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

Rosa A. Moscarella[1], Kevin C. Haudek[2], Jennifer K. Knight[3], Alexandria Mazur[1], Karen N. Pelletreau[4], Luanna B. Prevost[5], Michelle K. Smith[4], Matthew Steele[1], Mark Urban-Lurain[1], John E. Merrill[6]

[1]Collaborative Research in Education, Assessment, and Teaching Environments for the fields of Science, Technology, Engineering and Mathematics (CREATE4STEM), Michigan State University. [2]Department of Biochemistry and Molecular Biology, Michigan State University. [3]Department of Molecular, Cellular and Developmental Biology, University of Colorado Boulder. [4]School of Biology and Ecology, University of Maine. [5]Department of Integrative Biology, University of South Florida. [6]Department of Microbiology and Molecular Genetics, Michigan State University.

Paper presented at the 2016 Annual International Conference of the National Association for Research in Science Teaching (NARST), Baltimore, MD, April 14-17, 2016.

Author Note

Correspondence may be addressed to Rosa A. Moscarella (moscarel@msu.edu) at the CREATE for STEM Institute, Michigan State University, 620 Farm Lane, Room 115, East Lansing, MI, 48824.

Abstract

Understanding genetics is fundamental for biological literacy and is broadly recognized as one of the most difficult disciplines in biology. Although a number of published assessment instruments are available to assess students' genetics knowledge, it is unclear to what extent these instruments can measure students' understanding or reasoning. Constructed response (CR) questions, in which students craft responses to questions using their own words, may provide a more authentic assessment of students' understanding and reasoning. Automated scoring has made it possible to use CR assessments in large enrollment classes. We have created CR assessments in which students are asked to predict the effect of a DNA mutation on the processes of replication, transcription, and translation. We analyzed over 4,000 student responses, collected from introductory and upper-level biology classes across five universities in the U.S., using a computer scoring model, trained with 1,031 human scored responses.   Analyses of path maps of students' responses and written answers revealed students' alternative conceptions of each process. The results show that students' understanding of transcription is critical for the overall comprehension of the processes involved in the genetic information flow. Researchers may find our approach useful for evaluating overall students' performance in multiple-part CR questions. We suggest instructors focus on transcription as a keystone concept to understanding the processes underlying the flow of genetic information.

Keywords: lexical analysis, text analysis, automated scoring, genetics education, central dogma

Introduction

Genetics is a fundamental subject for biology literacy (AAAS, 2011), yet it is one of the most challenging areas for students to learn (Bahar, Johnstone, & Hansell, 1999; Marbach-Ad, 2001; M. K. Smith & Knight, 2012; M. K. Smith, Wood, & Knight, 2008; Wright, Fisk, & Newman, 2014). Students at all levels struggle to understand the nature of genetic information in the cells of an organism and how this information flows, is exchanged and stored (Lewis & Wood-Robinson, 2000; Newman, Catavero, & KateWright, 2012; Wood-Robinson, Lewis, & Leach, 2000; Wright et al., 2014). In particular, the central dogma of molecular biology, stated as "*information in nucleic acid can be perpetuated or transferred but the transfer of information into protein is irreversible*" (Crick, 1958), is one topic where students have persistent conceptual difficulties (Jensen, Kummer, & Banjoko, 2013; Newman et al., 2012; M. K. Smith & Knight, 2012; Wright et al., 2014).

Genetic information flow is considered a core concept in Vision and Change (AAAS, 2011), and the lack of understanding of genetics can impede the understanding of other core concepts, for example, the mechanisms of evolution (Kalinowski, Leonard, & Andrews, 2010; Klymkowsky, 2010; White, Heidemann, & Smith, 2013). Our efforts to improve students' understanding of genetics first requires assessing student knowledge.

A growing trend is to assess discipline literacy via multiple-choice concept inventories. These are valuable instruments that can be used to measure student initial knowledge and learning gains. Instruments to assess genetics knowledge include the Genetic Concept Assessment (GCA; M. K. Smith et al., 2008), the Genetic Literacy Assessment Instrument (Bowling et al., 2008), the Biology Concept Inventory (Klymkowsky & Garvin-Doxas, 2008), and the Molecular Biology Capstone Assessment (Couch, Wood, & Knight, 2015). However,

results from these multiple choice tests may not reflect students' actual knowledge because students have learned test taking strategies and prepare for multiple choice tests differently than they prepare for tests using constructed response questions (Stanger-Hall, 2012). Although multiple true/false questions can reveal heterogeneity of student thinking (Couch et al., 2015), it is unclear to what extent multiple choice tests in general can measure students' understanding or reasoning (J. I. Smith & Tanner, 2010). In contrast, constructed response (CR) questions may reflect student understanding more accurately than multiple choice questions because they require students to craft explanations using their own words (Bennett & Ward, 1993; Kuechler & Simkin, 2010). CR items are also a more authentic assessment tool that can potentially reveal novel student conceptions not previously considered and therefore unavailable among the options of multiple choice tests (Birenbaum & Tatsouka, 1987). However, an important limitation of CR assessments is the time and resources needed to evaluate student responses, especially in large enrollment classes. This challenge can be mitigated with the use of computerized lexical analysis and scoring algorithms (Ha, Nehm, Urban-Lurain, & Merrill, 2011; Haudek, Prevost, Moscarella, Merrill, & Urban-Lurain, 2012; J. J. Kaplan, K. C. Haudek, M. Ha, N. Rogness, & D. Fisher, 2014; Magliano & Graesser, 2012; Nehm, Ha, & Mayfield, 2012; Park, Haudek, & Urban-Lurain, 2015; L. Prevost, Haudek, Urban-Lurain, Merrill, & Ieee, 2012; Urban-Lurain, Prevost, Haudek, Henry, Berry, Merrill, et al., 2013).

