Automated Analysis Of Student Writing Reveals Student Thinking: An Innovative Assessment Methodology Built On Community Goals

Mark Urban-Lurain

Automated Analysis of Constructed Response (AACR) Research Group

http://aacr.crcstl.msu.edu

Division of Science and Mathematics Education

Michigan State University

241st ACS National Meeting & Exposition

Division of Chemical Education

NSF-Catalyzed Innovations

in the Undergraduate Curriculum

March 30, 2011

Acknowledgements

- National Science Foundation DUE-0736952, DUE-0243126, EHR-0314866, DUE-1022653
- Carnegie Corporation Grant #B7458
- The Vice Provost for Libraries, Computing and Technology, Michigan State University

Forging a National Network for Innovative Assessment Methods

Michigan State University
- Joseph Dauer
- Olga Eremina
- Emma Giese
- Laurissa Gulich
- Kevin Haudek
- Merle Heidemann
- Shauna Jones
- Jennifer Kaplan
- Julie C Libarkin (co-PI)
- Tammy Long (co-PI)
- Casey Lyons
- John Merrill (co-PI)
- Rosa Anna Moscarella
- Alan Munn
- Joyce Parker
- Brittany Shaffer
- Duncan Sibley
- Elena Bray Speth
- Mark Urban-Lurain (PI)
- Emily Geraghty Ward

The Ohio State University
- Ross Nehm (PI)
- Judy Ridgway (co-PI)
- Hendrick Haertig
- Minsu Ha

University of Colorado - Boulder
- Jennifer Knight (PI)

University of Washington
- Michelle Smith (co-PI)

Grand Valley State University
- Brittany Shaffer

Western Michigan University
- Mary Anne Sydlik (Evaluator)

CCLI Type II (NSF DUE-1022653)

Overview

- Background
- Example
- Collaborators’ prior work
- Project goals and directions
- Invitation to participate

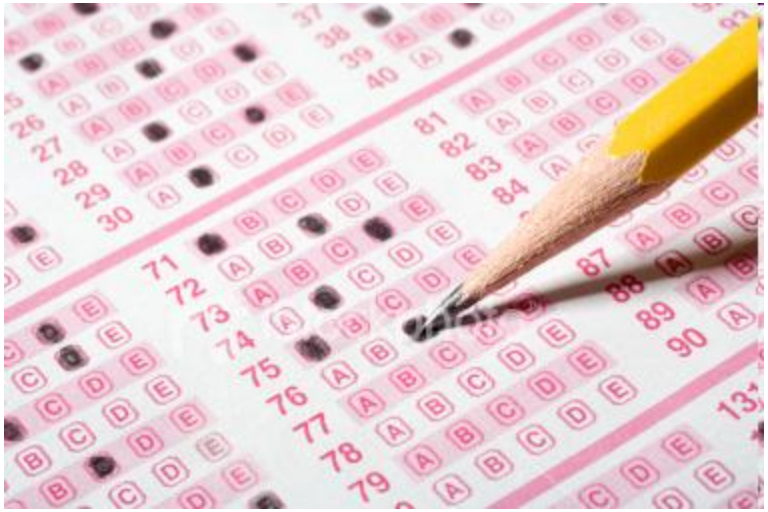
STEM Education Reform

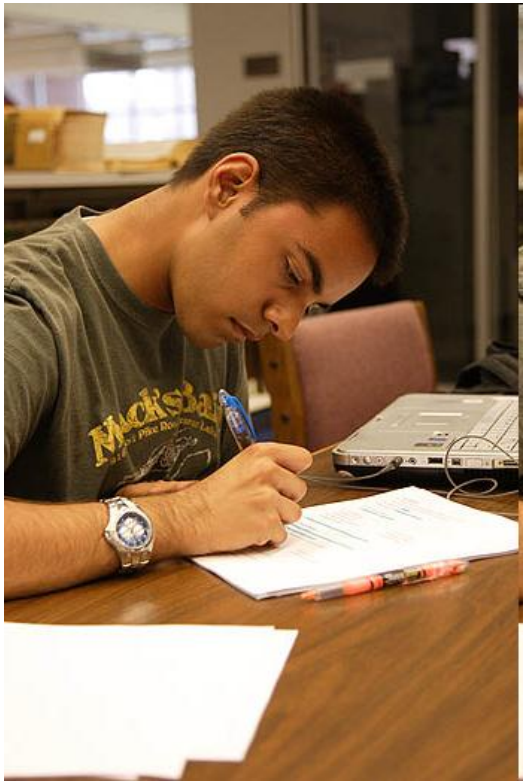
At all levels, science education needs to be redefined, with much less emphasis on the memorization of science facts and terms.

Closely related changes in the introductory science courses in college, emphasizing "science as a way of knowing," are the key to driving these reforms.

