Computerized Lexical Analysis Reveals Complexity of Student Ideas in Biology

John Merrill
Casey Lyons
Shauna Jones
Rosa Moscarella
Kevin Haudek
Mark Urban-Lurain

MICHIGAN STATE UNIVERSITY

CRCSTL
Center for Research on College Science Teaching and Learning

Society for the Advancement of Biology Education Research
SABER

Introduction: Use of constructed responses

- Students have complex ideas
- One limitation of multiple choice questions is the forced selection of a single idea
- Having students create their own explanations may better reveal their complex ideas

Introduction: Conceptual Change

- Students build new ideas upon existing knowledge
- This makes conceptual change difficult in that incorrect ideas are not easily replaced

Research Question

- How can we better reveal and understand students' complex ideas?
  - When students construct own answer, more likely to reveal mix of ideas
  - Impossible to analyze all students' submissions in very large courses
  - Can computer help?
    - **Lexical analysis** allows the processing of large numbers of student responses to reveal ***common patterns of ideas***

Why Research Photosynthesis?

- Photosynthesis a complex biological process
  - energy transformations
  - molecular rearrangements
  - structure/function relationships
- Existing diagnostic questions* and research into student difficulties

*DQC = Diagnostic Question Cluster

Methods

- Exam data from introductory cell biology course (n=391)
- Each student received one MC DQC and one constructed response
- Used 2 versions of the DQC questions that allowed a cross-over design
- Lexical analysis by SPSS Text Analytics for Surveys

Multiple Choice Questions

- Q. A mature maple tree can have a mass of 1 ton or more (dry biomass, after removing water), yet it starts from a seed that weighs less than 1 gram. Which of the following contributes most to this huge increase in biomass?

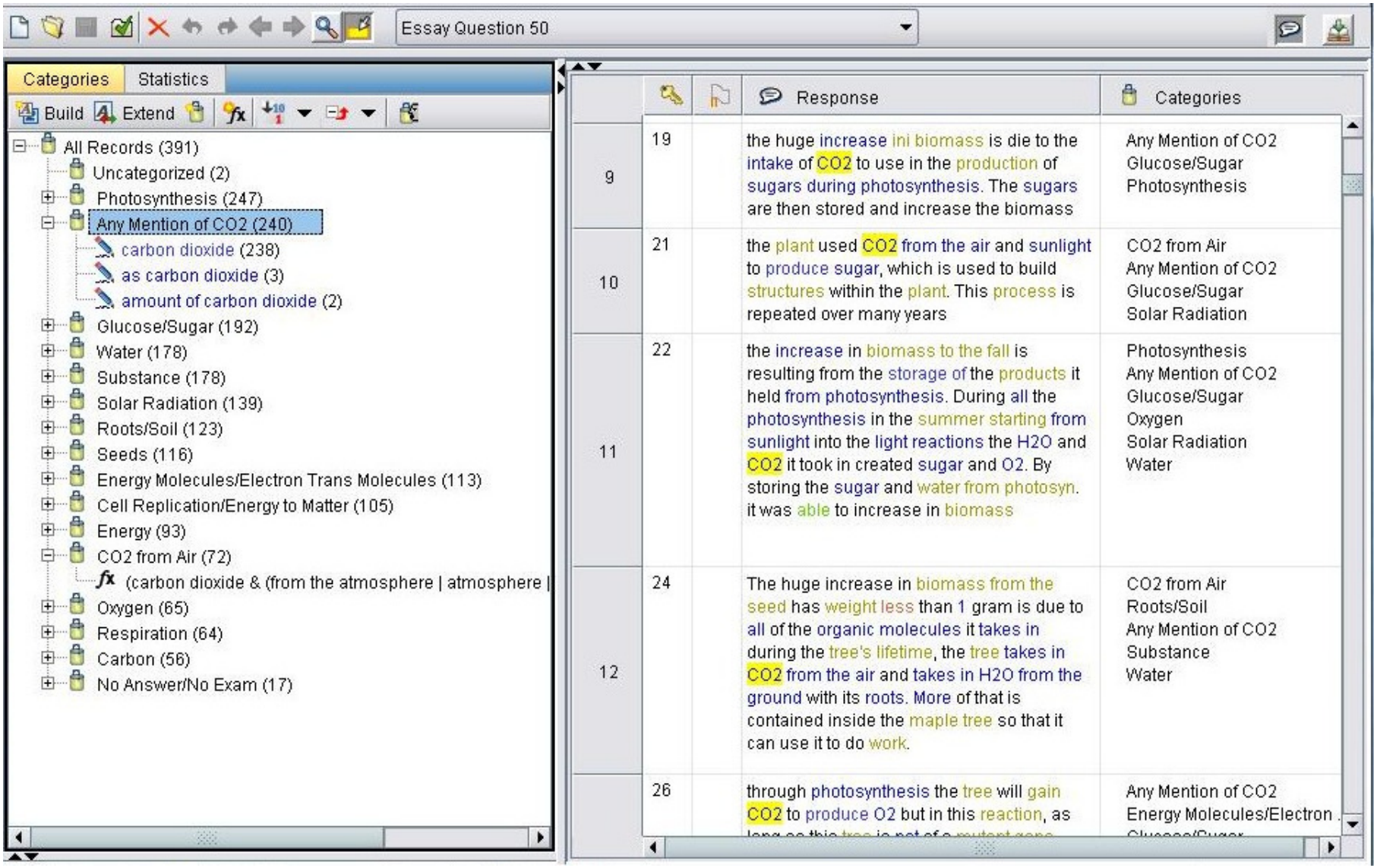
  A. Absorption of mineral substances from root (7.7%)
  B. Absorption of organic substances from soil via roots (12.7%)
  C. **Incorporation of CO2 gas from atmosphere into molecules by green leaves (59.4%)**
  D. Incorporation of H2O from soil into molecules by green leaves (7.7%)
  E. Absorption of solar radiation into the leaf (12.7%)