In this study we investigated college students' ability to correctly differentiate the central dogma processes of replication, transcription, and translation when asked to describe the effect of a mutation generating a stop codon. Previous studies have shown that college students believe that stop codons, which signal the end of polypeptide synthesis during translation, stop transcription, the process of producing RNA from information in DNA (M. K. Smith & Knight,

2012; Wright et al., 2014). We used computerized lexical analysis of CR questions to investigate the challenges undergraduate students have in understanding the impact of this mutation on replication, transcription and translation. We analyzed over 4000 written responses from a heterogeneous sample of undergraduate students. Our analytic approach can be adopted by other education researchers to investigate diverse topics related to teaching and learning science through the implementation of open-ended assessments. Instructors may find these results helpful to inform their teaching and help their students succeed mastering genetic concepts in their classes.

Methods

We created a three-part CR question based on two items from the GCA (M. K. Smith et al., 2008) about the effect of a mutation, which produces an early stop codon, on the processes of replication, transcription and translation (Fig. 1). Correct answers to the questions in Figure 1 should state that the mutation will not affect either the process of replication or transcription, although the product of both processes will carry the mutation. Translation, though, will be affected because the early stop codon would cause the process to terminate sooner than intended, producing a shorter polypeptide.

> 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?

Figure 1. CR questions based on two items from the GCA. The original multiple choice questions can be found in Smith and Knight (2012)

Responses were collected as online homework assignments from university students enrolled in introductory biology courses for science majors (BIO, 7 courses, five institutions, N=3230) and upper level genetics courses (GEN, 2 courses, two institutions, N=818) across five universities in the U.S., during the fall 2014 and spring 2015 semesters.

An instructional intervention about this topic/question was developed by some of the instructors responsible for the courses (Pelletreau et al., in preparation). The same intervention was presented to all participating courses. The questions were administered twice during each semester: post instruction about the central dogma but PRE intervention (BIO n=1610; GEN n=430, and POST intervention (BIO= 1620; GEN=388; hereafter refer to as PRE and POST). Students electronically submitted their responses to each part of the question in separate text boxes in an online Learning Management System. They were explicitly requested not to use any external resources but to answer the questions to the best of their knowledge. Students were awarded participation points for the completion of the assignment, regardless of the correctness of their answers. This study was designated exempt by the University Institutional Review Board (IRB# x10-577).

Students' written responses to each CR part were analyzed using computer models that predict human expert scoring by combining lexical and statistical analyses (Ha et al., 2011; Haudek et al., 2012; J. J. Kaplan, K. C. Haudek, M. Ha, N. Rogness, & D. G. Fisher, 2014; Nehm, Ha, et al., 2012). For this study, the lexical analysis was performed using IBM SPSS Text Analytics for Surveys v. 4.0.1 (IBM Corp). During lexical analysis, key words and phrases are extracted from the text of each response. The extraction procedure can be refined through the creation of custom libraries to include technical terminology that would not otherwise be recognized by the software's linguistic algorithms. Then, words/phrases with similar content and that are relevant to the question are grouped into conceptual *categories* that are defined by researchers with disciplinary expertise. The software uses the definitions of the *categories* to automatically categorize each student's response into one or more of the *categories* (Fig. 2). In this way, each response is now coded into a suite of *categories*. The set of categories created for the lexical analysis of each question part was used to analyze the data, which provides consistent results across all three questions. After each student's response is categorized into the various *categories*, those *categories* are used as independent variables in statistical analyses. In other words, the analysis of each student's response is based on the *categories* associated with their written answers.
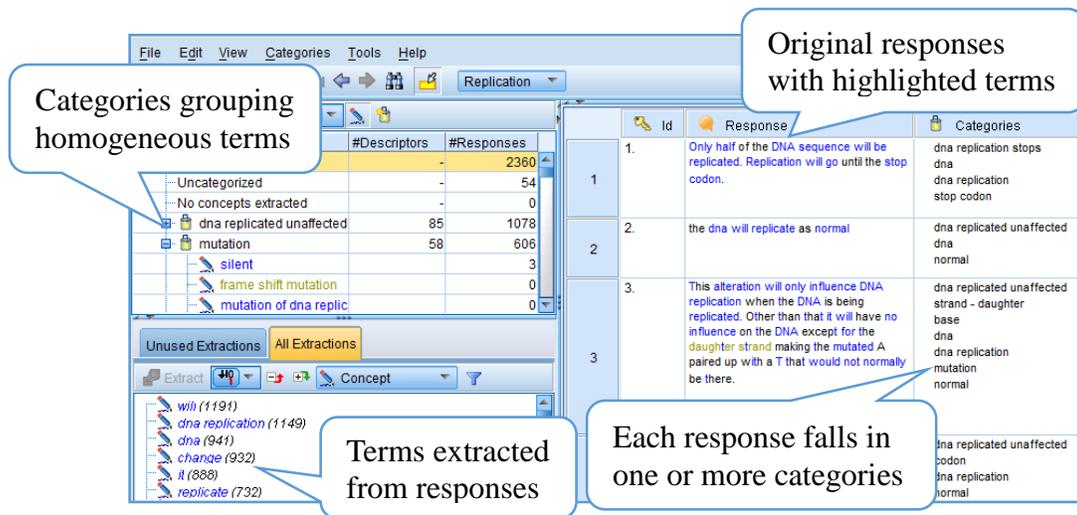
Figure 2. Screenshot of IBM SPSS Text Analytics for Surveys v. 4.0.1 (IBM Corp) to illustrate how words are extracted from student' answers (words in color in right panel) and grouped in categories (upper left panel) that are the variables analyzed in the ensemble scheme. Words extracted, or *terms* are shown in the lower left panel.

