Using Lexical Analysis to Explore Students' Written Responses to Genetics Concept Assessment-Derived Items

Luanna Prevost, Jenny Knight, Michelle Smith,
Kevin Haudek, John Merrill, Mark Urban Lurain

University of Colorado Boulder

1865 THE UNIVERSITY OF MAINE

MICHIGAN STATE UNIVERSITY

Acknowledgements

- Funding
  NSF DUE 1022653

- Automated Analysis of Constructed Response (AACR) research group
  - Michigan State
  - CU- Boulder
  - University of Maine
  - The Ohio State University
  - University of Georgia

- Website:

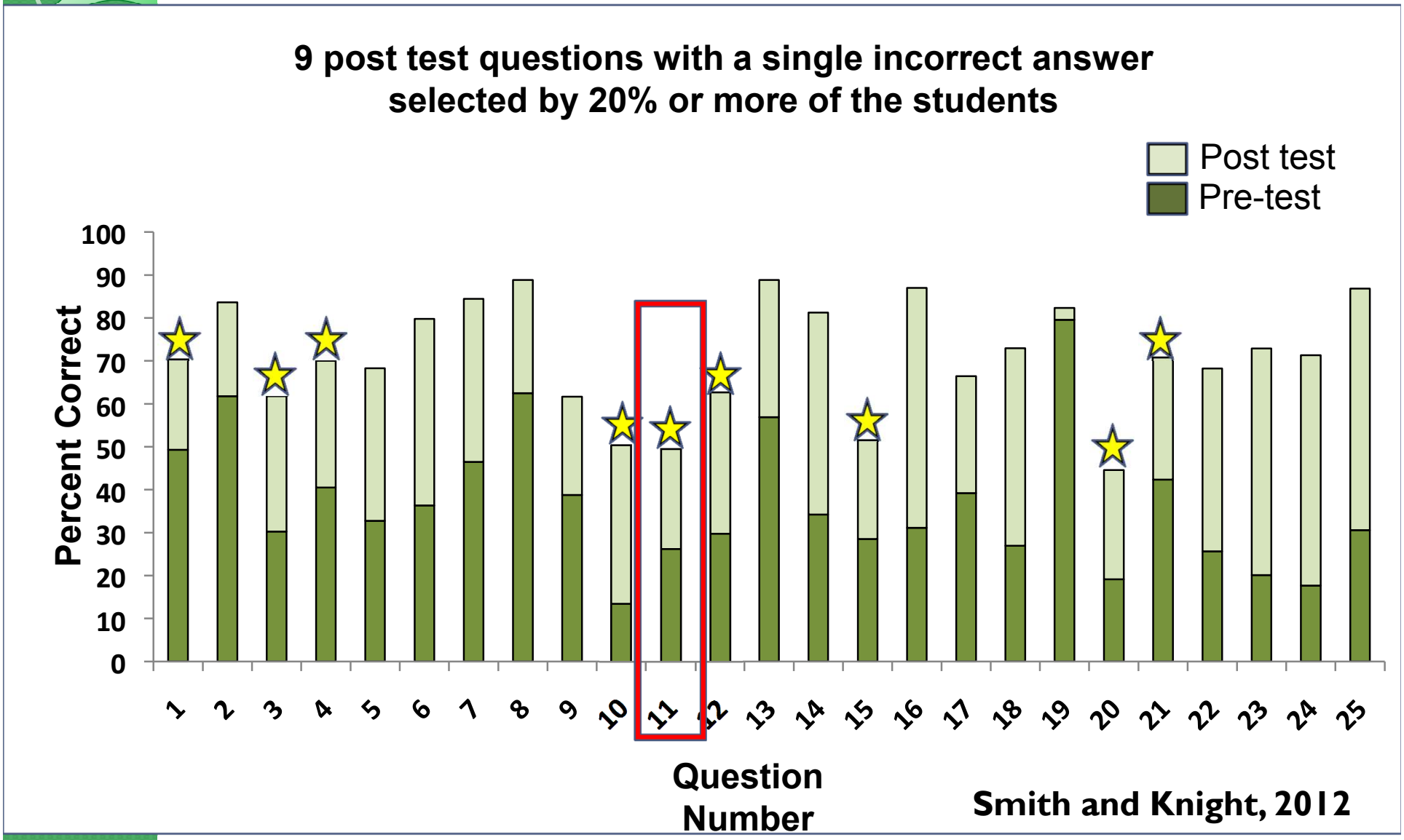
aacr.crcstl.msu.edu

Objectives

- Create models of student thinking
- Use lexical and statistical analysis to analyze students' writing

- Misconceptions previously identified by Genetics Concept Assessment (GGA) (Smith et al., 2008)

Subset of GCA questions are particularly difficult on the post test



Use the following mRNA codon key as needed to answer the next two questions (key omitted here):

The following DNA sequence (coding strand) occurs near the middle of the coding region of a gene.

