How Are Students Interpreting Constructed Response Questions? Using Computerized Lexical Analysis to Identify Key Concepts in Student Writing

Michele Weston, Kevin Haudek, Luanna Prevost, Casey Lyons, Mark Urban-Lurain, John Merrill, and Joyce Parker

Automated Analysis of Constructed Response (AACR) research group

MICHIGAN STATE UNIVERSITY

Acknowledgements

- This material is based upon work supported by the National Science Foundation (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 NSF.

Overview

- Benefits and challenges of constructed response questions
- Research question: Altering a photosynthesis question stem
- Using text analysis to analyze written responses
- Using an analytic scoring rubric to analyze written responses
- Comparison of Stem Version One to Stem Version Two

Using Constructed Response Questions

- Constructed response questions require students to create a written answer from their rationalization of the question and the concepts involved in it (Kuechler & Simkin 2010)

- Multiple choice questions require students to separate and favor presented choices (Birenbaum & Tatsouka 1987)

Kuechler, W.L., & Simkin, M.G. (2010). Why is performance on multiple-choice tests and constructed-response tests not more closely related? Theory and an empirical test. *Decision Sciences Journal of Innovative Education 8: 55–73.*

- Expert vs. novice reasoning
- A challenge to using constructed response questions is writing the stem so that it is meaningful to students

Ambiguity in Responses

Stem Version One (Fall 2009):

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.

The increase is due to a large absorption of minerals and organic materials through the soil then joining their benefits with H2O, O2 and sunlight to help reach maximum production.

The maple tree is able to increase the mass so much because a majority of the mass is due to CO2 gas from the atmosphere. The green leaves on the tree absorb the CO2 and increase the maple tree's mass.

Research Question

- What is the impact on students' written explanations when they are explicitly asked to identify a process in their response?

Methods

Stem Version One (V1) (Fall 2009):

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.

Stem Version Two (V2) (Fall 2010 & Fall 2011):

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 where this biomass comes from and by what process.

Study Design

- Cells and Molecules
  - Introductory biology course
  - Prerequisite is general chemistry
- Post-instruction on photosynthesis
- Given on an exam

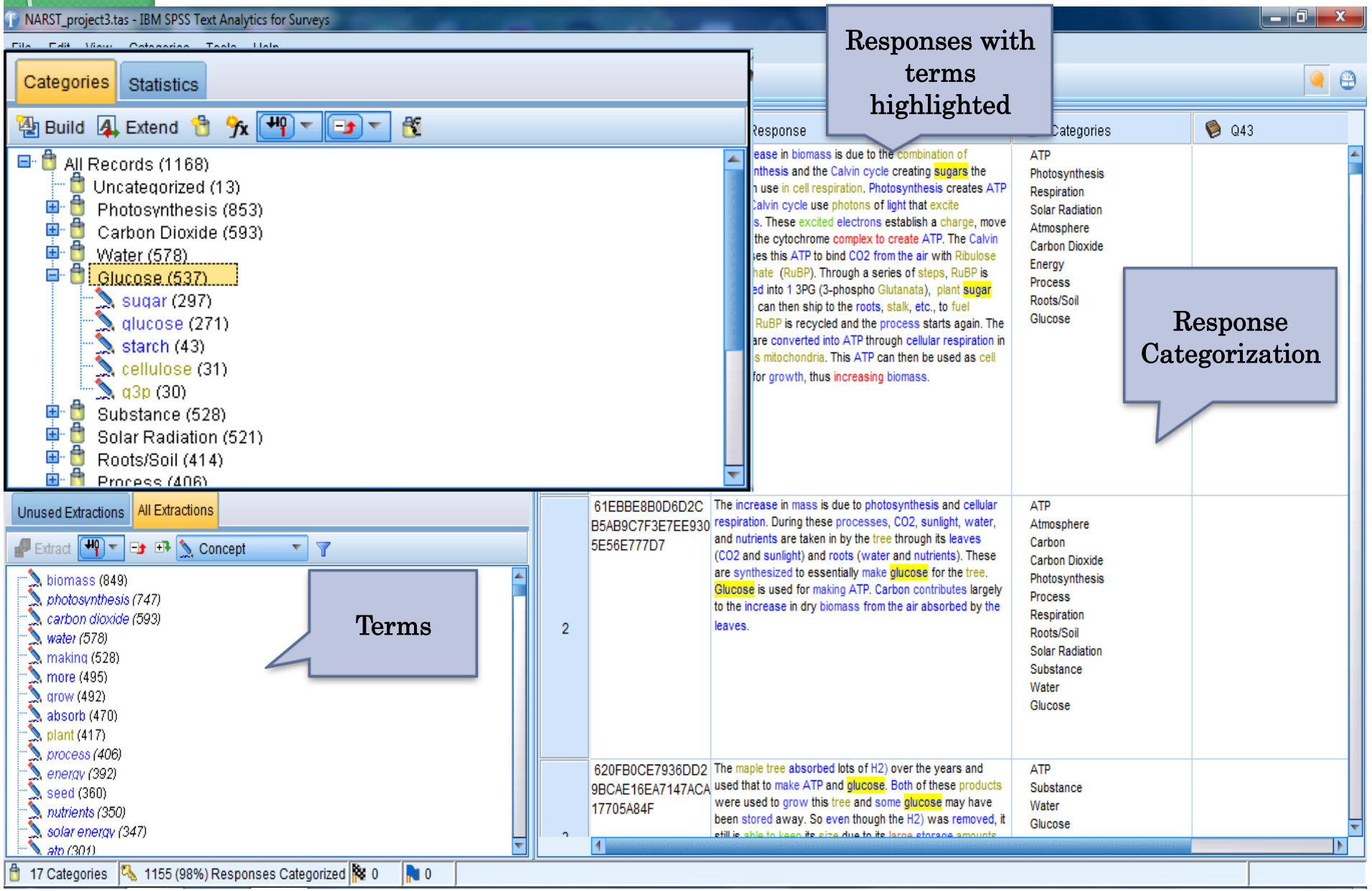
V1: A mature maple tree can have a mass of 1 ton or more (dry biomass, after removing the water), yet it starts from a seed that weighs less than 1 gram. Explain this huge increase in biomass.