We created a three bin holistic scoring rubric for each question that was used to score a set of responses (N = 1031) as completely correct, irrelevant/partially correct, or incorrect (L. B. Prevost, Smith, & Knight, in revision).

Table 1. Examples of students' responses for each rubric definition (explanation in italics)

Rubric Level	Replication	Transcription	Translation
Correct	It will not influence DNA replication. *Response suggests that replication is unaffected by the point mutation.*	There will be no change in transcription but the newly transcribed mRNA sequence will include the mutation. *Response states the point mutation will not cause transcription to end early.*	The amino acid sequence will be shorter than expected. *Response suggests that translation will end early at the new stop codon.*
Incomplete/ Irrelevant	This will cause a mutation. *Incomplete: The response only states that a mutation will occur.*	It could cause the DNA to be fixed or could cause mutations in the cell that could be bad. *Irrelevant: This response does not refer to transcription.*	You would end up getting a different kind of protein. *Incomplete: It is unclear how the protein will be different.*
Incorrect	DNA replication would stop. *Response suggests replication will stop once it reaches the point mutation.*	The mRNA will stop at this point and read no further, so the mRNA will be significantly shorter. *Response suggests that transcription will end early.*	Translation will proceed as normal. *Response suggests that the stop codon will not affect the process of translation.*

The statistical analysis consists of an ensemble of machine learning algorithms developed in R that is used to predict the scoring of student responses (dependent variable) based upon the categories assigned during lexical analysis (independent variables) (Jurka, Collingwood, Boydstun, Grossman, & van Atteveldt, 2012). The ensemble consists of nine classification algorithms, each of which is trained on the set of human scored student responses. When a new student response is fed to the system, each individual classification algorithm returns a set of membership probabilities for each rubric level. The rubric level that has the highest aggregate classification probability across the ensemble is assigned as the score for that student response. The interrater reliability between human scores and computer predictions was within the range

usually seen between human raters using well calibrated rubrics (Landis & Koch, 1977; Nehm, Ha, et al., 2012).

We used two approaches to visualize students' performance on each question part. First, we created stacked bar plots to show the proportion of answers in each rubric level for each question part PRE and POST. Second, we created path maps that trace all student responses through each question part. To do so, the responses of each student were paired by question part and rubric level, and the frequency of each question/rubric level pair was determined. These maps allow us to visualize the general pattern of students' performance by looking at the frequencies of responses in all combinations of correct, irrelevant/incomplete, and incorrect for the three parts of the questions.

Additionally, we performed a qualitative analysis of student's written responses, following an inductive approach (Bryman, 2012), to characterize alternative conceptions students have for each of the process encompassed in the central dogma.

Findings

Students in both BIO and GEN showed similar trends in the proportion of answers scored correct, incomplete/irrelevant or incorrect, although GEN had larger percentages of correct answers, for both PRE and POST (Fig. 3). In both courses, we found that the question that had the largest percentage of correct answers for the PRE data collection was translation, while transcription had the lowest.   POST data showed improvement in all three question parts for both courses. It is noteworthy that for the transcription data, though the proportion of correct answers increased, it still had the largest proportion of incorrect answers in both courses.

Moreover, for the BIO data the percentage of incorrect answers remained almost the same as PRE, showing little gain from instruction.
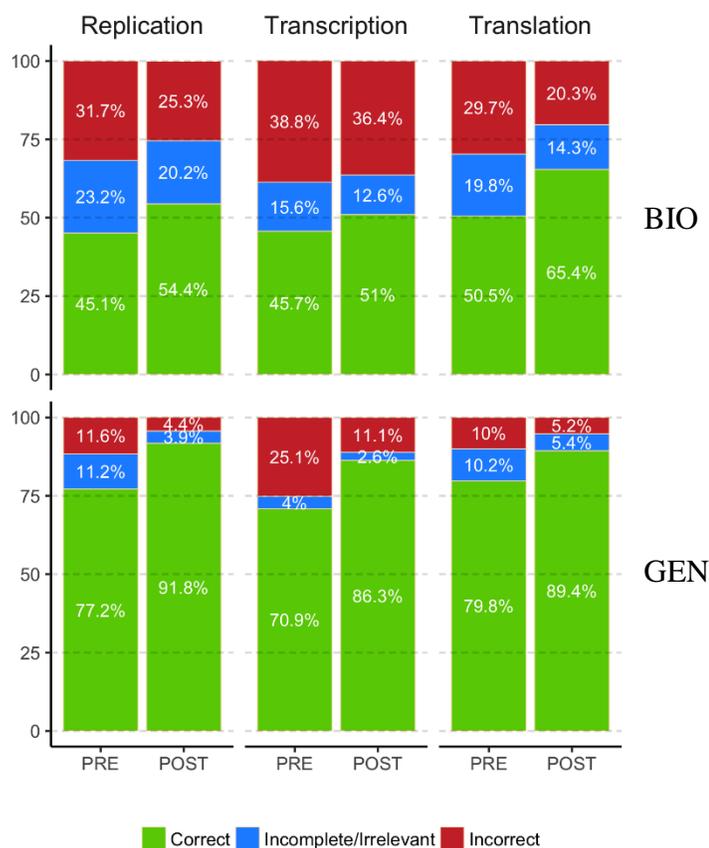

Figure 3. Stacked bar plots showing the percentages of answers in each rubric level, PRE and POST, for both BIO and GEN

We use "path maps" to represent the relationships among students' responses to each question part (i.e. correct, incomplete/irrelevant, incorrect; Figs. 4 and 5). The size and color of the circles represents the percentage of the students' responses from the total responses that falls in that category. The color and thickness of the arrows represent the percentage of responses from the total that were classified in the other categories.