DNA

Students think stop codon stops transcription

...has

...s in

the DNA. The first triplet of nucleotides AAU (underlined) is in frame for coding, and encodes Asparagine as the codon table above indicates.

mRNA

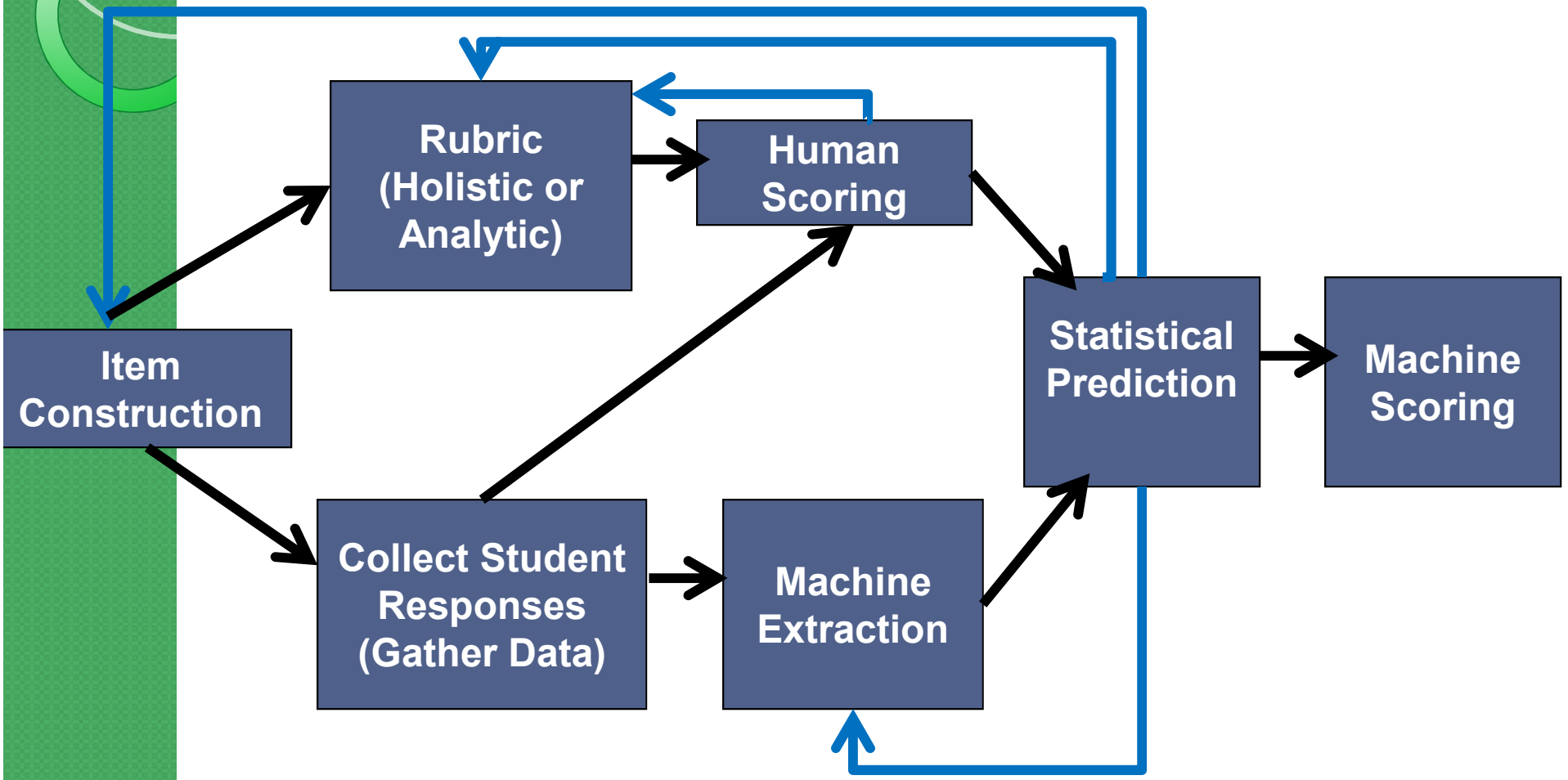
50 55 60 65

5'—A A U G A A U G G G A G C C U G A A G G A G —-3'

