

ARTICLE

Introductory biology undergraduate students' mixed ideas about genetic information flow

Juli D. Uhl¹  | Kamali N. Sripathi²  | Jenifer N. Saldanha^{1,3} |
 Rosa A. Moscarella⁴  | John Merrill⁵ | Mark Urban-Lurain¹  |
 Kevin C. Haudek^{1,6} 

¹CREATE for STEM Institute, Michigan State University, East Lansing, Michigan

²UC Davis Genome Center, University of California, Davis, Davis, California

³Lyman Briggs College, Michigan State University, East Lansing, Michigan

⁴Biology Department, University of Massachusetts Amherst, Amherst, Massachusetts

⁵Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan

⁶Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan

Correspondence

Juli D. Uhl, CREATE for STEM Institute, Michigan State University, 620 Farm Lane, Room 115 Erickson Hall, East Lansing, MI 48824, USA.
 Email: uhljuli@msu.edu

Funding information

National Science Foundation, Grant/Award Number: DUE grant 1323162

Abstract

The core concept of genetic information flow was identified in recent calls to improve undergraduate biology education. Previous work shows that students have difficulty differentiating between the three processes of the Central Dogma (CD; replication, transcription, and translation). We built upon this work by developing and applying an analytic coding rubric to 1050 student written responses to a three-question item about the CD. Each response was previously coded only for correctness using a holistic rubric. Our rubric captures subtleties of student conceptual understanding of each process that previous work has not yet captured at a large scale. Regardless of holistic correctness scores, student responses included five or six distinct ideas. By analyzing common co-occurring rubric categories in student responses, we found a common pair representing two normative ideas about the molecules produced by each CD process. By applying analytic coding to student responses preinstruction and postinstruction, we found student thinking about the processes involved was most prone to change. The combined strengths of analytic and holistic rubrics allow us to reveal mixed ideas about the CD processes and provide a detailed picture of which conceptual ideas students draw upon when explaining each CD process.

KEYWORDS

information flow and transfer, rubrics, student thinking, undergraduate

1 | INTRODUCTION

Several national groups have made calls to reform science education to emphasize critical thinking, rather than memorization of facts. These calls include identifying core concepts that all undergraduates should learn.^{1,2}

Vision and Change in Undergraduate Biology Education: A Call to Action (hereafter *Vision and Change*) identified five core concepts; among them is the genetic concept of information flow, exchange, and storage.³ Previous work identified gene expression as critical for understanding information flow.^{4,5}

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. Biochemistry and Molecular Biology Education published by Wiley Periodicals LLC on behalf of International Union of Biochemistry and Molecular Biology.

The Central Dogma (CD) of molecular biology provides a framework for understanding gene expression in terms of three processes, replication, transcription, and translation. Understanding the genetic aspects of information flow supports understanding of the other *Vision and Change* core concepts including evolution.^{6,7} Gene expression is an essential part of a genetics education which is increasingly focused on concepts foundational to “omics” (e.g., metabolomics, proteomics) research.^{8–10} Undergraduates on a variety of career paths will interact with genetic data and thus must understand the mechanism of gene expression, that is, the CD, as a basis for understanding the significance of genomic technologies.

Gene expression is difficult to learn. Understanding and distinguishing between the molecules produced and processes involved are challenging^{11–17} and some misconceptions persist after instruction.^{18,19} Technical genetics and biology vocabulary are a source of confusion.^{12,20,21} Students may incorrectly apply the three-nucleotide code (codons) to the processes, or do not understand how gene structure affects gene expression.^{14,19,22–24} Furthermore, students who have completed introductory genetics retain mostly low-level factual information; for example, the names of the bases or monomers that make up polymers.¹¹ Focusing on low-level knowledge relating to the CD is in conflict with educational reforms focused on practices and conceptual understanding.

To help expand student understanding of gene expression from memorized facts to conceptual understanding, instructors must first identify what students know using formative assessments.²⁵ Instructors might evaluate student understanding using multiple-choice (MC) assessment through the use of personal response systems (PRS or “clickers”). However, students can arrive at a correct MC response by ruling out possibilities or guessing. In such instances, MC questions may hinder critical thinking and inaccurately or incorrectly measure student learning.^{26–28} Constructed response (CR) questions require students to respond in their own words. Responses to CR questions are thought to better diagnose student misconceptions and can contain mixtures of normative and nonnormative ideas.^{26,27,29,30} Studies on CR instruments demonstrate that they are capable of eliciting key concepts and misconceptions and that student verbal responses in an interview closely match written responses.^{27,31} Including CR items in assessment instruments can provide instructors with an opportunity to more accurately assess student conceptual understanding.