We can focus on any one of the question parts (Fig. 4 focuses on transcription) and explore the patterns for the other question parts. Alternatively, we can focus on a rubric level (Fig. 5 focuses on correct responses) and investigate the patterns for the different question parts. For instance, the top panel of Figure 4 is focused on student's correct responses for transcription, i.e., 100% of those responses. From those students in BIO, 66% responded correctly to replication and 65% to translation. The percentage of students that responded either incomplete /irrelevant or incorrect to those question parts was <20%. In GEN the pattern is more pronounced: over 90% of the students that responded correctly to transcription also responded correctly to replication and translation. In BIO, from all the students whose transcription response was scored as incomplete/irrelevant, similar percentages responded to replication and translation either correct, incomplete/irrelevant or incorrect (left middle panel Fig. 4). Interestingly, a large percentage of students whose transcription answer was scored incorrect were scored correct for translation (BIO= 56%, GEN= 76%, bottom panel Fig. 4)
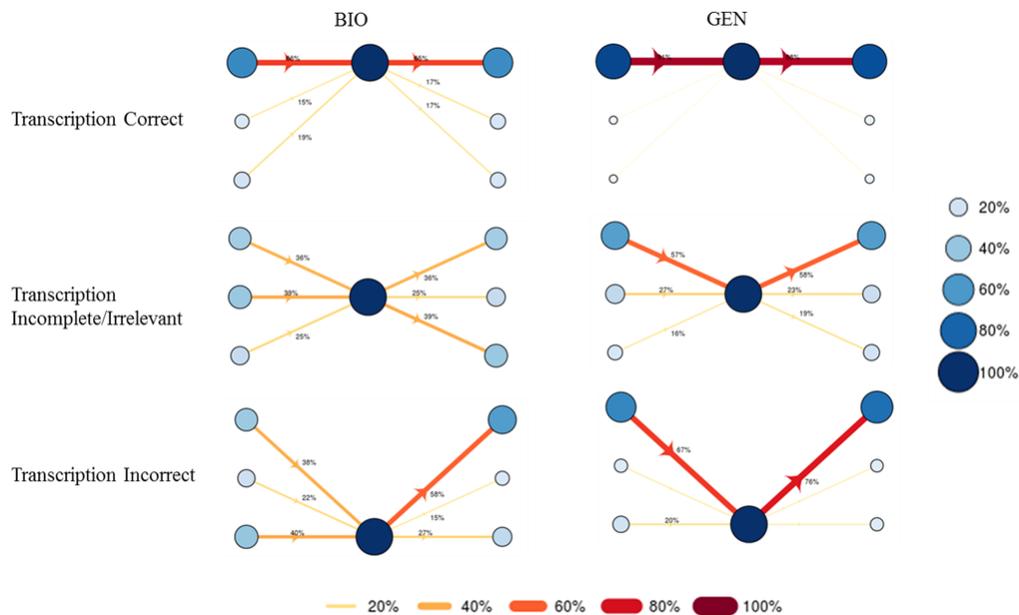
Figure 4. Path map showing patterns of students answers for transcription. Here, we show how students answered to both replication and translation when they answered to transcription either correct, incomplete/irrelevant, or incorrect. For instance, the bottom of the figure shows the path students followed when they answered to transcription incorrectly. For BIO, similar percentages of students answered to replication either correctly or incorrectly (~40%), while almost 60% responded correctly to translation.
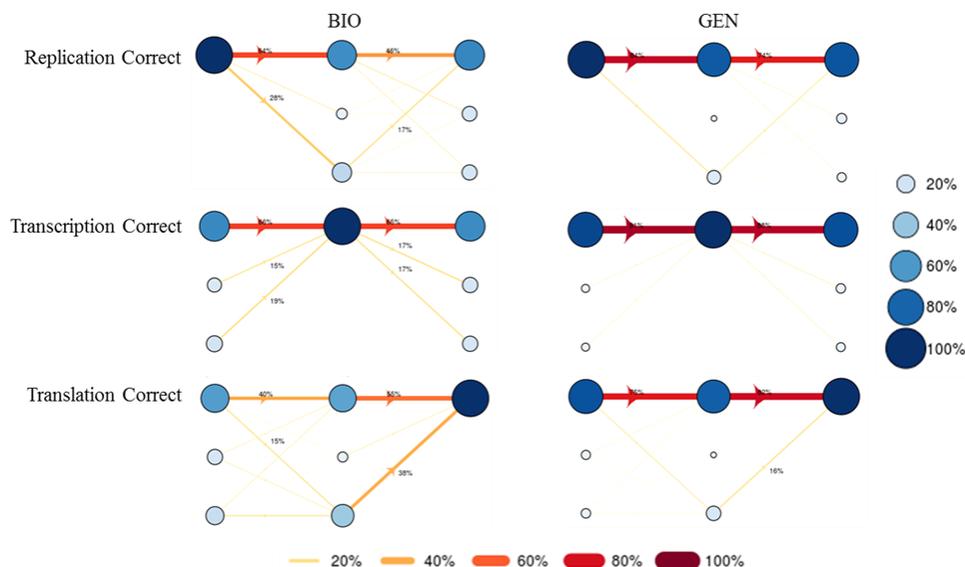
Figure 5. Path map showing the patterns of students correct answers for each process. The top of the figure shows the path students followed when they answered to replication correctly. For GEN, 84% of students who answered to replication correctly also answered correctly to transcription, and 74% responded correctly to translation. Less than 20% had responses either incomplete/irrelevant or incorrect in the other two processes.

