Comparing Formative Feedback Reports: Human and Automated Text Analysis of Constructed Response Questions in Biology

Michele Weston, Joyce Parker, and Mark Urban-Lurain

Automated Analysis of Constructed Response (AACR) research group
Constructed Response Questions vs. Multiple Choice

- 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)

- Students treat multiple choice questions and CR questions as different cognitive tasks (Stanger-Hall 2012)


Using Constructed Response Questions

- CR questions require a greater investment of time to evaluate than multiple choice questions

- To save time, instructors tend to read a sample of responses
Using Text Analysis to Speed Up Scoring

- Computerized text analysis (TA) can summarize the distribution of key concepts and misconceptions in student writing.

- TA identifies words and phrases that can be further analyzed with statistical modeling.
Background

• Interested in exploring the use of computerized text analysis to evaluate constructed-response assessments

• Photosynthesis project: Looking for differences in how students respond to different question stems

• Other projects in chemistry, thermodynamics, genetics, and evolution
Research Questions

• How do instructors analyze answers to CR questions?
Instructor Analyses

- Expert scorers experienced with reading student written responses
- Had not seen the responses to this question before
- Read for formative assessment of the students’ writing
Research Questions

- How do instructors analyze answers to CR questions?
- How do the results from computerized text analysis techniques compare with the instructors’ analyses?
Not all cells in plants (e.g. root cells) contain chlorophyll required for photosynthesis. How do these cells get energy? (Parker et al. 2012)

- Introductory biology course: Cells and Molecules
  - Prerequisite is general chemistry
- Given as a homework on an online course management system
- Post-instruction on cell metabolism
- 360 out of 468 students responded
Automated Analysis

Responses with terms highlighted

Response Categorization

Terms

During respiration, sugars produce ATP which is how these types of cells get energy.
Exploratory Analyses

Instructor Analyses

- A random sample of 50 responses was chosen to be read separately by the two instructors
- Instructor 1 tallied emergent ideas in the students’ writing
- Instructor 2 also kept track of key ideas

Automated Analysis

- Full 360-response dataset
- K-means cluster analysis
<table>
<thead>
<tr>
<th>Instructor 1</th>
<th>Description of Cluster</th>
<th>Instructor 2</th>
<th>Description of Cluster</th>
<th>Automated Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution of Responses</td>
<td></td>
<td>Distribution of Responses</td>
<td></td>
<td>Distribution of Responses</td>
</tr>
<tr>
<td>15%</td>
<td>1. Accurate description based on the transport of glucose and/or cellular respiration</td>
<td>22%</td>
<td>1. Sugar being transported</td>
<td>13%</td>
</tr>
<tr>
<td>20%</td>
<td>2. Transport of energy</td>
<td>14%</td>
<td>2. Energy being transferred throughout the plant</td>
<td>23%</td>
</tr>
<tr>
<td>14%</td>
<td>3. Roots drawing something from the soil for energy</td>
<td>22%</td>
<td>3. Energy comes from nutrients from the soil</td>
<td>17%</td>
</tr>
<tr>
<td>16%</td>
<td>4. Special processes such as C4 photosynthesis, Calvin Cycle, and Kreb’s Cycle</td>
<td>10%</td>
<td>4. Incorrect process such as C4 photosynthesis and heterotrophy</td>
<td>16%</td>
</tr>
<tr>
<td>2%</td>
<td>5. Force-dynamic style explanation</td>
<td>10%</td>
<td>5. Respiration as the process involved</td>
<td>12%</td>
</tr>
<tr>
<td>4%</td>
<td></td>
<td>6. Energy being transferred and nutrients from the soil</td>
<td>16%</td>
<td>6. Incorrect source of energy such as from other organisms</td>
</tr>
</tbody>
</table>
Exploratory Analysis: Clustering Based on Emergent Ideas

*These cells take energy from what is around it. For example the root cells take energy from the soil that it is in. Also some cells get their energy transported from the part of the plant that contains chlorophyll.*