Science and the World's Future, Bruce Alberts, MSU STEM Education Symosium

Assessment to Reveal Student Thinking





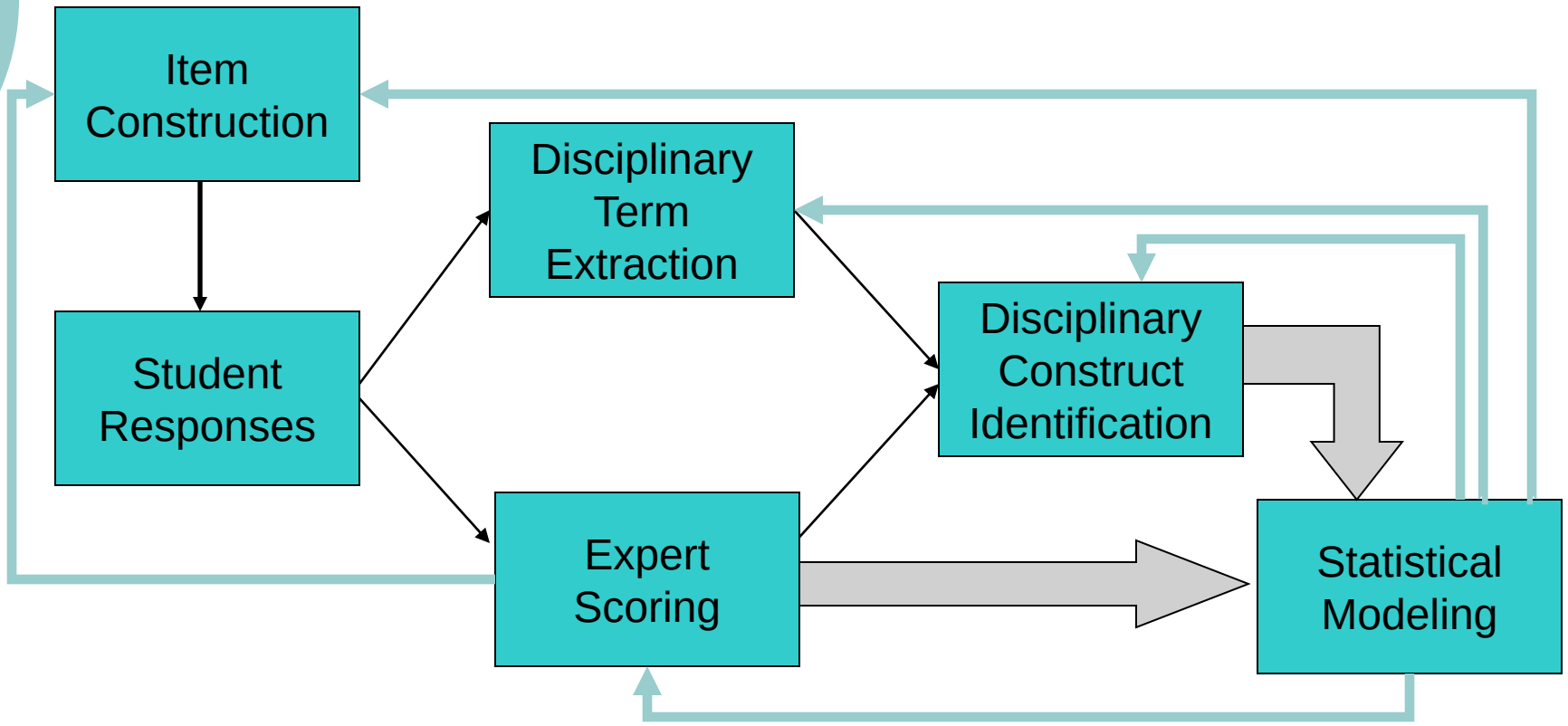
Theoretic Framework

- Conceptual Change
  - Role of prior knowledge in learning
    - Vosniadou, S., (2008)
- Conceptual barriers impair students' understanding complex processes in science
    - Pellegrino, J.W., Chudowsky, N., and Glaser, R. (2001); Tanner, K., and Allen, D. (2005)
- Student ideas
  - May be identified by students' use of language
  - *Constructed Response* questions can provide insight into student ideas
    - Bennett and Ward (1993); Birenbaum and Tatsouka (1987); Bridgeman (1992); Kuechler and Simkin (2010)

AACR Objectives

- Evaluate students' understanding of scientific concepts
  - Create models of student thinking
- Use linguistic and statistical analysis to analyze students' writing
  - Develop necessary libraries and resources
  - Validate by predicting expert ratings

Our Approach: Linguistic Feature-Based



Example: Chemistry of Biology

- Evaluate students' understanding of basic chemistry related to cellular and molecular biology
  - Free energy and acid/base chemistry
- Introductory Biology Cells and Molecules (BS111)
  - Large enrollment (400-500 / section)
  - General chemistry prerequisite

Functional Groups: Multiple Choice

Consider two small organic molecules in the cytoplasm of a cell, one with a hydroxyl group (-OH) and the other with an amino group (-NH_2).

Which of these small molecules (either or both) is most likely to have an impact on the cytoplasmic pH?

35%	A. Compound with amino group
45%	B. Compound with hydroxyl group
7%	C. Both
13%	D. Neither

Explain your answer

Sample Student Answers

The amino group can break down compounds faster and can therefore change the pH of the cytoplasm

Has a carboxyl group, is more acidic

The amino group is more basic and can change the pH better than the hydroxyl group.

The hydroxyl group doesn't affect the pH as much as an amino, which has a NH2.

The level of Hydrogen concentration defines the pH.

The amino group is an acid. It will cause the pH in the compound to rise.