Which of the following DNA mutations is almost certain to result in a shorter than normal mRNA?

a) A→G at position 50 2.0%
b) G→A at position 53 46.5%
c) C→A at position 58 1.6%
d) None of the above 49.5%

Automated analysis approach



Convert MC to open ended questions

The following DNA sequence occurs near the middle of the coding region of a gene.

DNA
5'-AATGAATG*GGAGCCTGAAGGA-3'

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

How will this influence :

1. Replication?
2. Transcription?
3. Translation?

Data collection

- 600 students at CU- Boulder and MSU taking genetics and introductory biology
- Post instruction on central dogma
- 314 responses analyzed
  - 213 - Model testing
  - 101- Model validation

Convert MC to open ended questions

The following DNA sequence occurs near the middle of the coding region of a gene.

DNA
5'-AATGAATG*GGAGCCTGAAGGA-3'

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

How will this influence :

1. Replication?
2. Transcription?
3. Translation?

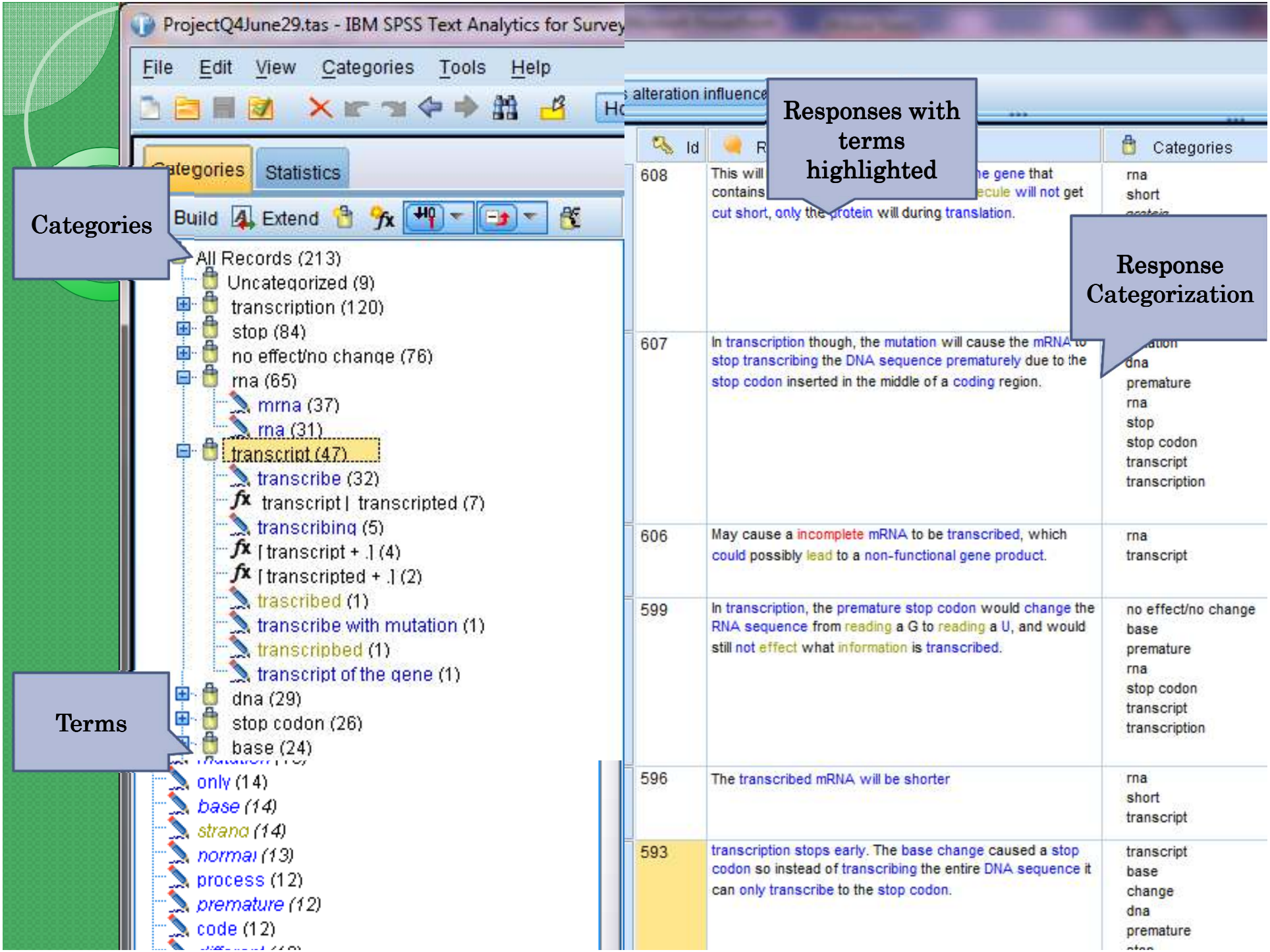
*

Expert rating using holistic rubric for each CR response question

- Two experts rated explanations from correct answers using 3-bin rubric
  - Bin 1: Correct **62.1%**
  - Bin 2: Partly correct explanations with errors in facts or reasoning **10.2%**
  - Bin 3: Totally incorrect **27.7%**

Text analysis

- Software
  - IBM SPSS Text Analysis for Surveys
- Procedure
  - Library Construction
  - Extraction
  - Categorization


ProjectQ4June29.tas - IBM SPSS Text Analytics for Survey
File Edit View Categories Tools Help
Categories Statistics
Build Extend
All Records (213)
Uncategorized (9)
transcription (120)
stop (84)
no effect/no change (76)
rna (65)
mrna (37)
rna (31)
transcript (47)
transcribe (32)
transcript | transcripted (7)
