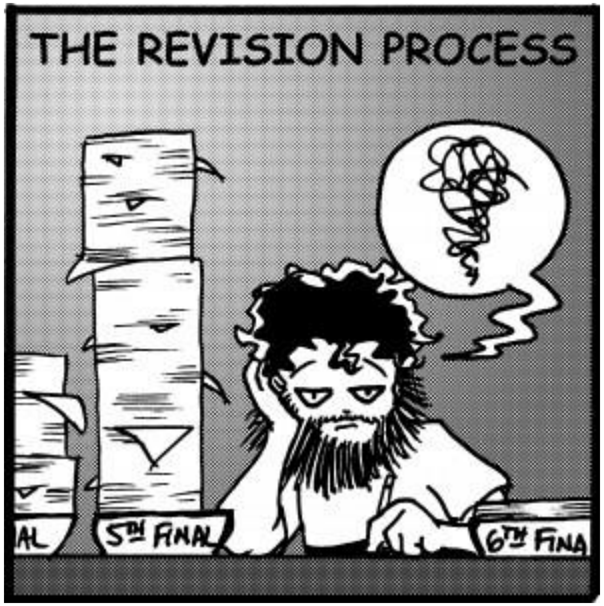
Objectives/ Planning

Content Specification

Item Development

Test administration

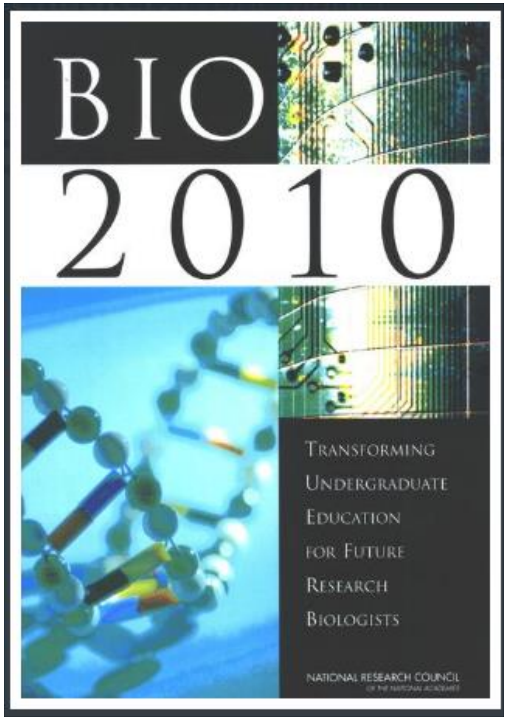
Validity Analysis

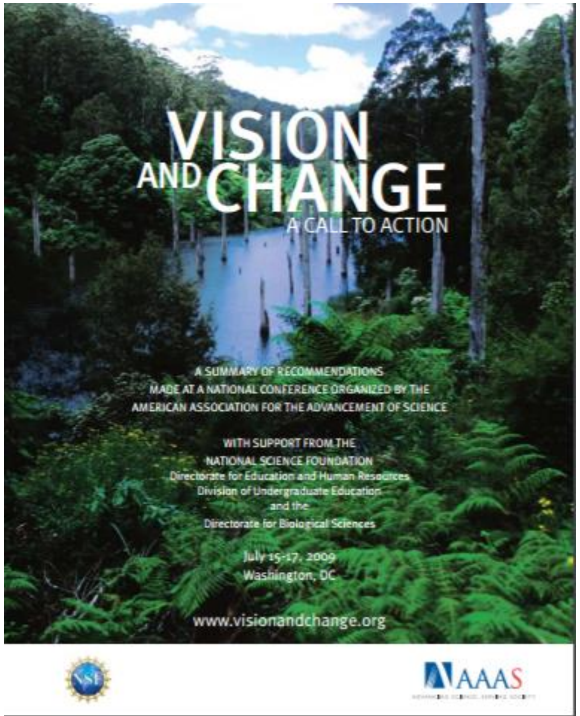


NIMBioS
National Institute for Mathematical and Biological Synthesis

NISER

Objectives/Planning





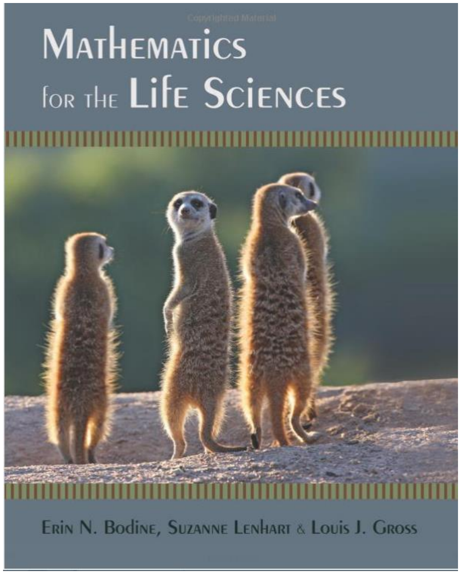
Literature review of concept inventories