- A similar question stem using corn and same distractors was also used

Constructed Response Prompt

Q: *A mature maple tree can have a mass of 1 ton or more (dry biomass, after removing the water), yet it starts from a seed that weighs less than 1 gram.* ***Explain this huge increase in biomass****.*

Lexical Analysis

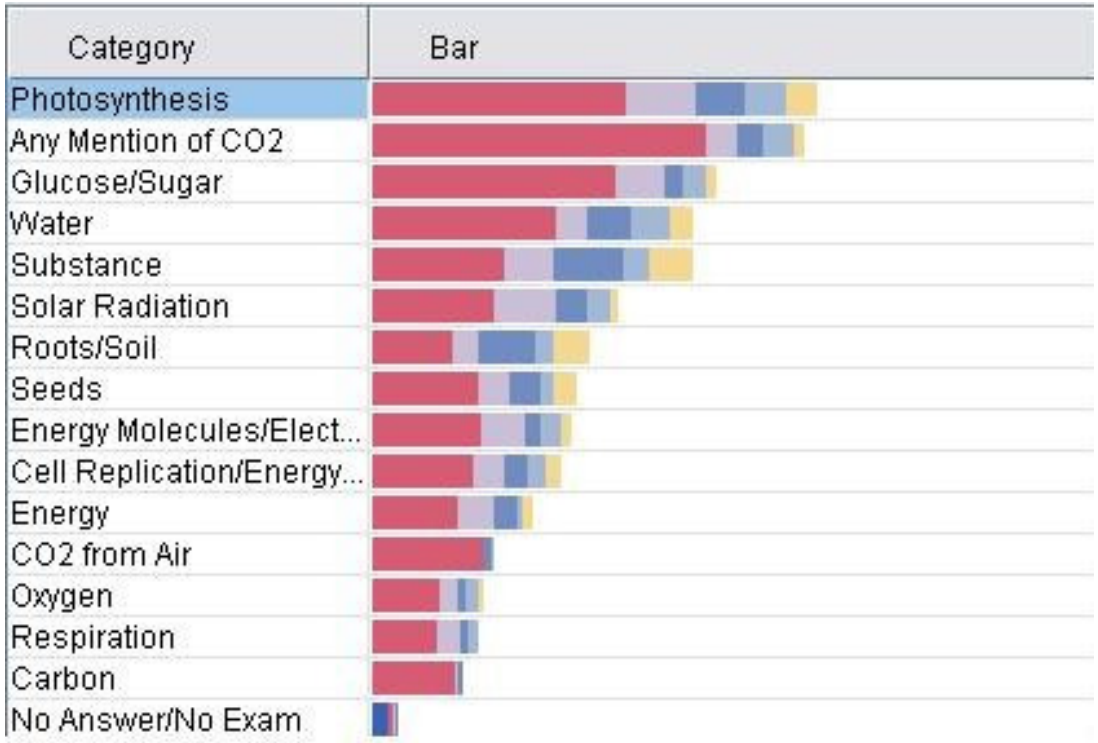


Frequencies of concepts expressed in Constructed Responses