transcribing (5)
[transcript + .] (4)
[transcripted + .] (2)
trascribed (1)
transcribe with mutation (1)
transcripbed (1)
transcript of the gene (1)
dna (29)
stop codon (26)
base (24)
only (14)
base (14)
strand (14)
normal (13)
process (12)
premature (12)
code (12)
Categories
Terms
Responses with terms highlighted
Response Categorization
Id
Categories
608
This will ... gene that contains ... ecule will not get cut short, only the protein will during translation.
rna
short
607
In transcription though, the mutation will cause the mRNA to stop transcribing the DNA sequence prematurely due to the stop codon inserted in the middle of a coding region.
dna
premature
rna
stop
stop codon
transcript
transcription
606
May cause a incomplete mRNA to be transcribed, which could possibly lead to a non-functional gene product.
rna
transcript
599
In transcription, the premature stop codon would change the RNA sequence from reading a G to reading a U, and would still not effect what information is transcribed.
no effect/no change
base
premature
rna
stop codon
transcript
transcription
596
The transcribed mRNA will be shorter
rna
short
transcript
593
transcription stops early. The base change caused a stop codon so instead of transcribing the entire DNA sequence it can only transcribe to the stop codon.
transcript
base
change
dna
premature

Predicting expert ratings using Discriminant Analysis

- Dependent Variable
  - Human Rating Rubric
- Independent Variables
  - Lexical categories

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_n X_n + \varepsilon$$

Categories important for predicting human ratings

Category	Standardized Discriminant Coefficients
Stop codon	-0.321
transcription	0.250
translation	-0.094
rna	-0.019
stop	-0.678
short	-0.663
no effect/no change	0.755