An instructor who identifies student conceptions can focus instructional methods and practices to support students' scientific thinking.³² Teaching toward conceptual change has been proposed as a means to guide instruction, and different types of assessment aid identification

of the complex way students think about a concept.³³ As CR items can reveal mixed student thinking, there is value in adding them to an assessment strategy. However, there are significant barriers to implementation of CR items, including time to develop and use a grading rubric, cost of grading, and interpretation of responses.^{34,35} These barriers can be overcome by using computer-automated analysis methods that provide rapid reports representing student thinking in a consistent manner. Such analysis can produce computer scores that are nearly equivalent to human scores for multiple biology topics,^{35–38} and can provide instructors with a rapid, in-depth look at student understanding.

Holistic or analytic rubrics can be used to evaluate CR items. Holistic rubrics assign a single score (e.g., correct, incomplete, or incorrect) to a response and are capable of capturing a wide variety of student ideas and evaluating multiple criteria at once.^{39,40} Instructors may find using holistic rubrics faster for grading purposes, but the scores may reflect a complex mixture of ideas including combinations of normative and nonnormative ideas.^{40,41} Analytic rubrics assign one or more codes for specific components (e.g., concepts or ideas) within a response. Analytic rubrics improve the reliability of scoring, student self-assessment, provide precise diagnostic information, and help focus instruction.^{39,42–44} Additionally, categories in analytic rubrics can co-occur and thus have the potential to reveal mixed student thinking. Studies that applied analytic and holistic approaches simultaneously to rate student writing demonstrated complementary effects that improve scoring reliability and validity and provide more detailed insight into student responses than each method alone.^{40,45,46} Used in combination, analytic and holistic scoring schemes may provide instructors with complementary ways to interpret student responses.

This study aims to obtain a fine-grained picture of students' explanations of the effect of a stop codon on each of the CD processes following a base change in a gene coding region. Building upon prior work¹⁴ we investigate the following research questions:

1. Which concept(s) do students include as part of an explanation about the effect of a changed DNA nucleotide base on each of the CD processes?
2. Which ideas are associated with holistically scored correct, irrelevant, or incorrect answers?
3. How does student thinking change after instruction, as evidenced by concept(s) prone to change after instruction?

For this purpose, we developed and applied an analytic rubric to analyze student responses describing the effect of a base change on each of the CD processes. We

then compared the analytic codes to previous holistic codes.¹⁴ This approach combines the strengths of a holistic rubric and analytic rubrics.

2 | RESEARCH METHODS

2.1 | Item and data collection

This study is an analysis of previously collected student responses to a CR item (see below) derived from a MC item in the Genetics Concept Assessment.^{14,19,47} We refer to the three-question item as the “stop codon” item for convenience. The item elicits student thinking about which CD process a stop codon affects, and can be used to identify student confusion about the three processes:

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?

The data used in this project are from two previously published sets of student responses to the questions. The first data set comprises responses from students in introductory cell and molecular biology courses and was originally analyzed using lexical analysis and human holistic scoring.¹⁴ The second data set was previously analyzed with an automated holistic scoring model to assess student learning after an instructional intervention in college biology courses.¹⁹ For the current analyses, we randomly selected 350 responses from each data set. The current study was designated exempt by Michigan State University's Institutional Review Board (IRB x10-577). For information about the demographics of the student population in the previous studies, see Prevost et al.¹⁴ and Pelletreau et al.¹⁹

2.2 | Rubric development and refinement

We developed three analytic rubrics, one for each question (namely replication, transcription, and translation) using an approach of emergent coding and expert review, that our group has used successfully for similar work.³⁰ Some of the ideas targeted by our rubrics were originally identified via lexical analysis, which groups and quantifies words

and phrases in text.¹⁴ In contrast to lexical analysis, categorization with an analytic rubric uses words and phrases in context of the entire response. Each analytic rubric category captures one conceptual idea, which may be expressed in a variety of ways. Every response was coded for each analytic category dichotomously for presence (1) or absence (0), thus multiple categories can co-occur within a single response. The rubrics were refined through an iterative process of independent scoring, coder discussion and disagreement resolution, revision of rubrics, and use of statistical text analysis software.⁴⁸ The final set of analytic categories for each CD process has a parallel structure to allow us to compare student thinking about the three processes. Detailed rubrics including coding rules and example responses are found in Tables S1–S3.