Representing the classifications of student responses using a path map reveals that students' understanding of transcription seems to be key for the full comprehension of the flow of genetic information (Fig. 4). As it can be seen in the path maps, a larger percentage of students with a correct answer for transcription also responded correctly to replication and translation than those who have transcription either incomplete/irrelevant or incorrect responses, in both courses (Fig. 4). For the BIO data set, >65% of students who responded correctly to transcription also responded correctly to replication and translation; while of students who responded correctly to replication, 64% responded correctly to transcription and 46% to translation; from those who responded correctly to translation, only 40% responded correctly to

replication and 55% to transcription. For the GEN data set the results are more homogeneous, but still a larger percentage of correct responses is observed in the group of students who responded correctly to transcription.

Qualitative analysis of written answers from the POST data focused on the trends observed in the path maps with >20% of students answers following a particular path. In each example, the responses for replication, transcription and translation were given by the same student. This analysis shows that students have several common alternative conceptions about each process. Some of these alternative conceptions are shown below (Rep=replication, Trc=transcription, Tsl=translation; co=correct, ii=incomplete/irrelevant, in=incorrect):

1. Some students think that a shorter protein at the end of translation is due to the fact that the mRNA was shortened during transcription because of the stop codon: **Rep-co:** "*The alteration will not affect DNA replication. DNA replication does not involve the use of STOP codons*" **Trc-in:** "*The RNA polymerase will stop coding prematurely. The mRNA strand will be short, and it will code for an incomplete protein.*" **Tsl-co:** "*The STOP codon will stop translation of the protein. Since this error already occurred previously in transcription, an incomplete protein will result from the already short mRNA strand.*" [Note: The rubric focuses on the responses to the particular process. In this case, the answer is still correct despite the wrong premise for the answer]

2. Students still can reach the correct answer for translation (the process will be halt and therefore a shorter protein will be made, likely dysfunctional) by having wrong ideas about the effect of such mutation during both replication and transcription: **Rep-in:** "*When the DNA polymerase gets to the A that was a G it will stop because it is now a stop codon.*" **Trc-in:** "*When the DNA polymerase gets to the newly formed stop codon*

transcription will stop leaving the rest of the DNA not coded for." **Tsl-co:** "*After the DNA has been transcribed (or the small part that has) translation will then translate the mRNA into a protein. The protein will not be the correct protein because it was stopped prematurely.*"

3. Other students thought a stop codon will stop transcription but did not clearly explain what the effect on translation would be: **Rep-co:** "*It will not affect it.*" **Trc-in:** "*Transcription will be stopped prematurely.*" **Tsl-ii:** "*It will be translated incorrectly.*"

4. There were students that thought that translation will not be affected by this mutation, although the other processes will: **Rep-in:** "*This will tell the DNA to stop before it really is supposed to. This will make it shorter than it is supposed to be.*" **Trc-in:** "*The mRNA will be shorter than it is supposed to be and will not contain all of the correct amino acids.*" **Tsl-in:** "*Translation will not be affected.*"

5. In other instances, students did not make a clear statement that a stop codon would stop translation prematurely, but instead referred to a change in the amino acid sequence: **Rep-co:** "*This alteration will not influence DNA replication.*" **Trc-co:** "*This alteration will not influence transcription.*" **Tsl-ii:** "*This alteration will influence the amino acid sequence made during translation.*"

6. Some other responses suggest that the effect of this mutation on replication is independent of the other two processes: **Rep-in:** "*it will stop the DNA from replicating any further*" **Trc-co:** "*it will change the G to an A so when it is being copied to RNA*" **Tsl-co:** "*the stop codon stops the translation into amino acids therefore creating a whole different amino acid than what it was supposed to be*"

7. Some responses suggested that none of the three processes could take place: **Rep-in:** "*Replication can't occur without transcription*" **Trc-in:** "*It will stop transcription prematurely, making the cell nonfunctional.*" **Tsl-in:** "*Translation can't occur*"

8. Some students believe that the short protein is caused by a replication that stop prematurely, producing a shorter DNA that was transcribed into a shorter mRNA and therefore translated into a shorter protein: **Rep-in:** "*The strand will be completed short*" **Trc-in:** "*It will only transcribe the first part of the DNA.*" **Tsl-co:** "*The RNA strand will be translated off a short DNA strand and will result in a shortened protein.*" [See note above]

9. Often, students answered to the effect of the mutation in translation without addressing how replication and transcription would be affected by this mutation: **Rep-ii:** "*If the DNA strand now reads 5' AATGAATGAGAGCCTGAAGGA3' Then the complementary mRNA strand will read 3' UUACUUACUCUCGGACUUCCU 5 The nonsense mutation will result in a nonfunctional protein*" **Trc-ii:** "*If this is a nonsense mutation, it will result in a shortened protein that is probably nonfunctional or if it is functional then it is highly impaired. (…)*" **Tsl-co:** *If this is a stop codon, then the translated protein will probably not be functional, as it will be much shorter and with less amino acids than intended*"

10. Some students' answers did not address the question; answers were irrelevant or incomplete: **Rep-ii:** "*It can be descibed as a nonsense mutation which doesn allow the correct amino acids to be made, creating an incomplete protein.*" **Trc-ii:** "*It creates an error in the mRNA replication, the mRNA will be make incorrect amino acids.*" **Tsl-ii:**

When the mRNA goes through translation to code for proteins needed it wont have all the correct amino acids to create the necessary proteins."