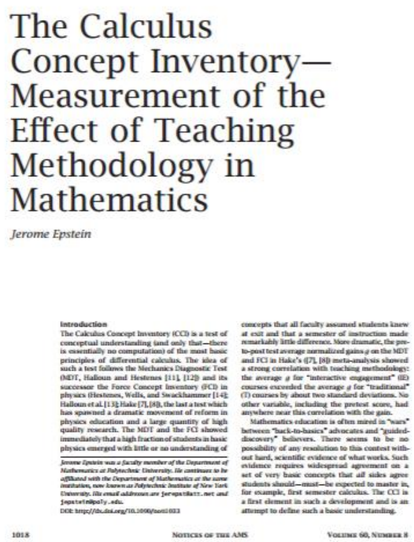
NIMBioS
National Institute for Mathematical and Biological Synthesis

NISER
NATIONAL INSTITUTE FOR STEM EVALUATION AND RESEARCH

Content Specification

- Calculus
  - Rates of change
  - Sums and Integration
  - Modeling
  - Interpreting Graphs and Data





The Calculus Concept Inventory—Measurement of the Effect of Teaching Methodology in Mathematics

Jerome Epstein

Introduction

The Calculus Concept Inventory (CCI) is a test of conceptual understanding (and only that—there is essentially no computation) of the most basic principles of differential calculus. The idea of such a test follows the Mechanics Diagnostic Test (MDT, Halloun and Hestenes [11], [12]) and its successor the Force Concept Inventory (FCI) in physics (Hestenes, Wells, and Swackhammer [14]; Halloun et al. [13]; Hake [7], [8]), the last a test which has spawned a dramatic movement of reform in physics education and a large quantity of high quality research. The MDT and the FCI showed immediately that a high fraction of students in basic physics emerged with little or no understanding of concepts that all faculty assumed students knew at exit and that a semester of instruction made remarkably little difference. More dramatic, the pre-to-post test average normalized gains g on the MDT and FCI in Hake's ([7], [8]) meta-analysis showed a strong correlation with teaching methodology: the average g for "interactive engagement" (IE) courses exceeded the average g for "traditional" (T) courses by about two standard deviations. No other variable, including the pretest score, had anywhere near this correlation with the gain.

Mathematics education is often mired in "wars" between "back-to-basics" advocates and "guided-discovery" believers. There seems to be no possibility of any resolution to this contest without hard, scientific evidence of what works. Such evidence requires widespread agreement on a set of very basic concepts that all sides agree students should—must—be expected to master in, for example, first semester calculus. The CCI is a first element in such a development and is an attempt to define such a basic understanding.

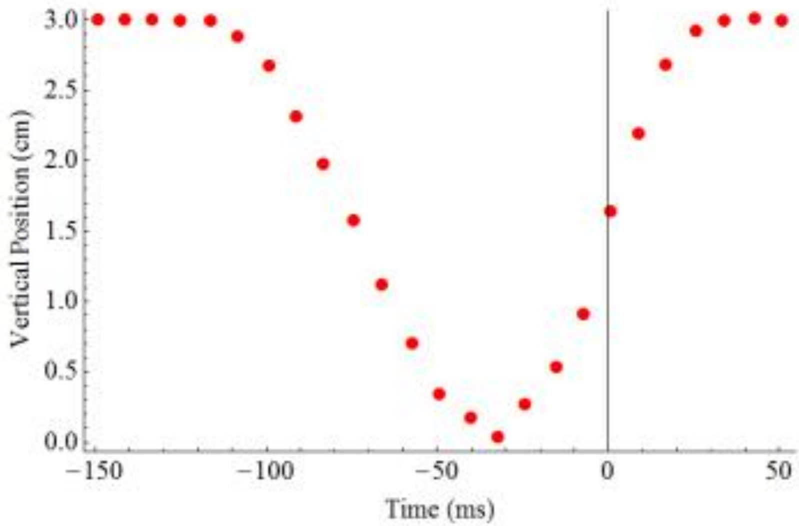
Jerome Epstein was a faculty member of the Department of Mathematics at Polytechnic University. He continues to be affiliated with the Department of Mathematics at the same institution, now known as Polytechnic Institute of New York University. His email addresses are [illegible] *and* [illegible].