Computer model correctly predicted 81.5 % of human ratings

Wilk's lambda = 0.408 $p<0.001$

Categories important for predicting human ratings

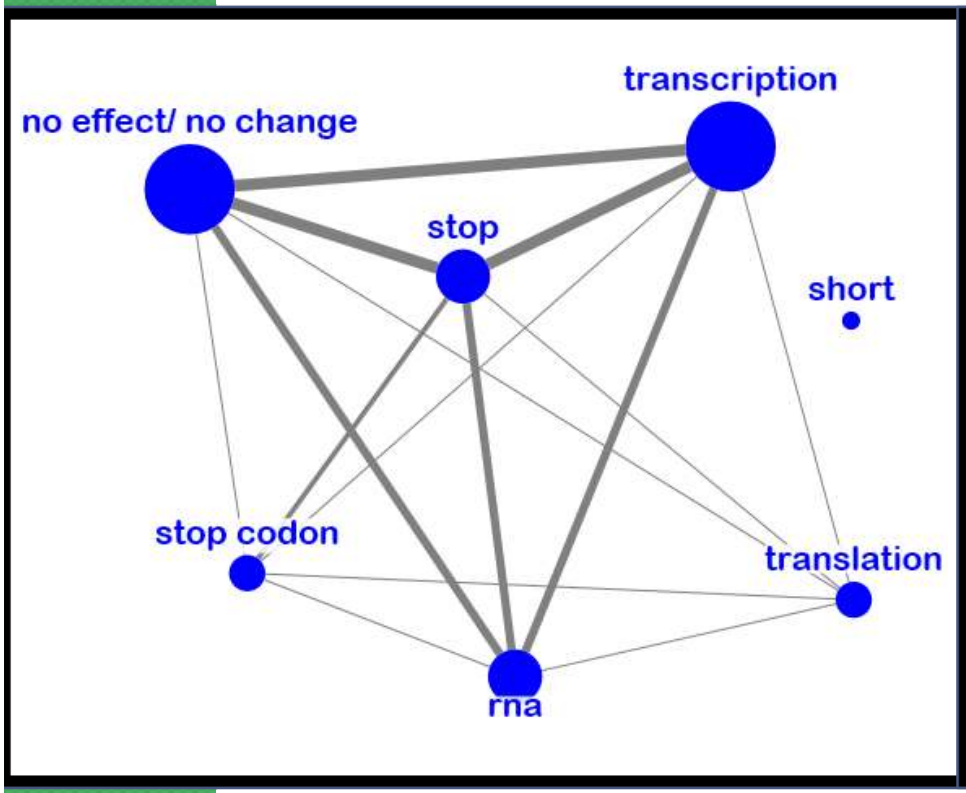
Category	Standardized Discriminant Coefficients
Stop codon	-0.321
transcription	0.250
translation	-0.094
rna	-0.019
stop	-0.678
short	-0.663
no effect/no change	0.755

Categories important for predicting human ratings

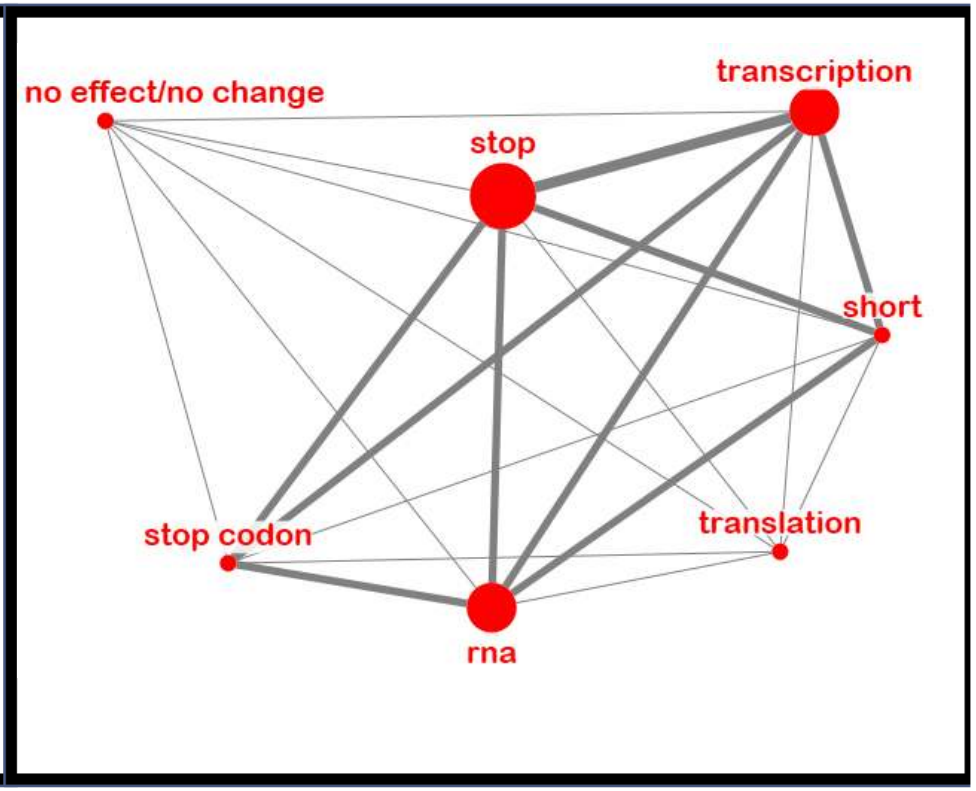
Category	Standardized Discriminant Coefficients
Stop codon	-0.321
transcription	0.250
translation	-0.094
rna	-0.019
stop	-0.678
short	-0.663
no effect/no change	0.755

Visualization of text analysis classifications

Correct Responses



Incorrect Responses



Constructed response questions

The following DNA sequence occurs near the middle of the coding region of a gene.