Instructor 1: Transport of energy

Instructor 2: Energy being transferred and nutrients from the soil

Automated Analysis: Energy being used from other parts of the plant
Developing an Analytic Rubric

<table>
<thead>
<tr>
<th>Instructor 1 Analytic Bins</th>
<th>Instructor 2 Analytic Bins</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Correct source</td>
<td>1. Transport sugars</td>
</tr>
<tr>
<td>2. Incorrect source/unspecified energy</td>
<td>2. Transport energy</td>
</tr>
<tr>
<td>3. Incorrect source/nutrients from soil</td>
<td>3. Transport ATP</td>
</tr>
<tr>
<td>5. Correct Process</td>
<td>4. Respiration</td>
</tr>
<tr>
<td>6. Incorrect Process</td>
<td>5. Dark Reactions</td>
</tr>
<tr>
<td></td>
<td>7. From surroundings</td>
</tr>
<tr>
<td></td>
<td>8. Transport products of photosynthesis</td>
</tr>
<tr>
<td></td>
<td>9. C4 photosynthesis</td>
</tr>
<tr>
<td></td>
<td>10. Heterotrophy</td>
</tr>
</tbody>
</table>
New Question Design

Predictive Model

Confirmatory Analysis

Human Coding

Rubric Development

Data Collection

Lexical Resource Development

Question Revision

Exploratory Analysis
Human scoring was done using Instructor 1’s rubric

An assistant was calibrated to Instructor 1 using 120 responses from a previous semester

Then, used that calibration to score more responses

- 360 responses (FS12) + 316 responses (FS10) = 676 responses total
- Discriminant analysis predicts human scoring
- 676 responses used to build model
- Leave-one-out system of classification
## Automated Scoring with an Analytic Rubric

<table>
<thead>
<tr>
<th>Rubric</th>
<th>Description</th>
<th>Human Scoring</th>
<th>Correctly Classified</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Correct Source</td>
<td>Any name for a product of photosynthesis</td>
<td>26%</td>
<td>94.2%</td>
</tr>
<tr>
<td>2. Incorrect source/unspecified energy</td>
<td>Energy or ATP being transported</td>
<td>26%</td>
<td>81.7%</td>
</tr>
<tr>
<td>3. Incorrect source/nutrients from soil</td>
<td>Energy comes from nutrients from the soil</td>
<td>19%</td>
<td>90.7%</td>
</tr>
<tr>
<td>4. Incorrect Source/water</td>
<td>Water without anything else or water with nutrients from the soil</td>
<td>11%</td>
<td>94.1%</td>
</tr>
<tr>
<td>5. Correct Process</td>
<td>Respiration or glycolysis</td>
<td>17%</td>
<td>96.8%</td>
</tr>
<tr>
<td>6. Incorrect Process</td>
<td>Various incorrect processes</td>
<td>12%</td>
<td>92.6%</td>
</tr>
</tbody>
</table>

n=360
Research Questions

- How do instructors analyze answers to CR questions?
  - Instructors who participated in this project did initial exploratory analyses to look for key ideas and misconceptions
  - Then, used those ideas to make analytic scoring rubrics
How do the results from automated text analysis techniques compare with the instructors’ analyses?

- Discriminant analysis predicted **above 80%** for all the rubrics and **above 90%** for five of them.

- 5 of 6 automated clusters overlapped with human clusters.
Time Requirements

- Exploratory analysis
  - Instructor 1, 11 minutes
  - Instructor 2, 45 minutes
  - Automated analysis, 15 minutes

- Analytic rubric scoring
  - Human scoring took approximately 5 hours after calibration
  - Automated analysis took less than 5 minutes
Future Direction

• Just-in-Time Teaching (JiTT) instructor formative feedback reports

Questions

- Michele Weston
  - westonmi@msu.edu

www.msu.edu/~aacr

Acknowledgements
We thank John Merrill for creating an instructor rubric.

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.