Fall 2009: 385 students

V2: 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 where this biomass comes from and by what process.

Fall 2010: 394 students
Fall 2011: 458 students

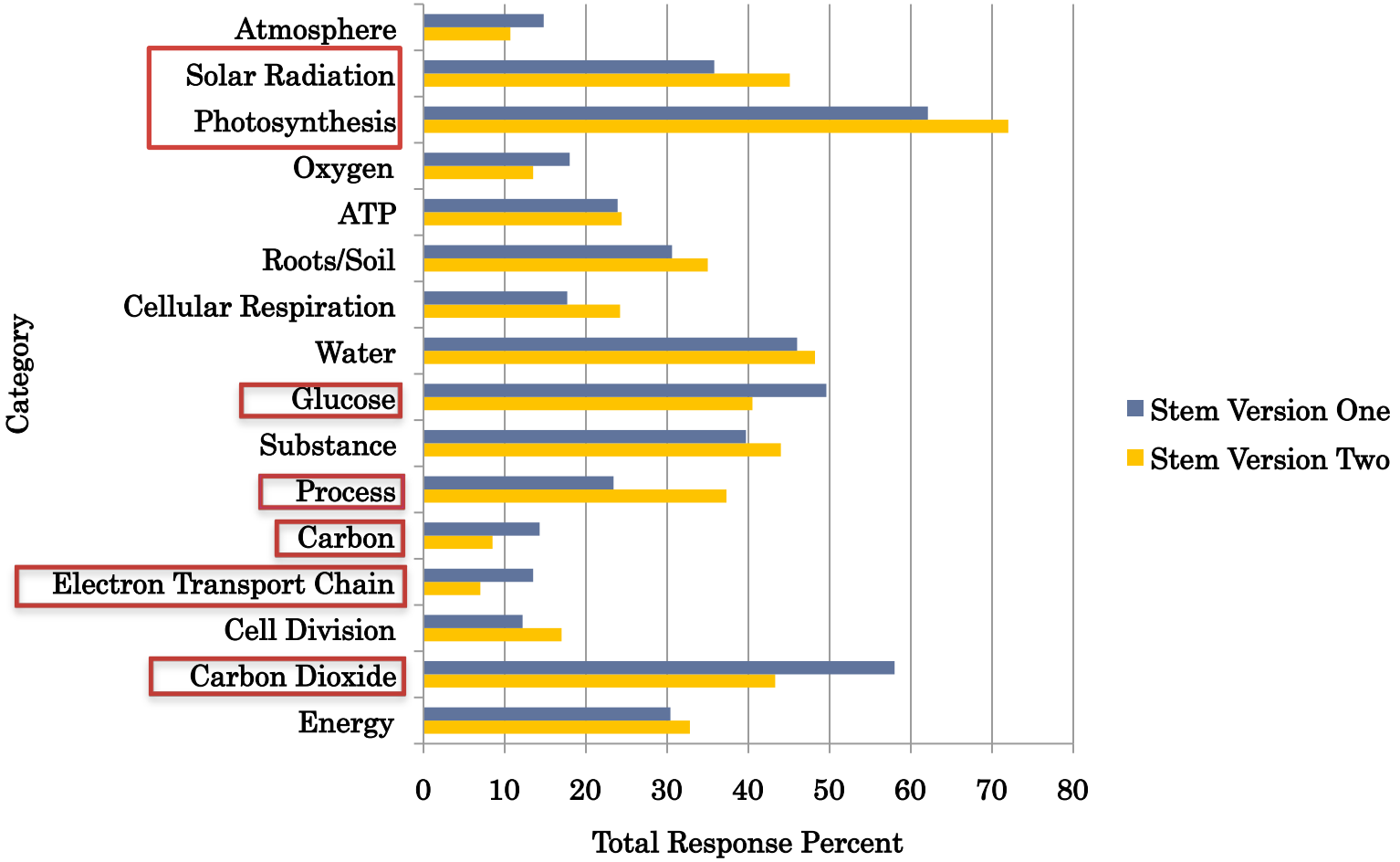
IBM SPSS Text Analytics for Surveys



Statistical Analysis

- Independent two-sample t-test on each category
- ANOVA test for collective means

Distribution of Categories



ANOVA

Category	Significance
Energy	.397
Carbon Dioxide	.000
Cell Division	.031
Electron Transport Chain	.000
Carbon	.002
"Process"	.000
Substance	.158
Glucose	.003
Water	.475
Respiration	.010
Roots/Soil	.139
ATP	.837
Oxygen	.041
Photosynthesis	.000
Solar Radiation	.002
Atmosphere	.042

Example Responses

- Stem Version One

The maple tree absorbed lots of H2O (sic) over the years and used that to make ATP and glucose. Both of these products were used to grow this tree and some glucose may have been stored away. So even though the H2O (sic) was removed, it still is able to keep its size due to its large storage amounts of nutrients, glucose and water.

The huge increase in biomass from the seed has weight less than 1 gram is due to all of the organic molecules it takes in during the tree's lifetime, the tree takes in CO2 from the air and takes in H2O from the ground with its roots. More of that is contained inside the maple tree so that it can use it to do work.

Purple Font = Processes

Red Font = Matter

- Stem Version Two

Process responsible for increased biomass is photosynthesis. The biomass came from the incorporation of CO2 gas from the atmosphere that was delivered to molecules in the green leaves, sunlight helps create glucose and nutrients moved to roots.

This biomass came from the processes of photosynthesis and the calvin cycle that occur within the leaves of these plants. In photosynthesis the plant takes in sunlight and uses this energy to produce ATP and NADPH for the calvin cycle. Then, in the calvin cycle the plant uses the ATP and NADPH, and combines it with CO2 from the air to produce sugar or starch. This starch accumulates in the plant leaving the plant weighing more than it was initially.