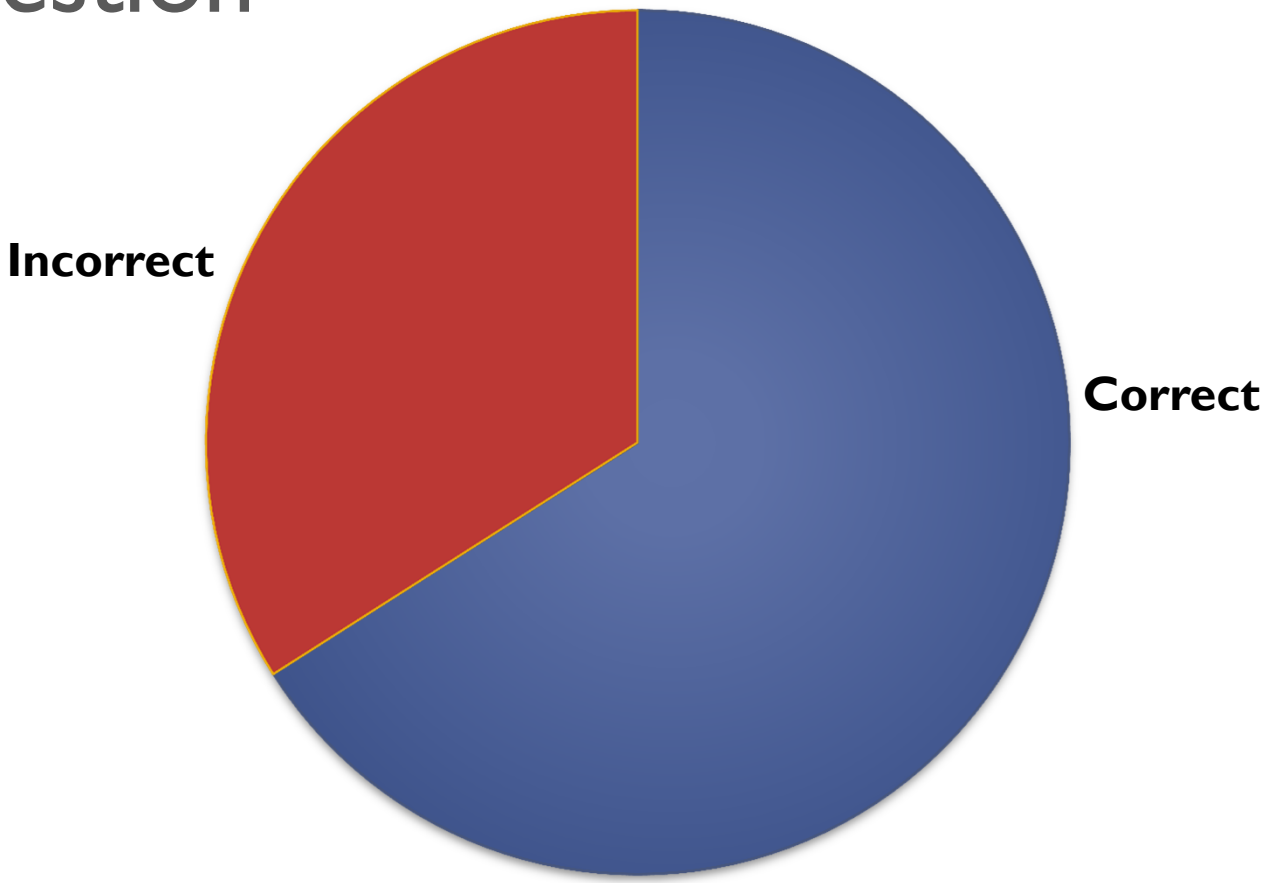
DNA
5'-AATGAATG*GGAGCCTGAAGGA-3'