2.3 | Human coding and machine learning

Paired coders coded batches of 50–100 responses for each of the three questions independently and met after each batch to discuss and resolve coding disagreements. Where agreements could not be reached between a pair of coders, responses were brought to a third coder to resolve. This process was iterated until paired coders reached sufficient overall agreement on all categories for each of the three rubrics (see below). The coders were blinded to the previous holistic codes to avoid bias. We used Cohen's kappa⁴⁹ as a measure of inter-rater reliability (IRR), with a kappa ≥ 0.6 as our threshold for satisfactory agreement.⁵⁰ All categories within the three rubrics had kappa ≥ 0.6 with one exception in the transcription rubric, *Protein or translation affected*, which had a kappa of 0.593 (Table S4). Once IRR thresholds were reached, the rubrics were finalized, and a single coder coded the remaining dataset. Additionally, we used a supervised ensemble machine learning algorithm on the coded responses for each of the three questions.^{24,51} The algorithm produces a set of predicted codes for each category as well as a probability that the predicted code is accurate. We leveraged the consistency of algorithmically predicted categories as an error check for human codes. Where an algorithmically predicted code did not match the human code and the human code was deemed to be incorrectly applied, the code was corrected to match the predicted code.

2.4 | Data and analysis of student ideas for the three CD processes

To address Research Question 1, we randomly selected 350 holistically expert-coded student responses from Prevost et al.¹⁴ and human-coded these responses using

our analytic rubric. In order to compare student ideas across the three CD processes, we named the rubric categories such that they capture similar ideas or reasoning for each process (Table 1). For example, the categories *Replication is not affected*, *Transcription is not affected*, and *Translation stops* are named *Correct effect on process* throughout the paper.

We represent co-occurrence of rubric categories as conditional frequency: a ratio of the frequency with which a given pair of categories co-occurs to the frequency of the category within the entire dataset, see (1).

$$\text{Conditional frequency} = \frac{\frac{N(\text{Category}|\text{Co-occurrent Category})}{N(\text{Co-occurrent Category})}}{\frac{N(\text{Category})}{N}}, \quad (1)$$

A conditional frequency of 1 indicates no association, or that two categories co-occur no more or less frequently than each category does individually within the entire data set. A conditional frequency of 0 indicates a negative association, or that the two categories very rarely or never co-occur. A conditional frequency greater than 1 indicates that when one of the two categories is present, the other is more likely to occur when compared to its occurrence in the entire response set.

To address Research Question 2, we analyzed the same random sample of 350 student responses from Prevost et al.¹⁴ We tallied the holistic and analytic codes in responses to determine the frequency of each analytic code within an assigned holistic code.

2.5 | Data and analysis of student responses before and after instruction

To address Research Question 3, we randomly selected 350 paired student responses from a preinstruction and postinstruction intervention condition (for a total of 700 responses) from a subsample of four institutions that participated in a previous study conducted by Pelletreau

et al.¹⁹ We coded these 700 responses using our analytic rubrics. Then, we categorized each set of paired responses the according to analytic category preintervention and postintervention in one of the following four groups: *Never used*, which represents students who did not include the idea in either their preinstruction or postinstruction responses; *Removed*, which represents students who included the idea preinstruction but did not include it postinstruction; *Maintained*, which represents students who included the idea in both their preinstruction and postinstruction responses; and *Added*, which represents students who did not include the idea preinstruction but included it in their postinstruction responses. We compared the proportion of each analytically categorized idea in student responses preinstruction and postinstruction using McNemar's test.⁵²

3 | RESULTS

3.1 | Research Question 1: Which concept(s) do students include as part of an explanation about the effect of the stop codon on each CD process?

We analytically coded 350 responses for each of the three questions, which were previously categorized using a holistic rubric.¹⁴ Each response can include zero, one, or more analytic categories. Most students' responses included the ideas *Correct effect on process* and/or *Incorrect effect on process* (a total of 69% of all responses in replication, 66% transcription, 66% translation; Figure 1). However, about half of the students included more than one idea in their response (mean number of categories per response = 1.6 for replication, 1.4 for transcription, and 1.7 for translation). In responses to the three questions, the most common product category is the *Identify product* category, identified by students in approximately equal proportions (36% replication, 32% transcription, 43% translation). The *Product is changed* category occurs

TABLE 1 Comparison of rubrics for the three Central Dogma questions

	Correct effect on process	Incorrect effect on process	Identify product	Product is changed	There is mutation	Affects another process
Replication	Replication is not affected	Replication stops or DNA is shorter	Product is DNA	The DNA is different	There is mutation	Protein or translation affected
Transcription	Transcription is not affected	Transcription stops or RNA is shorter	Product is RNA	The RNA is different	There is mutation	Protein or translation affected
Translation	Translation stops or protein is shorter	Translation is not affected	Product is protein	Protein structure and/or function is changed	There is mutation	Replication or transcription stopped