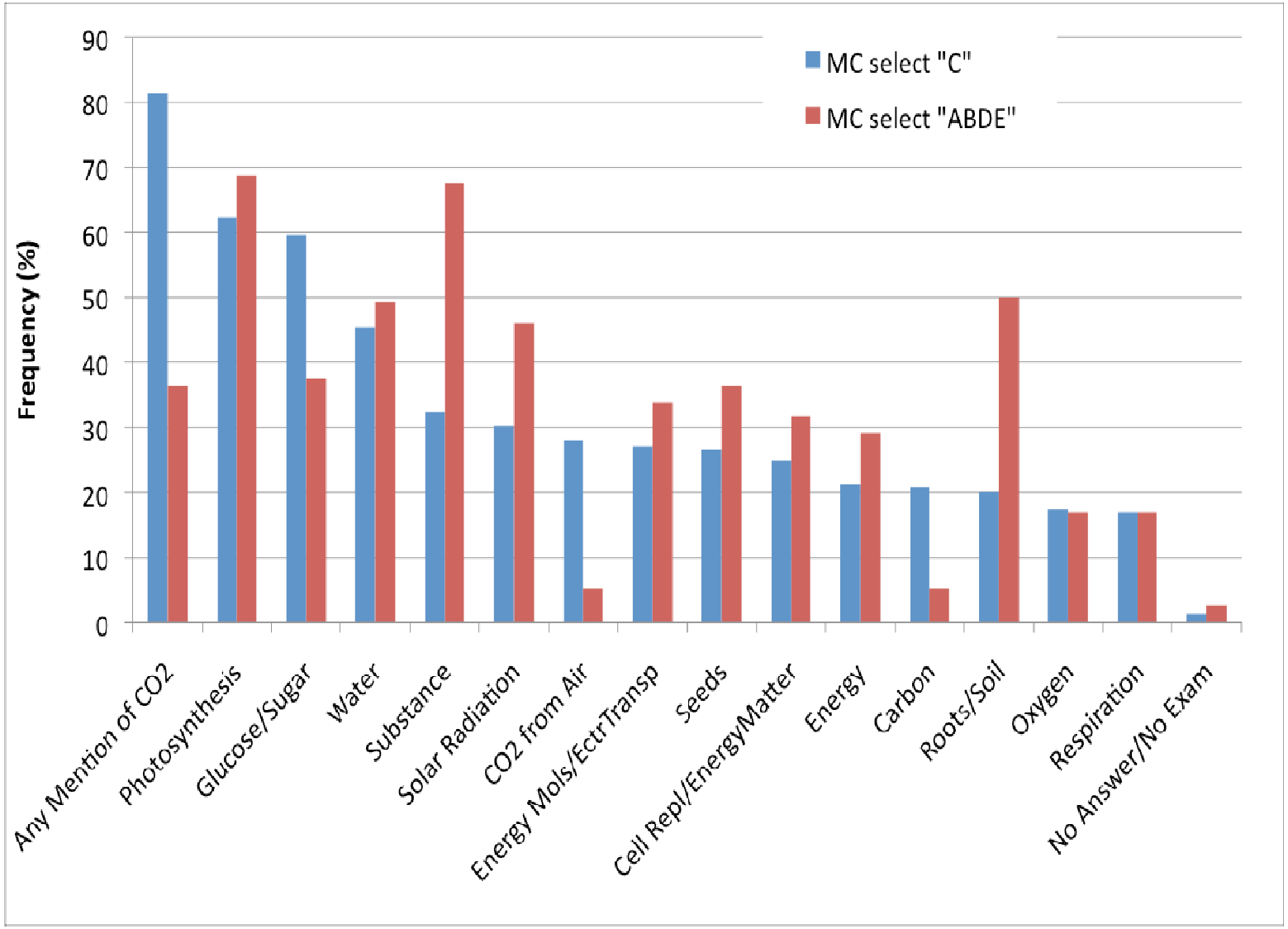
Category	Bar	Selection %
Photosynthesis		63.2
Any Mention of CO2		61.4
Glucose/Sugar		49.1
Water		45.5
Substance		45.5
Solar Radiation		35.5
Roots/Soil		31.5
Seeds		29.7
Energy Molecules/Elec...		28.9
Cell Replication/Energ...		26.9
Energy		23.8
CO2 from Air		18.4
Oxygen		16.6
Respiration		16.4
Carbon		14.3
No Answer/No Exam		4.3