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

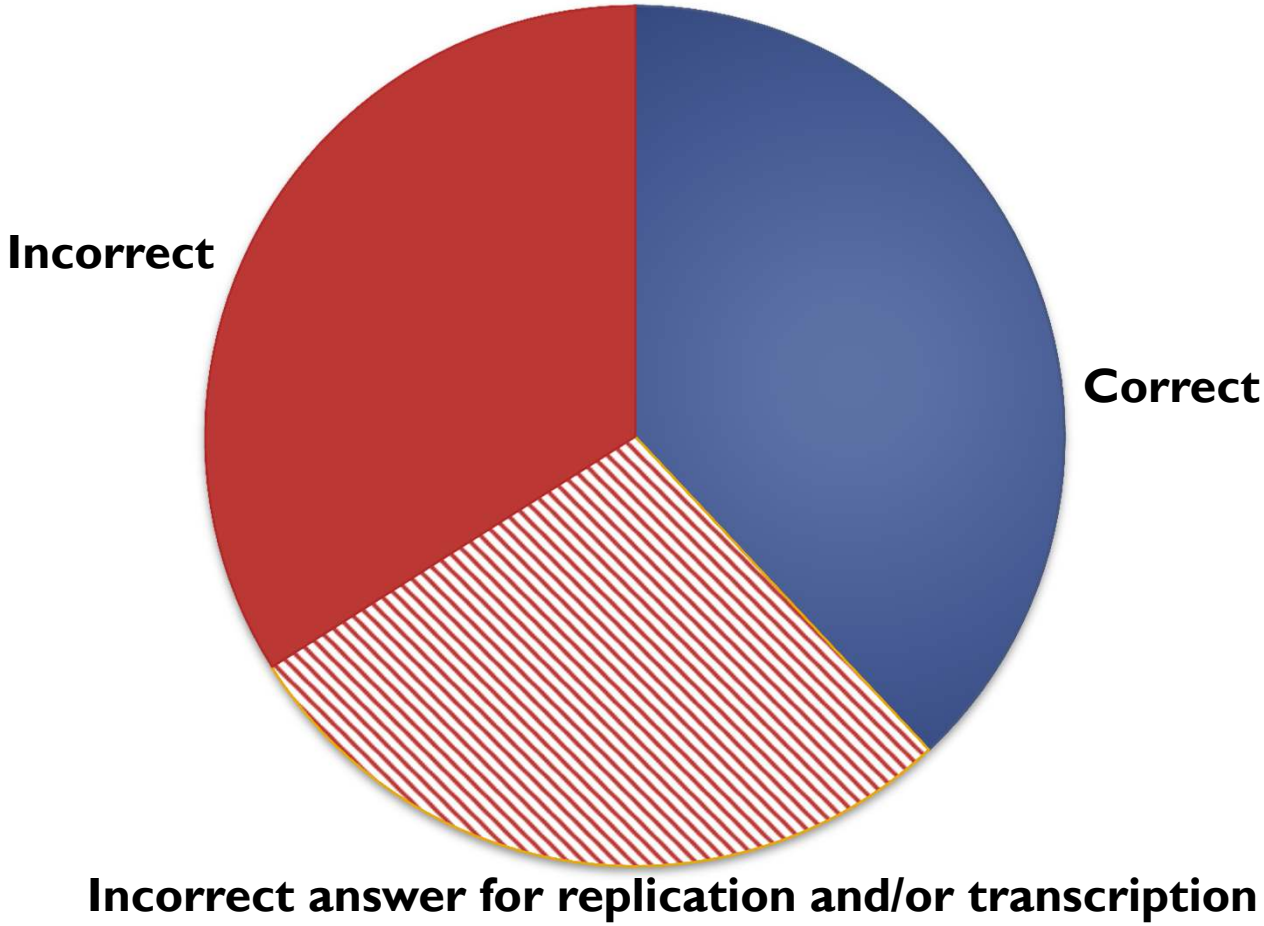
How will this influence

1. Replication?
2. Transcription?
3. Translation?

Students response to Translation question



Some students who answer translation correctly do so for the wrong reason



Summary

- Constructed response assessment with lexical and statistical analysis provide insight into student thinking

- This method enables us to trace student mistakes in central dogma

Future directions

- Investigate the use of broader questions to capture thinking about the steps of central dogma

- Investigate other models of student thinking about central dogma

Questions

Luanna Prevost
prevostl@msu.edu

Mark Urban-Lurain
urban@msu.edu

Website
aacr.crcstl.msu.edu