

# Insight into Student Thinking in STEM: Lessons Learned from Lexical Analysis of Student Writing

The Automated Analysis of Constructed Response (AACR) Research Group

Center for Engineering Education Research

Visit us at [aacr.crcstl.msu.edu](http://aacr.crcstl.msu.edu)

## Item Construction

Example from cell metabolism: Chemistry in introductory biology course

*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<sub>2</sub>). Which of these small molecules (neither or both) is most likely to have an impact on the cytoplasmic pH?*

Students made a selection and explained their answer.

## Rubric and Human Scoring

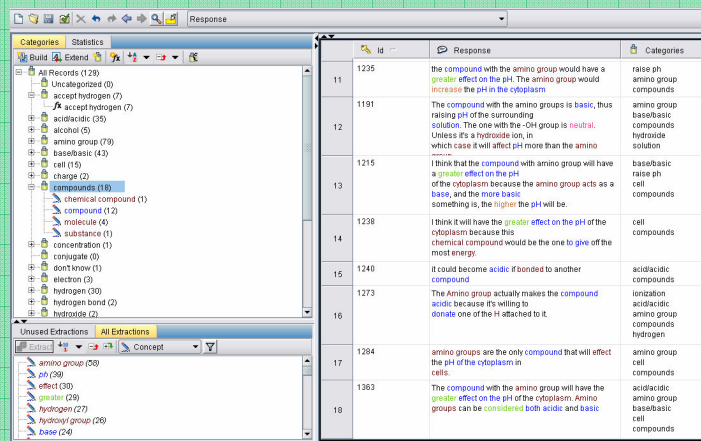
•Two expert scorers rated using 3-level rubric; agreement on 284 out of 323 (interrater reliability=0.92; intraclass correlation)

Level	Number	Rubric	Example
1	91	Totally correct explanation	<i>Amino groups act as a base and pick up a hydrogen ion from its surrounding solution.</i>
2	27	Partially correct explanation	<i>The amino group acts as a base. It will lower the pH of the cytoplasm toward base (8+).</i>
3	166	Totally incorrect or irrelevant explanation	<i>Amine has two H atoms it may give up, but hydroxyl has only one OH molecule it may give up.</i>

## Machine Extraction and Category Building

•Lexical analysis can process large numbers of responses easily

•Software identifies key terms and groups similar terms into categories (i.e. concepts)



IBM SPSS Text Analytics for Surveys software showing the terms extracted (lower left panel), categories (upper left panel), student responses (right panel). Each response is placed into one or more categories (rightmost column).

## Prediction Tool and Machine Scoring

•Computer generates prediction models using lexical categories

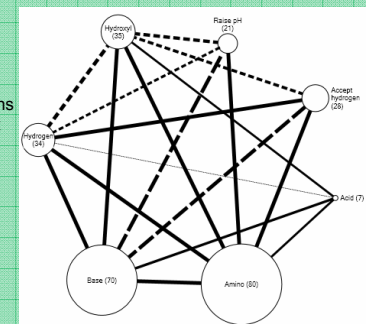
Category	Coefficient
Accept hydrogen	0.604
Acid	-0.433
Amino group	0.2
Base	0.799
Hydrogen	-0.326
Hydroxyl	-0.177
Raise pH	0.228

•Computer-expert interrater reliability =0.84 (intraclass correlation)

## Web Diagrams

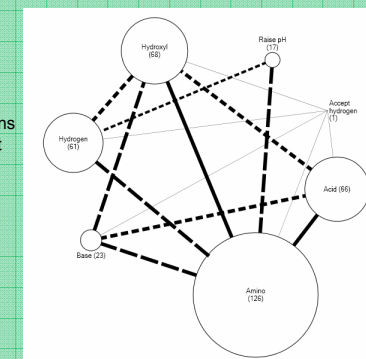
•A visualization of categories from machine extractions and computer prediction of human scoring

Explanations with Expert Rating 1



Solid: — Share 75 – 100 % of responses  
Dashed: - - - Share 50 – 74% of responses  
Dotted: . . . Share 25 – 49% of responses

Explanations with Expert Rating 3



## Acknowledgements

This material is based upon work supported by the National Science Foundation (DUE 0736952 and DUE 1022653). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the supporting agencies.

AACR consists of researchers from multiple universities with backgrounds in various STEM disciplines, linguistics, and educational research. At MSU:

Kevin Haudek John Merrill (co-PI)  
Andrew League Joyce Parker  
Fengjie Li Luanna Prevost  
Julie Libarkin (co-PI) Mark Urban-Lurain (PI)  
Tammy Long (co-PI) Michele Weston

Discipline	Instrument	University
Cell metabolism	Diagnostic Question Clusters	MSU
Evolution	ORI / EGALT / ACORNS	OSU
Genetics	Genetics Concept Assessment	CU-Boulder UMaine
Geoscience	Geoscience Concept Inventory	MSU
Statistics		UGA

**Theoretic Framework:** Conceptual Change (Vosniadou, 2008)

- Conceptual barriers impair students' understanding complex processes in science
- Importance of the role of prior knowledge in learning

## Student ideas

- May be identified by students' use of language (Pinker, 2007)
- Constructed Response questions can provide insight into student ideas

## Objectives

- Evaluate students' understanding of scientific concepts
- Use linguistic and statistical analysis to analyze students' writing (Deane, 2006)
  - Develop necessary libraries and resources
  - Validate by predicting expert ratings
- Model student thinking

