Automated Analysis Provides Insights into Students’ Challenges Understanding the Processes Underlying the Flow of Genetic Information

Rosa A. Moscarella, Kevin C. Haudek, Jennifer K. Knight, Alexandria Mazur, Karen N. Pelletreau, Luanna B. Prevost, Michelle K. Smith, Matthew Steele, Mark Urban-Lurain, John E. Merrill

Automated Analysis of Constructed Response Research Group

www.msu.edu/~aacr

NARST 2016 International Conference
Baltimore, April 14-17 2016
Question Development Cycle

1. Predictive Model
2. Confirmatory Analysis
3. Human Coding
4. Statistical Exploratory Analysis
5. New Question Design
6. Data Collection
7. Lexical Resource Development
8. Question Revision
9. Human Rubric Development
Outline

• Lexical analysis of constructed responses
• Biology literacy and genetics
• Computer predicted scores for Pre/Post responses
• Path maps
• Conclusions
• Next steps
Lexical analysis

Students responses → Lexical categories → Automated scoring

Feedback reports
Lexical analysis

Students responses → Lexical categories → Automated scoring

Feedback reports
The house where I grew up was big with yellow and white walls and had many rooms.
Lexical categories

The house where I grew up was big with yellow and white walls and had many rooms.
The house where I grew up was **big** with **yellow** and **white** walls and had many **rooms**.

<table>
<thead>
<tr>
<th>Size</th>
<th>Color</th>
<th>Vegetation</th>
<th>Interior</th>
</tr>
</thead>
<tbody>
<tr>
<td>big</td>
<td>green</td>
<td>garden</td>
<td>stairs</td>
</tr>
<tr>
<td>small</td>
<td>yellow</td>
<td>backyard</td>
<td>rooms</td>
</tr>
<tr>
<td>medium</td>
<td>white</td>
<td>roses</td>
<td>painting</td>
</tr>
</tbody>
</table>
**Automated scoring**

<table>
<thead>
<tr>
<th>Responses</th>
<th>Size</th>
<th>Color</th>
<th>Vegetation</th>
<th>Interior</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>The house were I grew up (...)</em></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
The flow of genetic information (...)  1  1  0  0

<table>
<thead>
<tr>
<th>Responses</th>
<th>DNA</th>
<th>No Effect</th>
<th>Codon</th>
<th>Stop</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>The flow of genetic information</em> (…)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Score**
- Correct
- Incomplete
- Incorrect
The flow of genetic information is considered a core concept (AAAS, 2011).

Genetics is fundamental for biology literacy.

- This knowledge is necessary to understand other core concepts, e.g., mechanisms of evolution (Kalinowski et al., 2010).
The following DNA sequence occurs near the middle of the coding region of a gene:

**DNA 5' A A T G A A T G G* G A G C C T G A A G G A 3'**

There is a G to A base change at the position marked with an asterisk. Consequently, a codon normally encoding an amino acid becomes a stop codon.

1. How will this alteration influence **DNA replication**?
2. How will this alteration influence **transcription**?
3. How will this alteration influence **translation**?
Computer predicted scores

Intro Bio for Majors
(Pre n= 1610; Post n=1620)

Genetics
(Pre n= 430; Post n=388)
Computer predicted scores

Intro Bio for Majors
(Pre n= 1610; Post n=1620)

Genetics
(Pre n= 430; Post n=388)
Transcription path maps Intro Bio for Majors

1. Correct
2. Incomplete/Irrelevant
3. Incorrect
Transcription path maps Intro Bio for Majors

1. Correct
2. Incomplete/Irrelevant
3. Incorrect
Transcription path maps Intro Bio for Majors

1. Correct
2. Incomplete/Irrelevant
3. Incorrect
Most common misconception path
Replication: DNA would continue to replicate but it would replicate the mutation. Correct

Transcription: With a nonsense mutation, the translation of mRNA stops early. Incorrect

Translation: Nonsense mutation causes premature termination of translation. Correct
Stop-codons-stop-everything path
Stop-codons-stop-everything path

Replication: I believe the DNA would not fully replicate. Incorrect

Transcription: transcription would stop at this place and consequent in an incomplete protein. Incorrect

Translation: translation would not be complete because the stop codon prevented the rest of the genetic message from being communicated to the ribosomes. There is a gap and therefore an incomplete protein being formed. Correct
Conclusion

- This analysis (N=3230) would not have been possible without the automated analysis of constructed responses.
- Path maps are very useful to illustrate students performance in multiple parts questions.
- Transcription is a keystone concept to understand the flow of genetic information.
  - Emphasis on differentiating DNA replication and gene expression.
  - Improving the instruction of this topic in introductory biology.
Next steps

- Analysis of students interviews to better understand learning difficulties with transcription
  - Use and knowledge of technical terminology
  - mRNA synthesis and purpose of transcription
  - Initiation and termination of transcription
  - Intrinsic student’s motivation
Acknowledgments

- AACR-MSU Group
  - Mark Urban-Lurain
  - John Merrill
  - Melanie Cooper
  - Carl Lira
  - Kevin Haudek
  - Andrea Bierema
  - Anne-Marie Hoskinson
  - Rosa Moscarella
  - Matthew Steele
  - Alexandria Mazur
  - Tanner Foster

This material is based upon work supported by the National Science Foundation (DUE grants: 1323162 and 1347740). 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.
Contacts

- Automated Analysis of Constructed Response Research Group
  www.msu.edu/~aacr
- Rosa A. Moscarella
  moscarel@msu.edu