11. Some students responded correctly to replication and translation but their answer to transcription was unclear: **Rep-co:** "*No effect on the replication process, except that the strand formed from this template strand will have a T where it otherwise would have had a C.*" **Trc-ii:** "*Could affect boundary between intron/exon, resulting in more/less of the gene being expressed as mRNA than should be.*" **Tsl-co:** "*mRNA with the new Stop codon will cease translation earlier than otherwise would have. Protein will be shorter than it should be (truncated protein).*"

Discussion

In this study we used automated scoring to score a three-part constructed response question in several large enrollment biology courses, both introductory and upper division, across five universities in the U.S. We have demonstrated that this methodology makes it feasible to integrate writing in the classroom for formative assessment (Ha et al., 2011; Haudek et al., 2012; Jennifer J. Kaplan et al., 2014; Moscarella et al., 2008; Nehm, Beggrow, Opfer, & Ha, 2012; Nehm, Ha, et al., 2012; L. Prevost et al., 2012; Urban-Lurain et al., 2009; Urban-Lurain, Prevost, Haudek, Henry, Berry, & Merrill, 2013; Weston, Haudek, Prevost, Merrill, & Urban-Lurain, 2014), which accomplishes two objectives: 1) helping students organize their knowledge and 2) providing instructors with richer insight about student thinking in the domain. In this article we elaborate on the second objective, namely using this type of assessment to obtain greater insight into students understanding of a particular subject- in this case, genetic information flow.

Genetics is a complex subject that students struggle to master (e.g., Bahar et al., 1999; Marbach-Ad, 2001). Our results not only corroborate that students have difficulties differentiating among the key molecular processes that underlie the central dogma, both in introductory biology and upper level genetics courses, but also provide greater insight into student reasoning and misconceptions. Through this analysis we were able to identify the process of transcription as a critical concept for the full understanding of the flow of genetic information.

We came to this conclusion about transcription after distinguishing two patterns in the data: First, the fact that transcription held the largest proportion of incorrect responses even after intervention (Fig. 3). Many students believe that a stop codon stops transcription, a persistent misconception previously noticed by Smith and Knight (2012). Second, the pattern that made clear that transcription may be a keystone concept emerged from the path maps. More than 65% of the students who responded correctly to the transcription part of the question in BIO also responded correctly to replication and translation. For GEN, this percentage is above 88% (Fig. 4). This pattern is not seen in students who answered correctly either replication or translation. In those cases, the percentage of students who have also a correct response for the other two processes is much lower (Fig. 5).

An interesting conceptual issue found in our results, not previously reported in the literature, was that some students thought that a stop codon could also stop replication. This was a more prevailing problem in the PRE data. When students were first presented with a problem in which they had to predict the effect of a mutation in replication, two potential problems were revealed: 1) students do not have a clear conception of what codons are and 2) failed to correctly predict the impact of mutations on the processes of replication, transcription and translation.

Students' struggle with understanding what codons are could help explain why students harbor the idea that stop codons terminate the synthesis of RNA during transcription. While other assessments target this common misconception (eg., GCA; M. K. Smith et al., 2008), the CR assessments allow us to explore what students think about stop codons and their effect on the different processes involved in the central dogma, offering a broader picture of diverse student conceptions.

The design of appropriate instructional interventions requires knowing the nature of students' learning difficulties. The path maps make it possible to visualize students' alternative conceptions. This approach, along with the qualitative analysis of written responses, gave us a unique insight into the overall students' performance across all the parts of the question. The path maps allow instructors to analyze students' responses to multiple-part questions and rapidly identify specific concepts that may be affecting students' global comprehension of a science phenomenon. This information can be used to plan instructional interventions and to inform curriculum development.

Our results suggest that transcription may require special instructional attention. We are currently investigating the details of students' struggles when learning transcription. We are analyzing interviews of students with different levels of genetics understanding to characterize students' conceptions, misconceptions and thinking about the processes of replication, transcription and translation. Meanwhile, we recommend that instructors use the results of this study and explicitly address in their classes the distinction between DNA replication and protein synthesis and have students predict and evaluate the effects of mutations at different levels. Students should be able to describe the processes and products of transcription and translation,

know how these processes are initiated and terminated, and explain how the products of transcription are used in translation.

Comparing the results obtained for BIO and GEN exposes the need of finding ways to make genetics more accessible for introductory biology students. McDonnell, Barker & Wieman (2016) compared two groups of students, both taught by the same instructor using the same active learning practices and only differing in the sequence in which concepts and definitions were presented. They found that that the group in which the presentation of concepts preceded the definitions showed a better understanding of the new concepts. An interesting hypothesis for future inquiry is to investigate whether reducing emphasis on definitions in lower level classes will help students to focus more on the deep understanding of the content. This needs to be aligned with teaching practices and assessments that promote higher level thinking (Momsen, Long, Wyse, & Ebert-May, 2010). Considering that introductory biology courses might be the only courses in which some students are exposed to genetic concepts, along with the fact that genetic literacy is gaining increased importance, instructors should consider what students need to learn to be literate in genetics. We join the call of Duncan, Rogat, & Yarden (2009), Elmesky (2013), Shea & Duncan (2013), and Wright et al. (2014) for the development of learning progression schemes to help instructors improve their students' learning of genetics.