Purple Font = Processes

Red Font = Matter

Text Analysis Results	
Significant Increase	Significant Decrease
Photosynthesis Electron Transport Chain Solar Radiation "Process"	Glucose Carbon Carbon Dioxide

The biomass added can be contributed to the production of ATP during photosynthesis and the use of ATP in the production of glucose using water and CO2 during cellular respiration.

the biomass comes from the plant's ability to produce its own food. Using sunlight as energy and water and fertilizer/nutrients the plant is able to produce ATP and starch which allows it to grow. The energy is put toward mitosis and cell growth, causing the plant to grow in size and weight

The absorption of CO2, sunlight, water and other nutrients through the soil are used and cause cells to grow and replicate.

Scoring with Analytic Rubric

- Detects important concepts and misconceptions in student writing
- All 1,232 responses were human scored

Correct Concepts

1. Correct Process
Photosynthesis, Calvin Cycle

2. Correct Source
Carbon dioxide

3. Correct Product
Glucose

4. Glucose Added as Biomass
Glucose is stored which adds to plant biomass

5. Water as Source
Water is an input of photosynthesis

Misconceptions

6. Incorrect Process
Respiration, cell division, light reactions, solar radiation, electron transport chain

7. Incorrect Source
ATP, energy, Nitrogen, Oxygen, Carbon

8. Incorrect Product
ATP, energy, carbon dioxide

9. Glucose Used as Energy
Glucose is used for energy to power other reactions

10. Incorrect Soil/Nutrients
Nutrients from the soil are the primary source of biomass

Comparison of Concepts to Lexical Categories

Analytic Rubric Results	
Significant Increase	Significant Decrease
Correct Process Incorrect Process Incorrect Source Water as Source Incorrect Source Soil/Nutrients	Correct Source Correct Product Glucose Added As Biomass