DOI: http://dx.doi.org/10.1090/noti1033

1018 Notices of the AMS Volume 60, Number 8

Item Development (Example 1)

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]

NIMBioS National Institute for Mathematical and Biological Synthesis

NISER NATIONAL INSTITUTE FOR STEM EVALUATION AND RESEARCH

Test Review and Administration

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

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

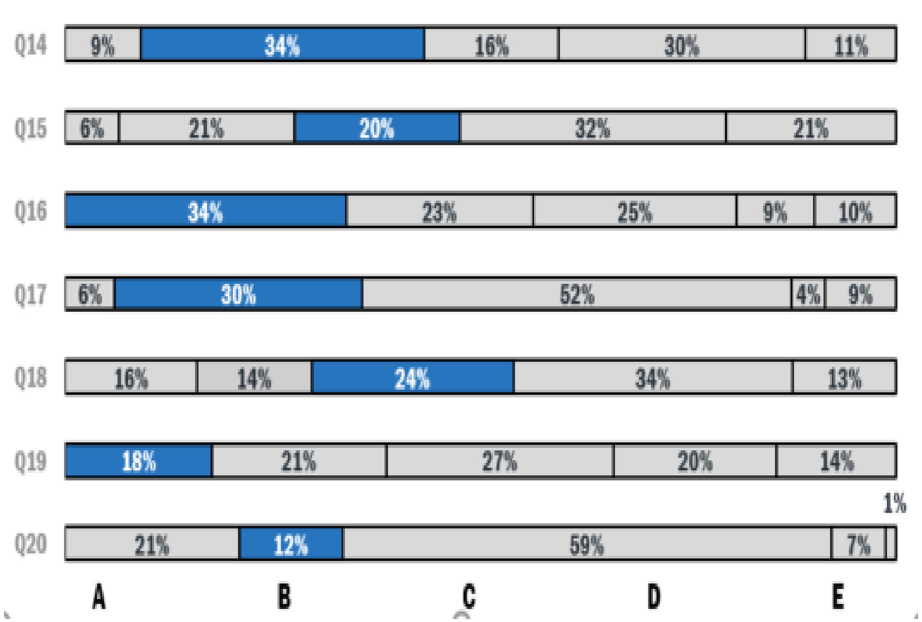
Validity Analysis

- 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**

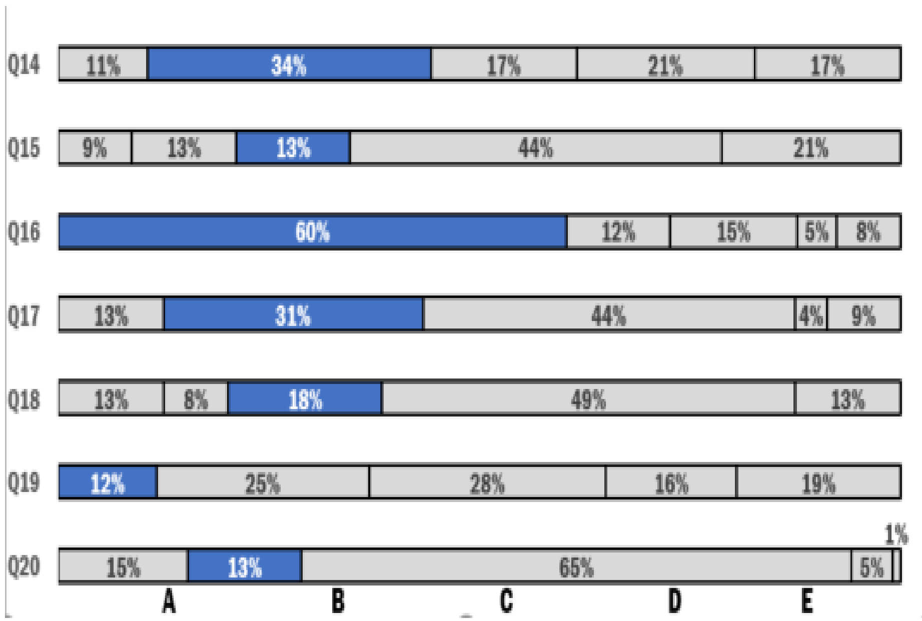
Example distribution of QBCI responses

Correct responses are in blue

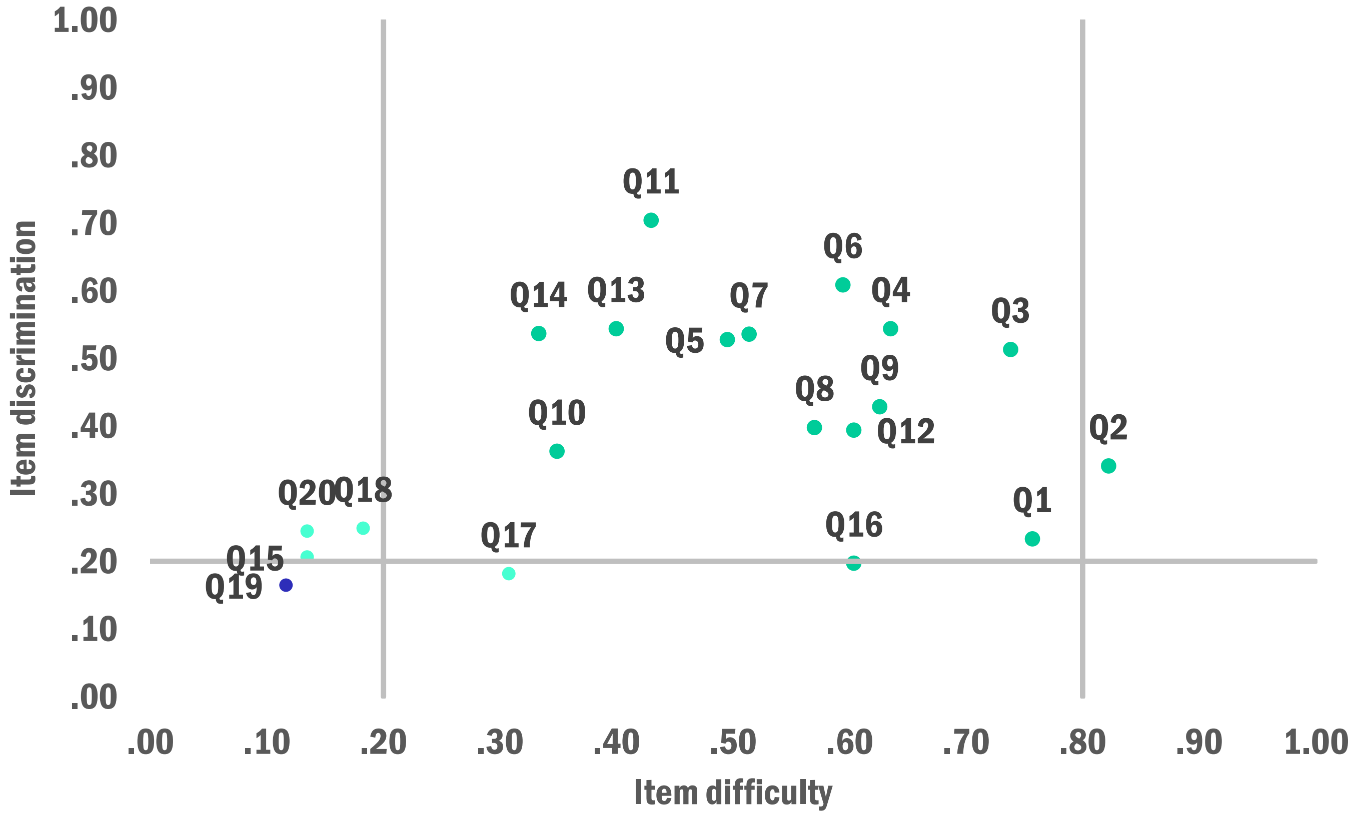
Pre-administration of the QBCI