References

AAAS. (2011). Vision and change: A call to action. Washington, DC: American Association for
  the Advancement of Science.

Bahar, M., Johnstone, A. H., & Hansell, M. H. (1999). Revisiting learning difficulties in biology.
  Journal of Biological Education, 33(2), 84-86. doi:10.1080/00219266.1999.9655648

Bennett, R. E., & Ward, W. C. (1993). *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment*. Hillsdale, NJ: L. Erlbaum Associates.

Birenbaum, M., & Tatsouka, K. K. (1987). Open-ended versus multiple-choice response formats - It does make a difference for diagnostic purposes. *Applied Psychological Measurement, 11*, 329-341.

Bowling, B. V., Acra, E. E., Wang, L., Myers, M. F., Dean, G. E., Markle, G. C., . . . Huether, C. A. (2008). Development and evaluation of a genetics literacy assessment instrument for undergraduates. *Genetics, 178*(1), 15-22. doi:10.1534/genetics.107.079533

Bryman, A. (2012). *Social Research Methods (4 ed.)*. Oxford: Oxford University Press.

Couch, B. A., Wood, W. B., & Knight, J. K. (2015). The Molecular Biology Capstone Assessment: A Concept Assessment for Upper-Division Molecular Biology Students. *Cbe-Life Sciences Education, 14*(1). doi:10.1187/cbe.14-04-0071

Crick, F. H. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology, 12*, 138-163.

Duncan, R. G., Rogat, A. D., & Yarden, A. (2009). A Learning Progression for Deepening Students' Understandings of Modern Genetics Across the 5th-10th Grades. *Journal of Research in Science Teaching, 46*(6), 655-674. doi:10.1002/tea.20312

Elmesky, R. (2013). Building Capacity in Understanding Foundational Biology Concepts: A K-12 Learning Progression in Genetics Informed by Research on Children's Thinking and Learning. *Research in Science Education, 43*(3), 1155-1175. doi:10.1007/s11165-012-9286-1

Ha, M. S., Nehm, R. H., Urban-Lurain, M., & Merrill, J. E. (2011). Applying Computerized-Scoring Models of Written Biological Explanations across Courses and Colleges: Prospects and Limitations. *Cbe-Life Sciences Education, 10*(4), 379-393. doi:10.1187/cbe.11-08-0081

Haudek, K. C., Prevost, L. B., Moscarella, R. A., Merrill, J., & Urban-Lurain, M. (2012). What Are They Thinking? Automated Analysis of Student Writing about Acid-Base Chemistry in Introductory Biology. *Cbe-Life Sciences Education, 11*(3), 283-293. doi:10.1187/cbe.11-08-0084

Jensen, J. L., Kummer, T. A., & Banjoko, A. (2013). Assessing the Effects of Prior Conceptions on Learning Gene Expression *Journal of College Science Teaching, 42*(4), 82-91.

Jurka, T. P., Collingwood, L., Boydstun, A. E., Grossman, E., & van Atteveldt, W. (2012). RTextTools: Automatic Text Classification via Supervised Learning. R package version 1.3.9. Retrieved from http://CRAN.R-project.org/package=RTextTools

Kalinowski, S. T., Leonard, M. J., & Andrews, T. M. (2010). Nothing in Evolution Makes Sense Except in the Light of DNA. *Cbe-Life Sciences Education, 9*(2), 87-97. doi:10.1187/cbe.09-12-0088

Kaplan, J. J., Haudek, K. C., Ha, M., Rogness, N., & Fisher, D. (2014). Using lexical analysis software to assess student writing in statistics. (Vol. 8): *Technology Innovations in Statistics Education*.

Kaplan, J. J., Haudek, K. C., Ha, M., Rogness, N., & Fisher, D. G. (2014). Using lexical analysis software to assess student writing in statistics. *Technology Innovations in Statistics Education, 8*(1).

Klymkowsky, M. W. (2010). Thinking about the Conceptual Foundations of the Biological Sciences. *Cbe-Life Sciences Education, 9*(4), 405-407. doi:10.1187/cbe.10-04-0061

Klymkowsky, M. W., & Garvin-Doxas, K. (2008). Recognizing student misconceptions through Ed's Tools and the biology concept inventory. *Plos Biology, 6*(1), 14-17. doi:10.1371/journal.pbio.0060003

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*(1), 55-73.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Lewis, J., & Wood-Robinson, C. (2000). Genes, chromosomes, cell division and inheritance - do students see any relationship? *International Journal of Science Education, 22*(2), 177-195. doi:10.1080/095006900289949

Magliano, J. P., & Graesser, A. C. (2012). Computer-based assessment of student-constructed responses. *Behavior Research Methods, 44*(3), 608-621. doi:10.3758/s13428-012-0211-3

Marbach-Ad, G. (2001). Attempting to break the code in student comprehension of genetic concepts. *Journal of Biological Education, 35*(4), 183-189. doi:10.1080/00219266.2001.9655775

McDonnell, L., Barker, M. K., & Wieman, C. (2016). Concepts First, Jargon Second Improves Student Articulation of Understanding. *Biochemistry and Molecular Biology Education, 44*(1), 12-19. doi:10.1002/bmb.20922

Momsen, J. L., Long, T. M., Wyse, S. A., & Ebert-May, D. (2010). Just the Facts? Introductory Undergraduate Biology Courses Focus on Low-Level Cognitive Skills. *CBE Life Science Eucation, 9*(4), 435-440. doi:10.1187/cbe.10-01-0001

Moscarella, R. A., Urban-Lurain, M., Merritt, B., Long, T., Richmond, G., Merrill, J., . . . Wilson , C. (2008). *Understanding undergraduate students' conceptions in science: Using lexical analysis software to analyze students' constructed responses in biology.* Paper presented at the National Association for Research in Science Teaching Annual International Conference, Baltimore, MD.