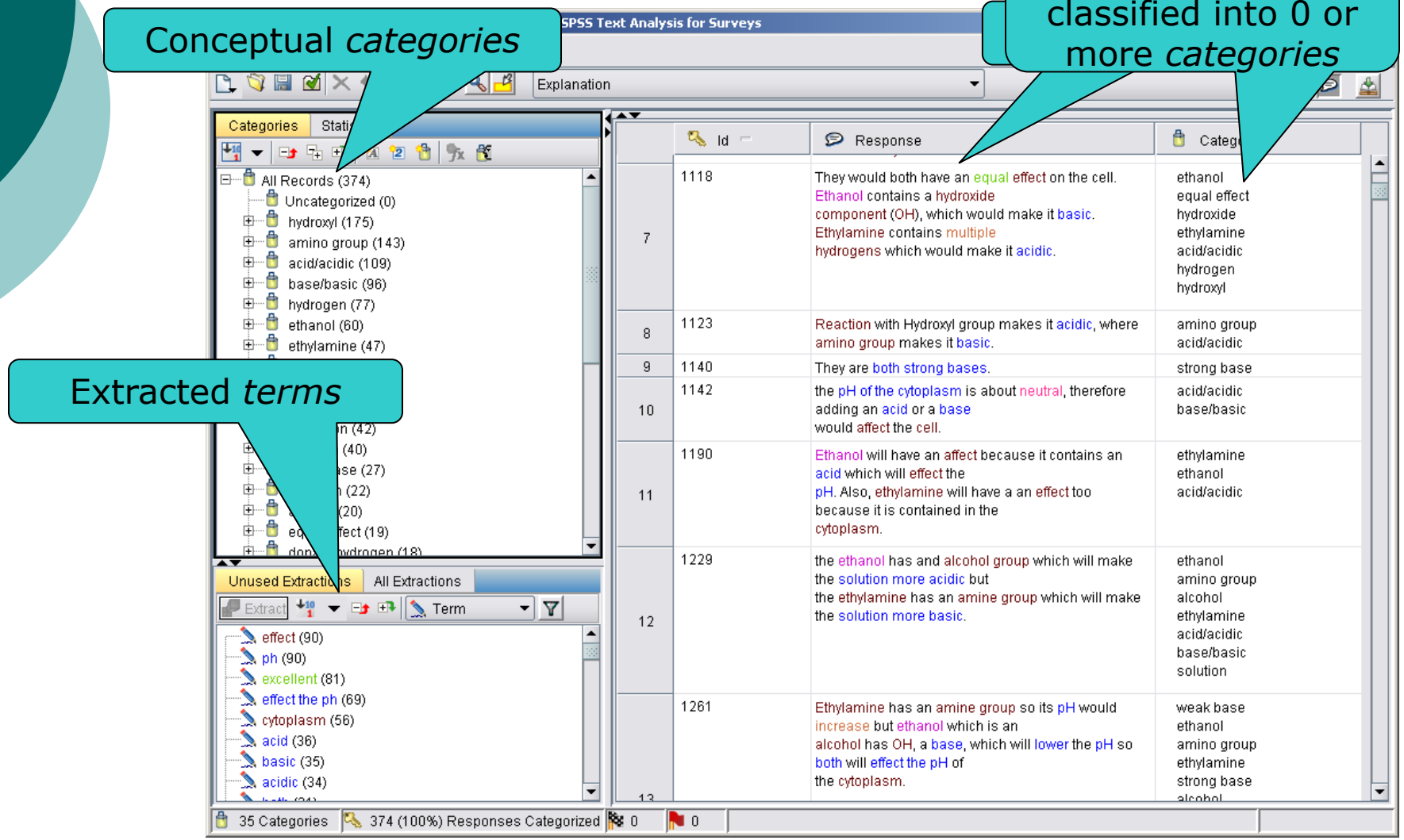
Hydroxyl is a base.

Expert Ratings of Explanations

- Two experts rated explanations from correct answers using 3-level rubric
  - 36% Level 1: Correct explanations of functional group chemistry (may include correct supporting reasoning)
  - 12% Level 2: Partly correct explanations with errors in facts or reasoning
  - 51% Level 3: Totally incorrect/irrelevant response