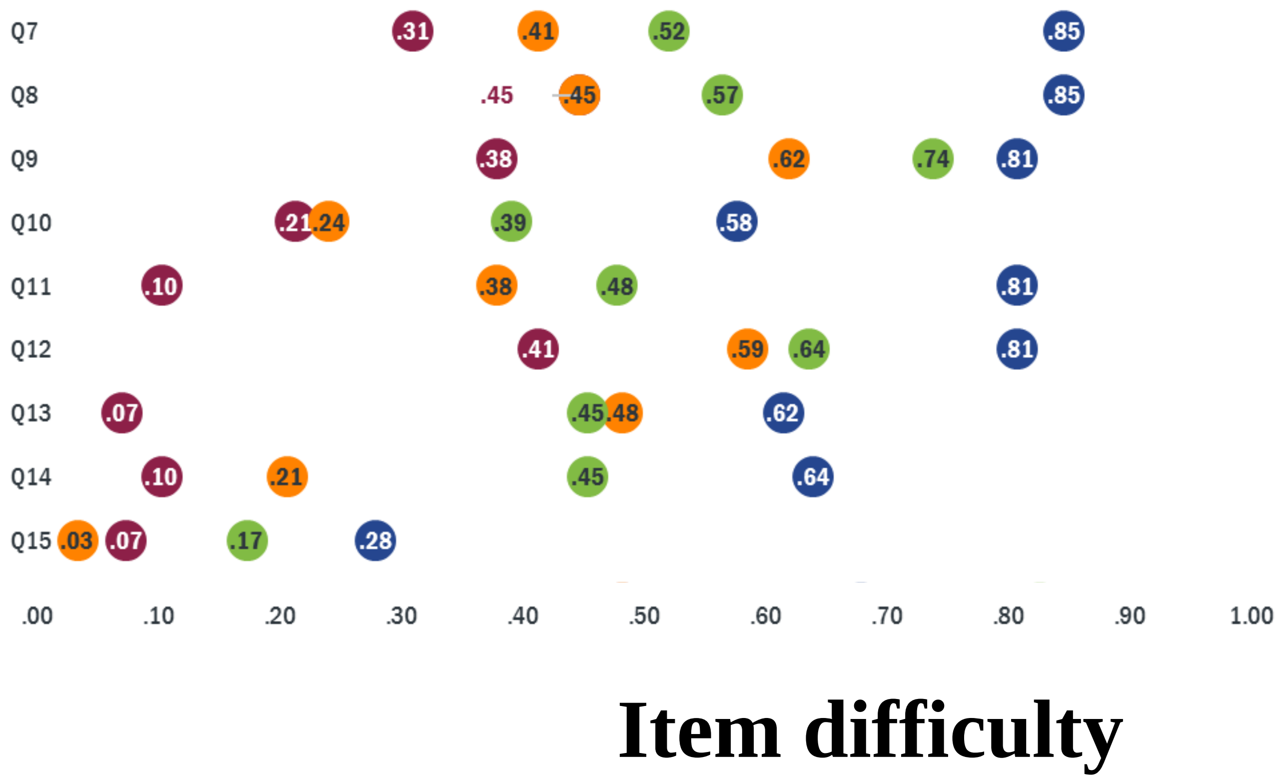
Post-administration of the QBCI



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).

NISER
NATIONAL INSTITUTE FOR STEM EVALUATION AND RESEARCH

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

NIMBioS
National Institute for Mathematical and Biological Synthesis

References and Further Reading

- Quantitative Biology
  - National Research Council, *BIO2010: Transforming Undergraduate Education for Future Research Biologists*. 2003, Washington, DC: The National Academies Press.
  - Brewer, C.A. and D. Smith, *Vision and change in undergraduate biology education: a call to action.* American Association for the Advancement of Science, Washington, DC, 2011.
  - Bodine, E.N., S. Lenhart, and L.J. Gross, *Mathematics for the Life Sciences*. 2014: Princeton University Press.
- Concept Inventories
  - Epstein, J. (2006). The calculus concept inventory. *National STEM Assessment, Washington, DC*, 60-67.
  - Jarrett, L., Ferry, B., & Takacs, G. (2012). Development and validation of a concept inventory for introductory-level climate change science.
- Test development
  - Downing, S. M., & Haladyna, T. M. (2006). *Handbook of test development*. Lawrence Erlbaum Associates Publishers.
- Rasch analysis
  - Boone, W. J., Staver, J. R. & Yale, M. S. (2014). Rasch analysis in the human sciences. Springer, Dordrecht, Netherlands.
  - TEAMS Project Webinars: Diagnosing Your Survey Using Rasch Modeling (Level 1/2): Karen Drill and Erin Stack: http://teams.mspnet.org/