Text Analysis Results	
Significant Increase	Significant Decrease
Photosynthesis Electron Transport Chain Solar Radiation "Process"	Glucose Carbon Carbon Dioxide

Purple = Correct Process Blue = Correct Product Red = Correct Source

Association Rules

Concept	Category	Coverage	Accuracy
Correct Source	Carbon Dioxide	50.9%	94.4%
Correct Process	Photosynthesis	73.6%	96.7%
Correct Process	Solar Radiation	42.1%	80.1%
Correct Product	Glucose	46.1%	95.7%

Witten, Ian H., and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques.* San Francisco: Elsevier Inc., 2005. Print.

Using Text Analysis and Human Scoring Together

- Text analysis shows patterns in data that would be hard to see otherwise
- Human scoring can be used to inform our category creation and refinement for text analysis

Comparison of Stem Version One to Stem Version Two

Analytic Rubric Results	
Significant Increase	Significant Decrease
Correct Process Incorrect Process Incorrect Source Water as Source Incorrect Source Soil/Nutrients	Correct Source Correct Product Glucose Added As Biomass

Text Analysis Results	
Significant Increase	Significant Decrease
Photosynthesis Electron Transport Chain Solar Radiation "Process"	Glucose Carbon Carbon Dioxide

Conclusion

- Even small changes in wording can influence how students respond.
- If the question is written so that it is clear to novices the amount of scientific detail they should provide in their answer, then they will provide more meaningful responses.

Future Work

- Refine categories
  - Ex. Different ways of using sunlight
- Machine learning to make scoring models of analytic rubric

Questions

- Michele Weston
  - westonmi@msu.edu

- http://aacr.crcstl.msu.edu

Category	Total Sum of Squares	Total df	F	Significance
Energy	268.4	1231	.718	.397
Carbon Dioxide	307.5	1231	22.970	.000
Cell Division	161.4	1231	4.653	.031
Electron Transport Chain	101.0	1231	13.947	.000
Carbon	114.0	1231	9.640	.002
"Process"	272.3	1231	23.660	.000
Substance	301.4	1231	1.998	.158
Glucose	302.5	1231	9.005	.003
Water	307.2	1231	.511	.475
Respiration	212.5	1231	6.589	.010
Roots/Soil	274.9	1231	2.191	.139
ATP	226.4	1231	.042	.837
Oxygen	155.8	1231	4.176	.041
Photosynthesis	263.9	1231	12.312	.000
Solar Radiation	300.519	1231	9.352	.002
Atmosphere	130.221	1231	4.138	.042

Alignment with Interview Data

- Interviews probed students' thoughts post-photosynthesis instruction
- In the interviews, two students explained their answer as being either scientific or "logical"
- Examples from our study show instances of informal reasoning similar to what was obtained through interviews
- Ex. *The increase (sic) in biomass is due to the tree (sic) gradually growing over the years. As it grows, the trunk gets larger, producing more mass and the branches extend and grow in length and diameter, also increasing mass*

Parker, J.M., Anderson, C.W., Heidemann, M., Merrill, J., Merritt, B., Richmond, G., & Urban-Lurain, M. (2012). Exploring undergraduates' understanding of photosynthesis using Diagnostic Question Clusters. *CBE Life Sciences Education 11(1): 47-57.*

Discriminant Analysis

Analytic Rubric Concept	Percent Correctly Classified Cases	Kappa Statistic
Correct Process	95.6%	.946
Incorrect Process	93.6%	.931
Correct Source	92.0%	.898
Incorrect Source	95.1%	.951
Incorrect Source Soil/ Nutrients	88.6%	.866
Water as Source	90.2%	.878
Correct Product	97.8%	.971
Incorrect Product	85.3%	.846
Glucose Added to Biomass	88.2%	.880
Glucose Added to Energy/ATP	88.2%	.876

SVM Scoring Model

Analytic Rubric Concept	Percent Correctly Classified Cases	Kappa Statistic
Correct Process	91.6%	.900
Incorrect Process	89.1%	.885
Correct Source	92.4%	.903
Incorrect Source	92.4%	.924
Incorrect Source Soil/ Nutrients	87.5%	.845
Water as Source	89.9%	.805
Correct Product	97%	.961
Incorrect Product	50.7%	.466
Glucose Added to Biomass	87.7%	.875
Glucose Added to Energy/ATP	81.7%	.804