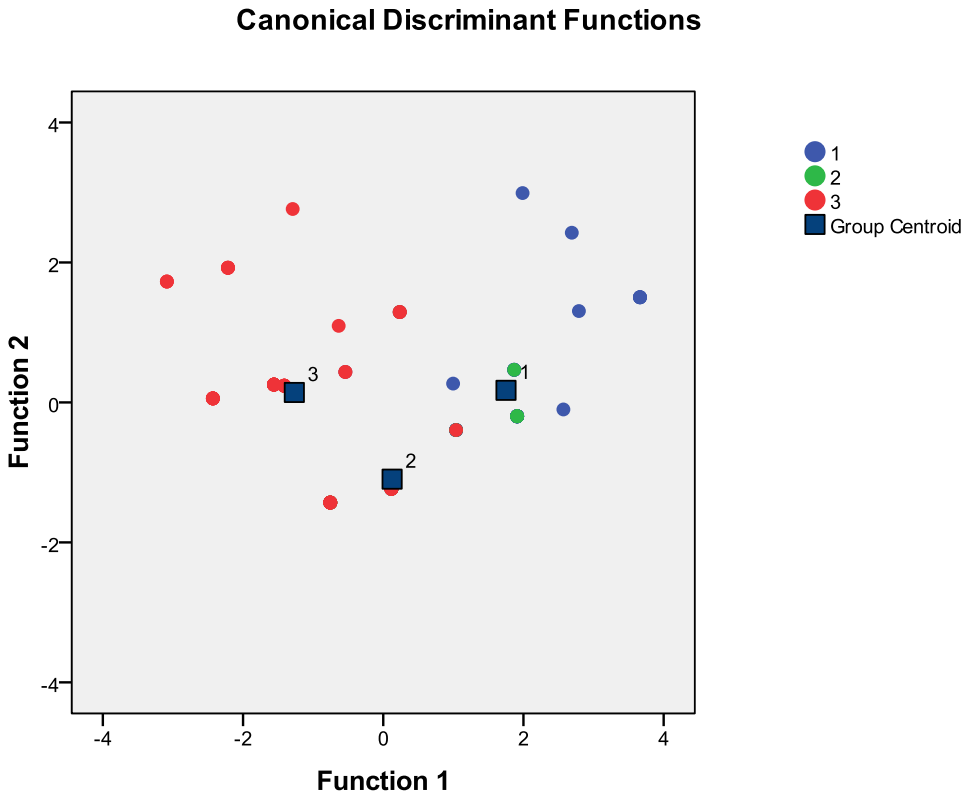
Inter-rater reliability = .92

SPSS Text Analysis for Surveys

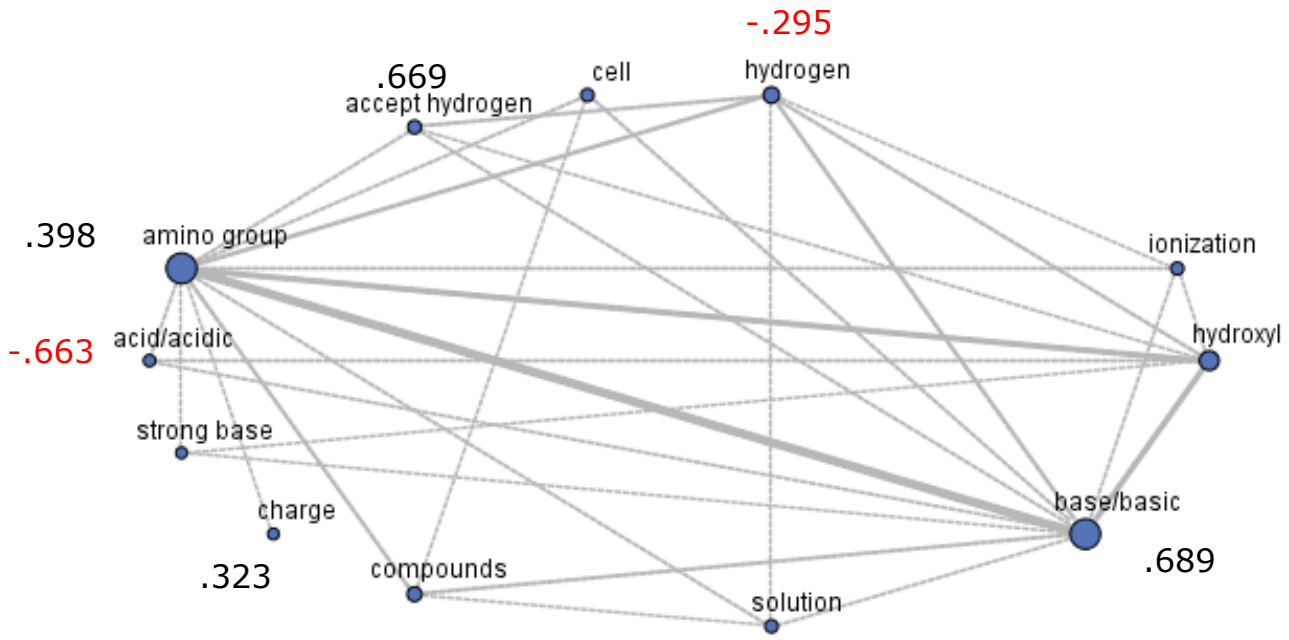


Statistical Classification: Discriminant Analysis

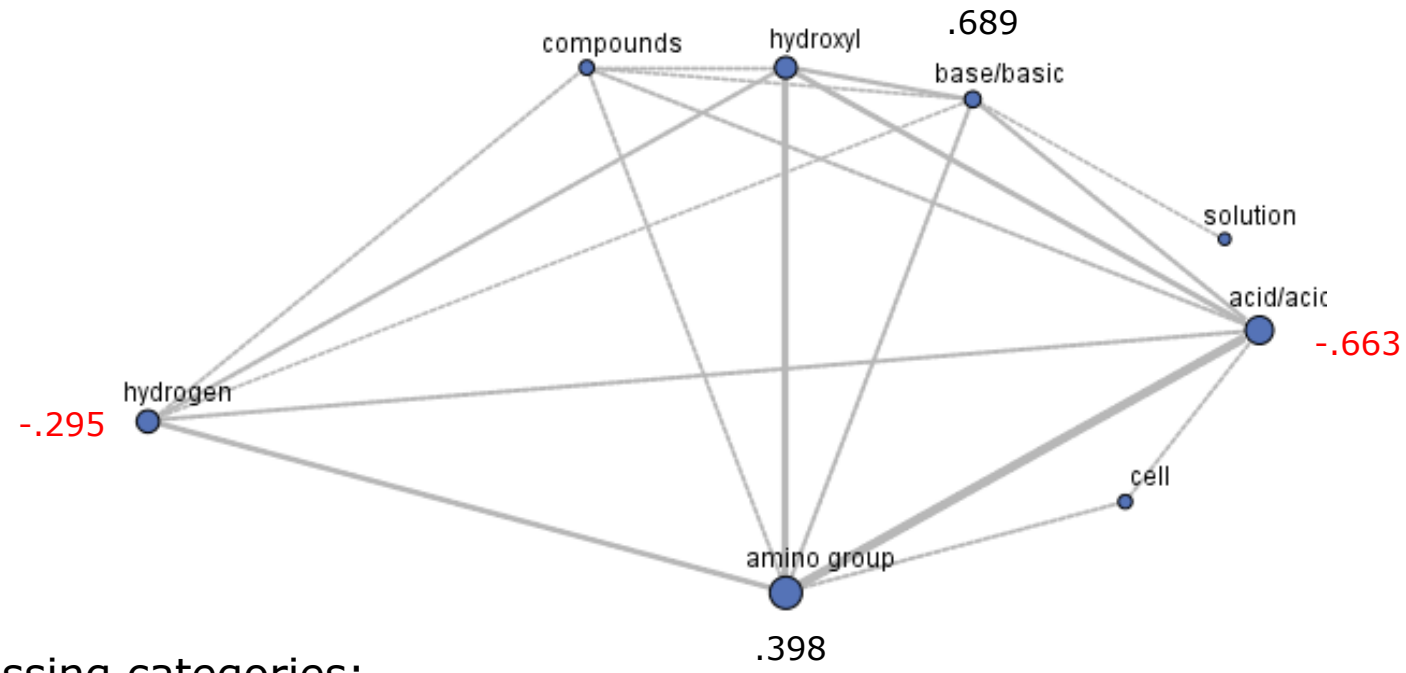
	Function	
Category	1	2
base/basic	.689	.398
acid/acidic	-.663	.590
amino group	.398	.090
hydrogen	-.295	.751
charge	.323	.149
accept hydrogen	.669	.009