Students' explanations reveal a more complex picture than multiple choice

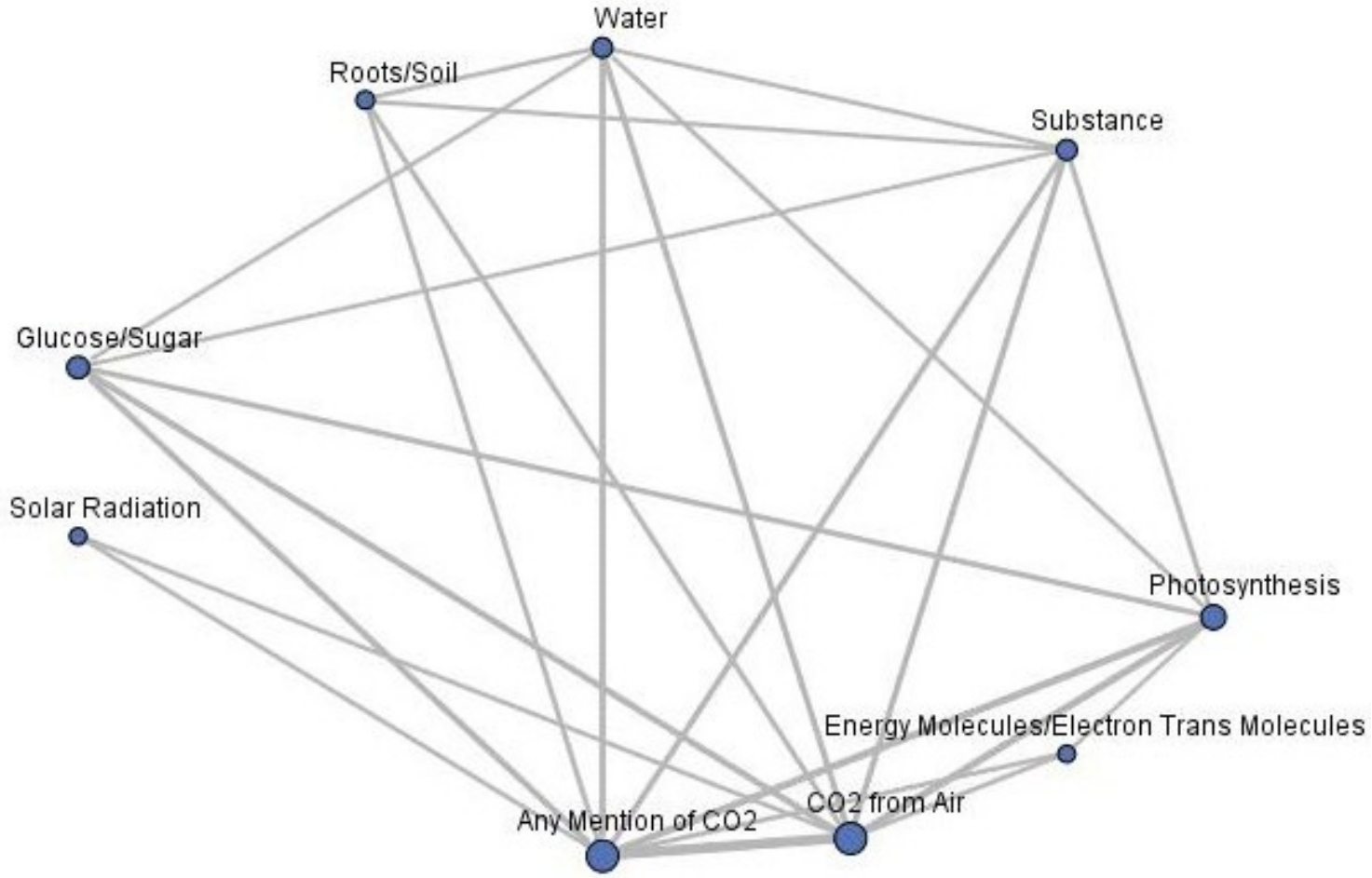
- Concepts in constructed response coded by MC choice



MC Selection vs. CR Categories



Students' concept heterogeneity revealed through written explanations



Conclusions

- Lexical analysis can reveal patterns of concepts present in large number of responses
- Students have complex and heterogeneous ideas
- Constructed response provides a unique view of this complexity that can be missed by multiple choice items

Conclusions

- Nothing particularly unique about analyzing student writing, *per se*
- Ability to accurately analyze LARGE numbers quickly at low cost **is unique**
- Formative feedback to instructor about patterns of ideas of whole class allows **rapid instructional response** (JiTT)
- Even reading lots of responses would be unlikely to reveal these patterns in responses – exceptional research tool

Next Steps

- Use lexical categories as predictors of human scoring
  - Different rubric types
- Creation and implementation of DQC-type instrument as opposed to questions

Acknowlegements

MICHIGAN STATE UNIVERSITY

- Collaborators
  - Automated Analysis of Constructed Response (AACR) Group
  - Diagnostic Question Cluster (DQC) Group
  - Joyce Parker, Duncan Sibley

- Funding
  - National Science Foundation
    - DUE 0736952 and DUE 1022653

Visit our group webpage: aacr.crcstl.msu.edu