Nehm, R. H., Beggrow, E. P., Opfer, J. E., & Ha, M. S. (2012). Reasoning About Natural Selection: Diagnosing Contextual Competency Using the ACORNS Instrument. *American Biology Teacher, 74*(2), 92-98. doi:10.1525/abt.2012.74.2.6

Nehm, R. H., Ha, M. S., & Mayfield, E. (2012). Transforming Biology Assessment with Machine Learning: Automated Scoring of Written Evolutionary Explanations. *Journal of Science Education and Technology, 21*(1), 183-196. doi:10.1007/s10956-011-9300-9

Newman, D. L., Catavero, C. M., & KateWright, L. (2012). Students Fail to Transfer Knowledge of Chromosome Structure to Topics Pertaining to Cell Division. *Cbe-Life Sciences Education, 11*(4), 425-436. doi:10.1187/cbe.12-01-0003

Park, M., Haudek, K., & Urban-Lurain, M. (2015). *Computerized lexical analysis of students' written responses for diagnosing conceptual understanding of energy*. Paper presented at the NARST 2015 Annual International Conference, Chicago, IL.

Pelletreau, K. N., Andrews, T., Armstrong, N., Bedell, M., Dastoor, F., Dean, N., . . . Smith, M. K. (in preparation). Using a clicker-based case study to untangle student thinking about the stages of the central dogma. *CourseSource*.

Prevost, L., Haudek, K., Urban-Lurain, M., Merrill, J., & Ieee. (2012). Examining student constructed explanations of thermodynamics using lexical analysis *2012 Frontiers in Education Conference*.

Prevost, L. B., Smith, M. K., & Knight, J. K. (in revision). Using student writing and lexical analysis to reveal student thinking about the role of stop codons. *CBE-Life Science Education*.

Shea, N. A., & Duncan, R. G. (2013). From Theory to Data: The Process of Refining Learning Progressions. *Journal of the Learning Sciences, 22*(1), 7-32. doi:10.1080/10508406.2012.691924

Smith, J. I., & Tanner, K. (2010). The Problem of Revealing How Students Think: Concept Inventories and Beyond. *Cbe-Life Sciences Education, 9*(1), 1-5. doi:10.1187/cbe.09-12-0094

Smith, M. K., & Knight, J. K. (2012). Using the Genetics Concept Assessment to Document Persistent Conceptual Difficulties in Undergraduate Genetics Courses. *Genetics, 191*(1), 21-+. doi:10.1534/genetics.111.137810

Smith, M. K., Wood, W. B., & Knight, J. K. (2008). The Genetics Concept Assessment: A New Concept Inventory for Gauging Student Understanding of Genetics. *Cbe-Life Sciences Education, 7*(4), 422-430. doi:10.1187/cbe.08-08-0045

Stanger-Hall, K. F. (2012). Multiple-choice exams: An obstacle for higher-level thinking in introductory science classes. *CBE-Life Sciences Education, 11*(3), 294-306.

Urban-Lurain, M., Moscarella, R. A., Haudek, K. C., Giese, E., Sibley, D. F., & Merrill, J. E. (2009). *Beyond multiple choice exams: Using computerized lexical analysis to*

understand students' conceptual reasoning in STEM disciplines. Paper presented at the ASEE/IEEE Frontiers in Education Conference, San Antonio, TX.

Urban-Lurain, M., Prevost, L., Haudek, K. C., Henry, E. N., Berry, M., & Merrill, J. E. (2013). Using Computerized Lexical Analysis of Student Writing to Support Just-in-Time Teaching in Large Enrollment STEM Courses *2013 IEEE Frontiers in Education Conference*.

Urban-Lurain, M., Prevost, L., Haudek, K. C., Henry, E. N., Berry, M., Merrill, J. E., & Ieee. (2013). Using Computerized Lexical Analysis of Student Writing to Support Just-in-Time Teaching in Large Enrollment STEM Courses *2013 Ieee Frontiers in Education Conference*.

Weston, M., Haudek, K. C., Prevost, L. B., Merrill, J. E., & Urban-Lurain, M. (2014). *Examining the impact of question surface features on students' answers to constructed response questions in biology*. CREATE4STEM Mini Conference. Michigan State University. East Lansing, MI.

White, P. J. T., Heidemann, M. K., & Smith, J. J. (2013). A New Integrative Approach to Evolution Education. *BioScience, 63*(7). doi:10.1525/bio.2013.63.7.11

Wood-Robinson, C., Lewis, J., & Leach, J. (2000). Young people's understanding of the nature of genetic information in the cells of an organism. *Journal of Biological Education, 35*(1), 29-36. doi:10.1080/00219266.2000.9655732

Wright, L. K., Fisk, J. N., & Newman, D. L. (2014). DNA -> RNA: What Do Students Think the Arrow Means? *Cbe-Life Sciences Education, 13*(2), 338-348. doi:10.1187/cbe.CBE-13-09-0188