Complexity of Student Ideas Expert Rated 1



Complexity of Student Ideas Expert Rated 3



Missing categories:

Charge

Accept hydrogen

Predicting Expert Ratings

	Computer Predicted Rating		
Expert Rating	1	2	3
1	**82.9**	12.2	4.9
2	21.4	**42.9**	35.7
3	6.9	12.1	**81.0**

- 77% of the cases scored correctly, $p < .001$
- Expert/computer inter-rater reliability Intraclass correlation = 0.835

Automated Analysis of Constructed Response Concept Inventories to Reveal Student Thinking: Forging a National Network for Innovative Assessment Methods

CCLI – II DUE 0736928

Cellular Metabolism (Urban-Lurain & Merrill)

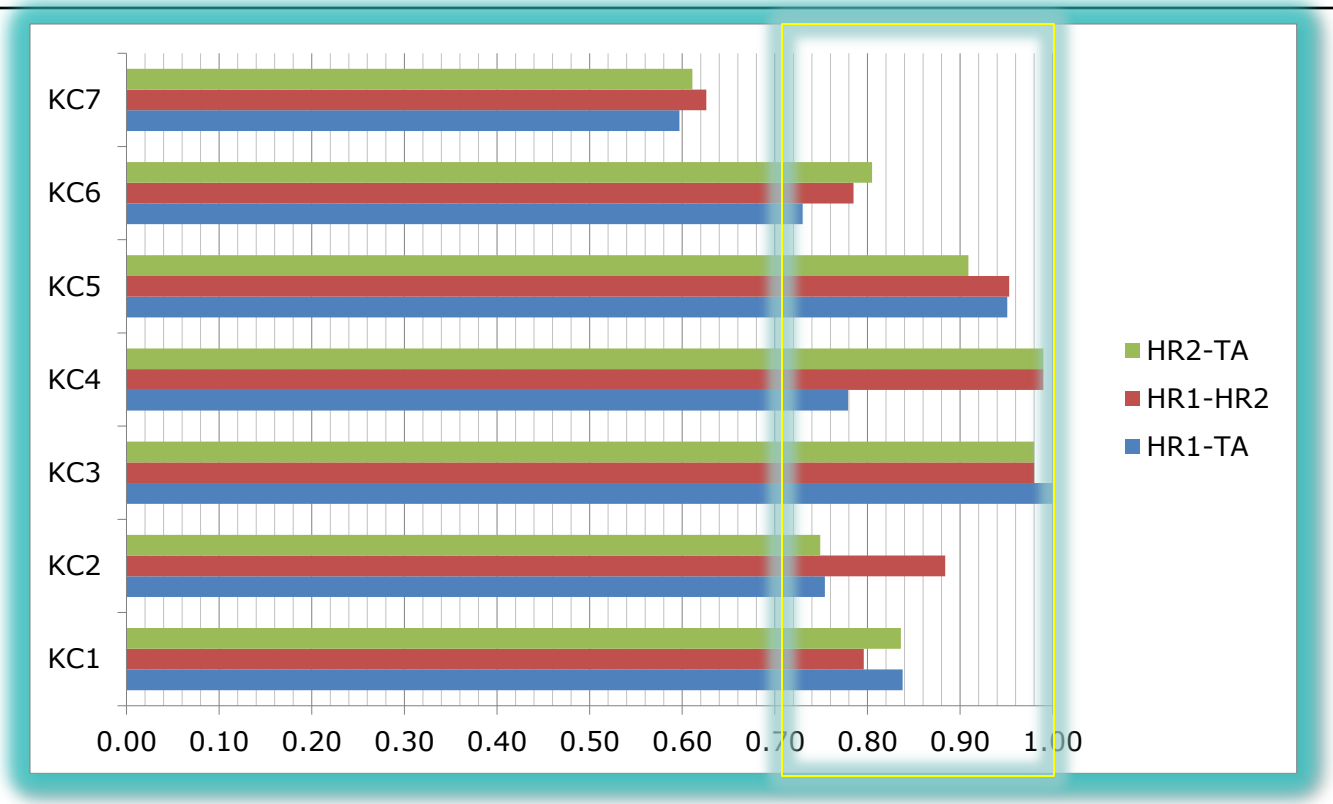
- Foundation: Diagnostic Question Clusters (DQC)
  - DUE-ASA 0243126 and CCLI 0736947
- Trace matter and energy in biological systems
- Understanding organismal ecological phenomena in terms of cellular metabolism

Evolution and Natural Selection (Nehm, Ha & Haertig)

- Open Response Instrument (ORI) and Evolutionary Gain and Loss Test (EGALT)
- Construct-grounded approach
  - 3 "Core Concepts"
    - Variation
    - Heritability of variation
    - Differential survival/reproduction
  - 4 "Key Concepts" used by experts
    - Biotic potential
    - Natural resources
    - Differential survival
    - Change in population

Evolution and Natural Selection
Human Raters (HR1, HR2) vs. Computer (TA)

Core and key concepts of natural selection



Kappa values

Target

Evolution and Natural Selection Compare Lexical Analysis Approaches

SPSS Text Analytics for Surveys

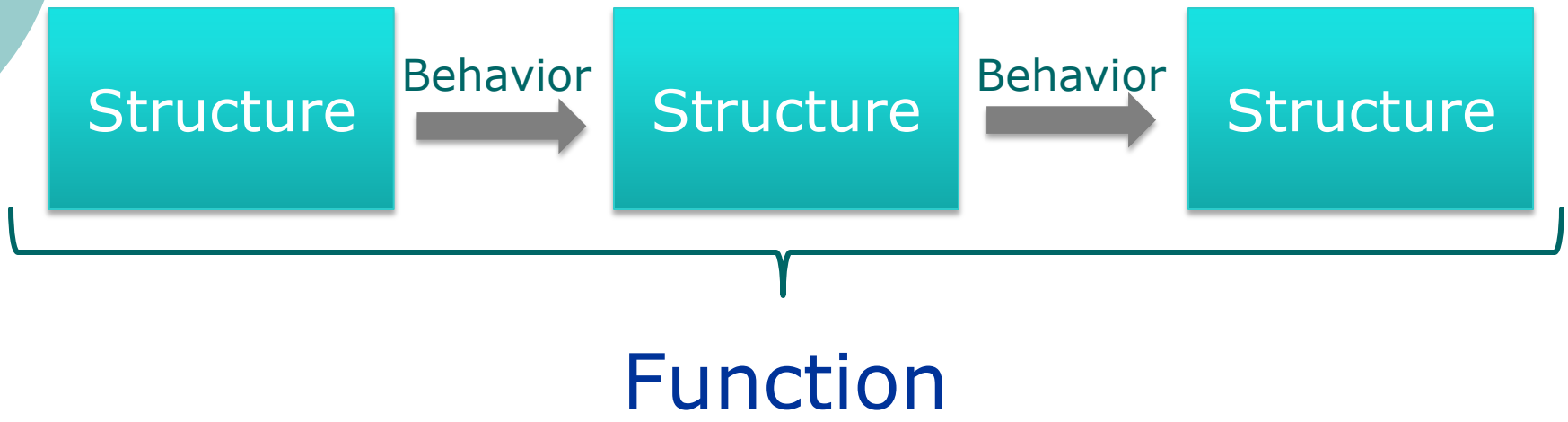
- Developed for open-ended web-based market research
- Supports exploratory, iterative development of lexical resources
- Manual creation of disciplinary libraries

Summarization Integrated Development Environment (SIDE)

- Developed for discourse analysis online discussions
- Machine-learning classification techniques
- “Black box”

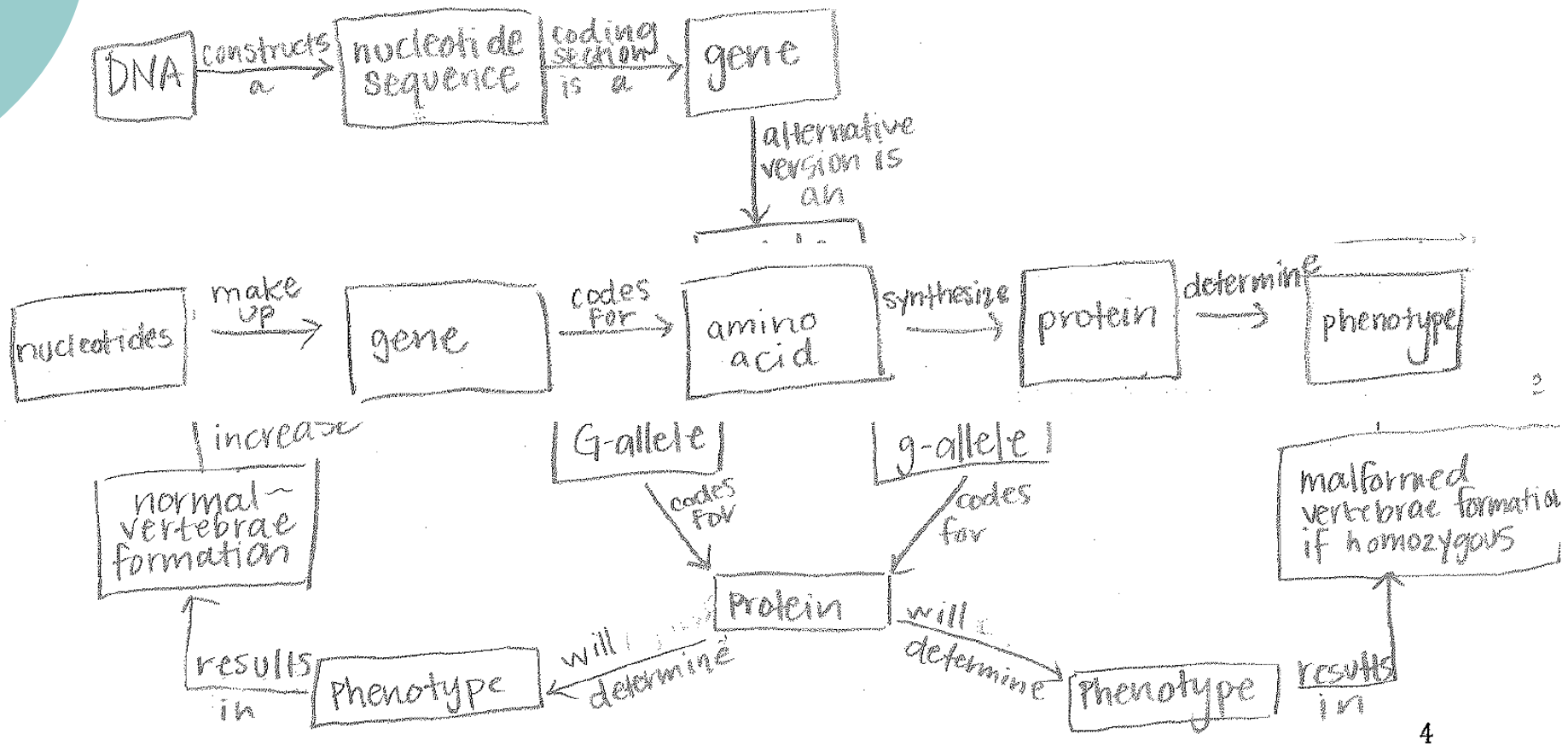
Models in Introductory Biology (Long)

- Structure-Behavior-Function (SBF)



Models in Introductory Biology
Analyze Changes in Student Models

Relationship Between Genotype and Phenotype



Genetics Concept Assessment (GCA) (Smith & Knight)

- Iterative development process
  - Review literature
  - Interview genetics faculty and students to explore misunderstandings.
  - Develop and administer a pilot assessment.
  - Eliminate jargon, write distracters with student-supplied incorrect answers, revise easy questions.
  - Validate and revise through student interviews and input from faculty experts at several institutions.

Genetics Concept Assessment (GCA) Constructed Response Research

- Create constructed response items for persistently difficult topics
  - Nature and consequences of mutations
  - DNA content of cells
  - Allele representation on chromosomes undergoing meiosis and mitosis
- Collecting data
  - University of Washington
  - University of Colorado – Boulder
  - Michigan State University

Genetics Concept Assessment (GCA) Research Questions

- Can lexical analysis be used to accurately score genetics concept assessment questions?
- Will student responses reveal the same persistent misunderstandings if the questions are asked in a short answer format?
- Are there some genetics concepts where multiple-choice and short answer response questions are similarly effective?

Geoscience Concept Inventory (GCI) (Libarkin)

- GCI WebCenter database of 813 student alternative conceptions about Earth Systems
- GCI
  - Evaluate learning in entry-level geoscience courses
  - Correlates strongly with individual expertise in geosciences
  - Rasch analysis to compare large number of items for equivalence

Geoscience Concept Inventory Lexical Analysis Research

- Exploring lexical analysis across diverse items to identify student misconceptions
- Role of lexical analysis in construction, revision, and validation of MC items

Statistics / Lexical Ambiguity (Kaplan)

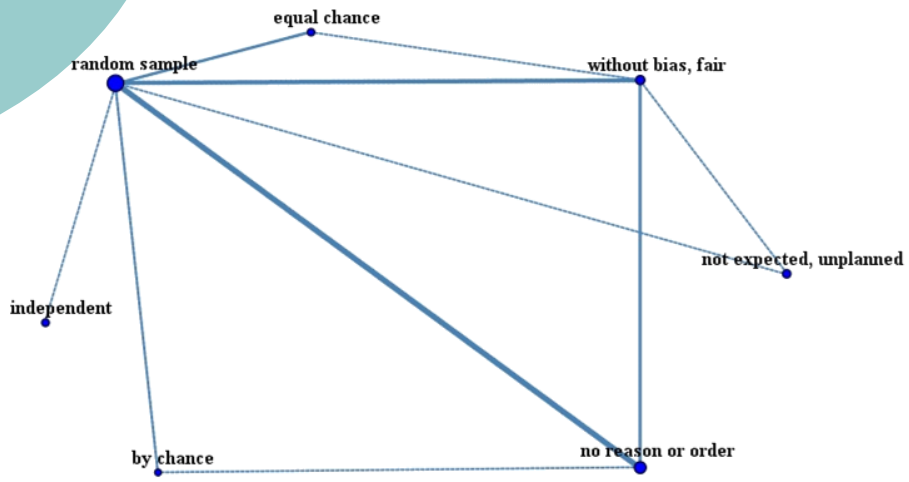
- Lexical Ambiguity
  - Domain-specific words similar to common English words
- Statistics
  - Random
  - Association
  - Correlation
  - Bias
  - Skew
- Barrier to learning, particularly in introductory courses

Lexical Ambiguity
Random

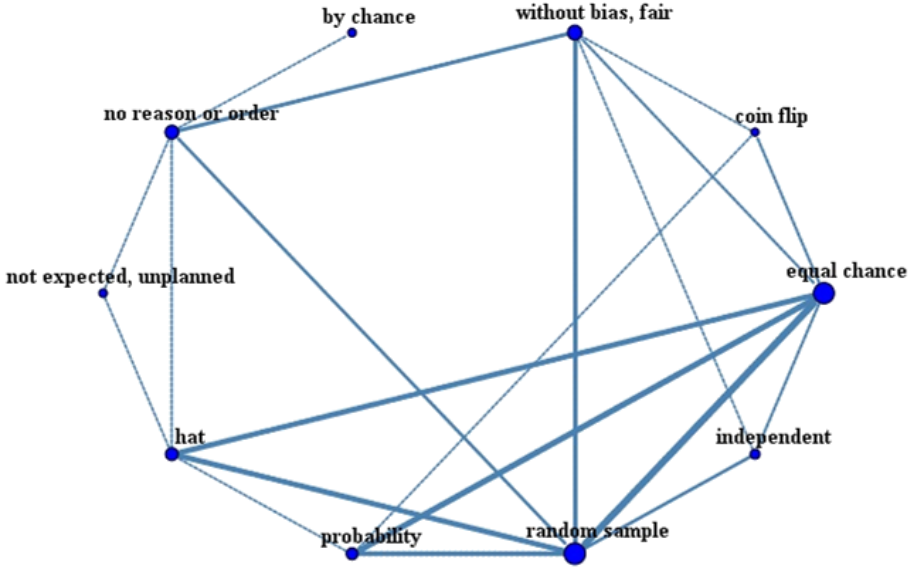
- 49%: An occurrence that is unplanned, unexpected or haphazard
- 17% Without criteria, plan or prior knowledge
- 8% Without pattern
- 4% Without bias

Lexical Ambiguity
Random

Pre-Instruction



Post-Instruction



Project Goals and Direction

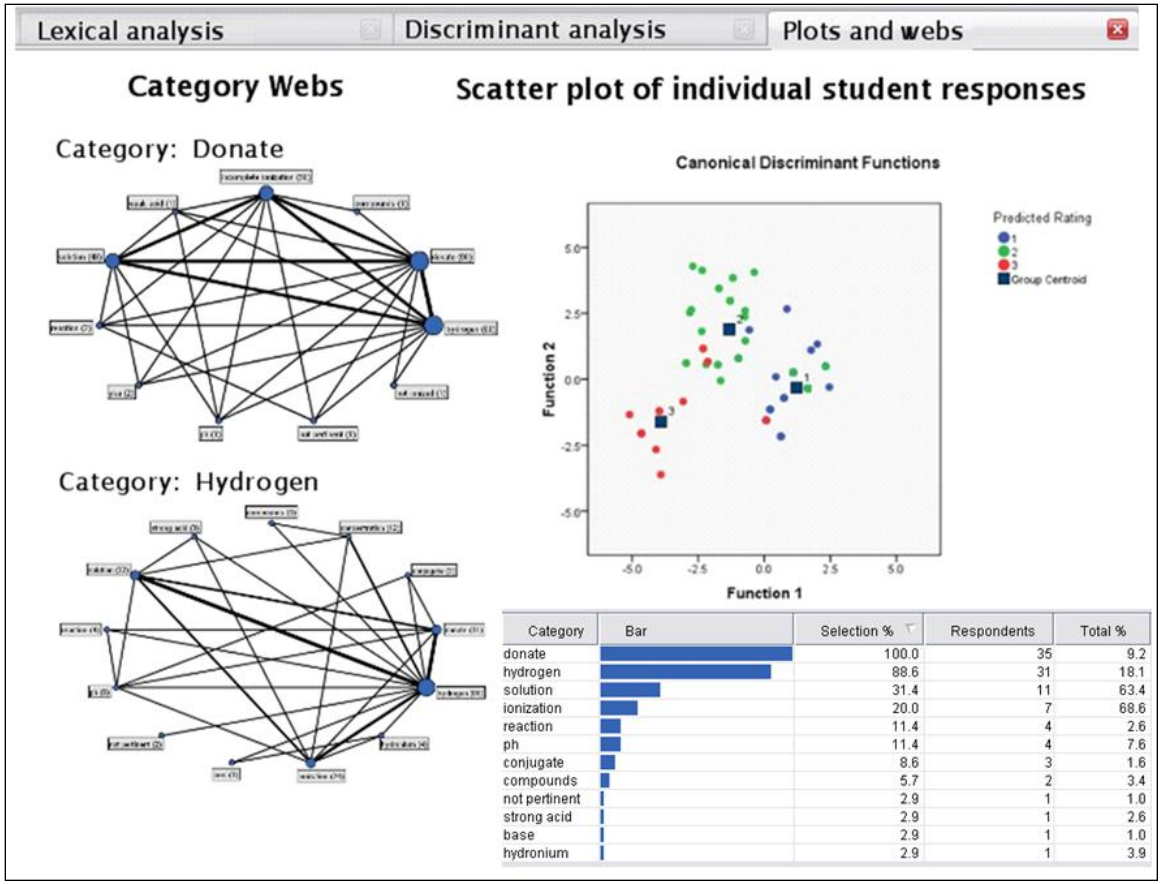
AACR
Building on Community Goals

- Create constructed-response concept inventory questions in each topic
- Create lexical resources for each topic
- Evaluate student responses using expert scoring rubrics
- Develop statistical classification functions to predict expert ratings
- Validate automated analyses
- Disseminate questions and resources
- Build community of researchers and teachers exploring these techniques

AACR
Research Questions

- Are constructed-response items always needed to uncover student thinking?
- Are lexical analysis protocols generalizable?
- What are the relative strengths and weakness of different automated analysis techniques?
- How well do these techniques predict expert scoring?
- How can text analysis inform rubric creation?
- How can linguistics enhance lexical analysis research in STEM fields?

Future Work
Web Portal



Category	Bar	Selection %	Respondents	Total %
donate		100.0	35	9.2
hydrogen		88.6	31	18.1
solution		31.4	11	63.4
ionization		20.0	7	68.6
reaction		11.4	4	2.6
ph		11.4	4	7.6
conjugate		8.6	3	1.6
compounds		5.7	2	3.4
not pertinent		2.9	1	1.0
strong acid		2.9	1	2.6
base		2.9	1	1.0
hydronium		2.9	1	3.9

AACR
Forging a National Network

- Seeking collaborators
  - Pilot items and collect data
  - Develop, evaluate and/or apply scoring rubrics
  - Suggest other concepts, inventories or questions
  - Join online discussions

Seeking Two Post-Doctoral Researchers

- Ph.D. STEM or STEM Education
- Strong commitment to undergraduate STEM education
- Mix of some set of these skills:
  - Publishing and presenting research
  - STEM education research
  - Assessment of learning
  - Computational linguistics
  - Natural language processing
  - Statistics/computer modeling
  - Database management
  - Web site development

Questions

Mark Urban-Lurain

Division of Science and Mathematics Education

Michigan State University

urban@msu.edu

aacr.crcstl.msu.edu