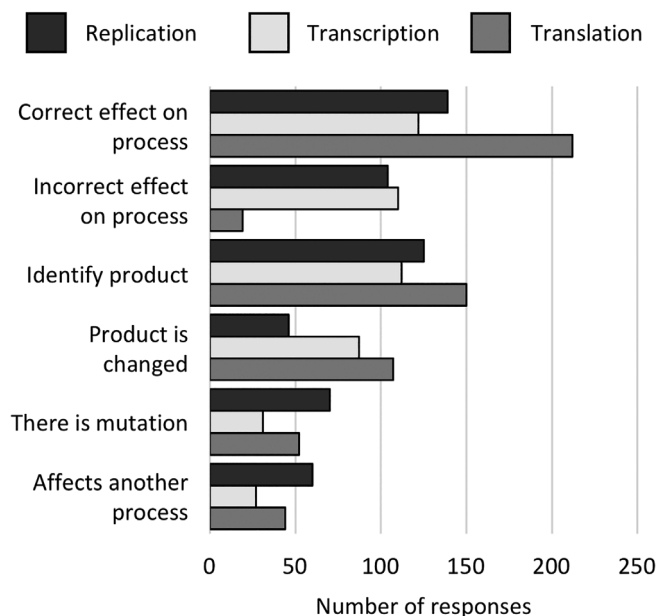


FIGURE 1 Student thinking about each of the Central Dogma (CD) processes. The most frequent analytic categories in student responses refer to the effect on the process (correct and incorrect). Number of responses = 350 for each process. Note that the total number of ideas may be greater than 350, as ideas can co-occur in a response

more frequently in responses to the translation question than to the other two questions.

To explore the combination of ideas included in student responses, we looked at co-occurrence of analytic categories for each question using conditional frequency (Figure 2). Because the most common categories in student responses are those that describe an effect on the process, we first looked at ideas that co-occur with process descriptions. Not surprisingly, students rarely discuss the *Incorrect effect on process* together with the *Correct effect on process* for any of the three questions (conditional frequency = 0.02 replication, 0.03 transcription, 0.26 translation). In the rare case that these ideas co-occur, students' responses contain statements that may be perceived as contradictory, such as this response to the translation question "The gene will be able to translate all of the strand however it still remains that the stop codon causes the remainder of the steps to be stopped" (underlining added by authors to indicate coded words). The *Product is changed* category also tends to have a negative association with the *Incorrect effect on process* idea for all three questions, indicating that students rarely discuss an effect other than the product being shorter (conditional frequency = 0.15 replication, 0.00 transcription, 0.34 translation).

Next, we looked for frequently co-occurring ideas in student responses. We found a positive association

between the ideas *Identify product* and *Product is changed* for all three questions (conditional frequency = 2.07 replication, 2.20 transcription, 1.61 translation). Responses to the replication and transcription questions with this pair of ideas tend to describe a changed base on the newly synthesized DNA or RNA strand. For example, a student response to the replication question states, "It will have a T on the complementary strand instead of a C". Responses to the translation question with this pair of ideas occasionally describe a changed amino acid but more often describe a change in protein function. For example, a student wrote, "This will create a non-functional protein. Stopping translation early removes the necessary amino acids to make the protein function." In responses to the replication and transcription questions, the ideas *There is mutation* and *Affects another process* also frequently co-occur. (conditional frequency = 1.67 replication, 1.71 transcription). While these ideas co-occur in responses to the translation question, they are nearer to a neutral association (conditional frequency = 1.22). Finally, we note that one idea used in replication responses (*Product is changed*) rarely co-occurs with either *Affects another process* (conditional frequency = 0.13) or with *There is mutation* (conditional frequency = 0.11). That is, students who describe the expected change to the replication product do not include effects on translation or a mutation in their responses.

3.2 | Research Question 2. Which ideas are associated with holistically scored *Correct*, *Irrelevant*, or *Incorrect* answers?

We associated the analytic categories for each response applied in this work with the original holistic codes, where responses were categorized as *Correct*, *Incomplete/Irrelevant* (*Irrelevant* hereafter), or *Incorrect*, see Reference 14. *Correct* responses state or imply that the process was not affected (replication and transcription) or that the process stopped (translation) and *Incorrect* state or imply the inverse effect. *Incomplete/Irrelevant* responses contain some, but not all, correct information, or do not refer to the process in question. For brevity in this paper, we will refer to this holistic code as *Irrelevant*. A surprising number of ideas were represented in each holistic code for each process. On average, 5.8 distinct analytically categorized ideas were found in each holistic code (range 5–6, see Figures 3, S1, and S2, Table 2).

The most common analytic category within the holistic *Correct* code was *Correct effect on process* (76% replication, 71% transcription, and 94% translation). These percentages indicate that the presence of one analytic category does not fully explain the holistic code, so we looked at the number

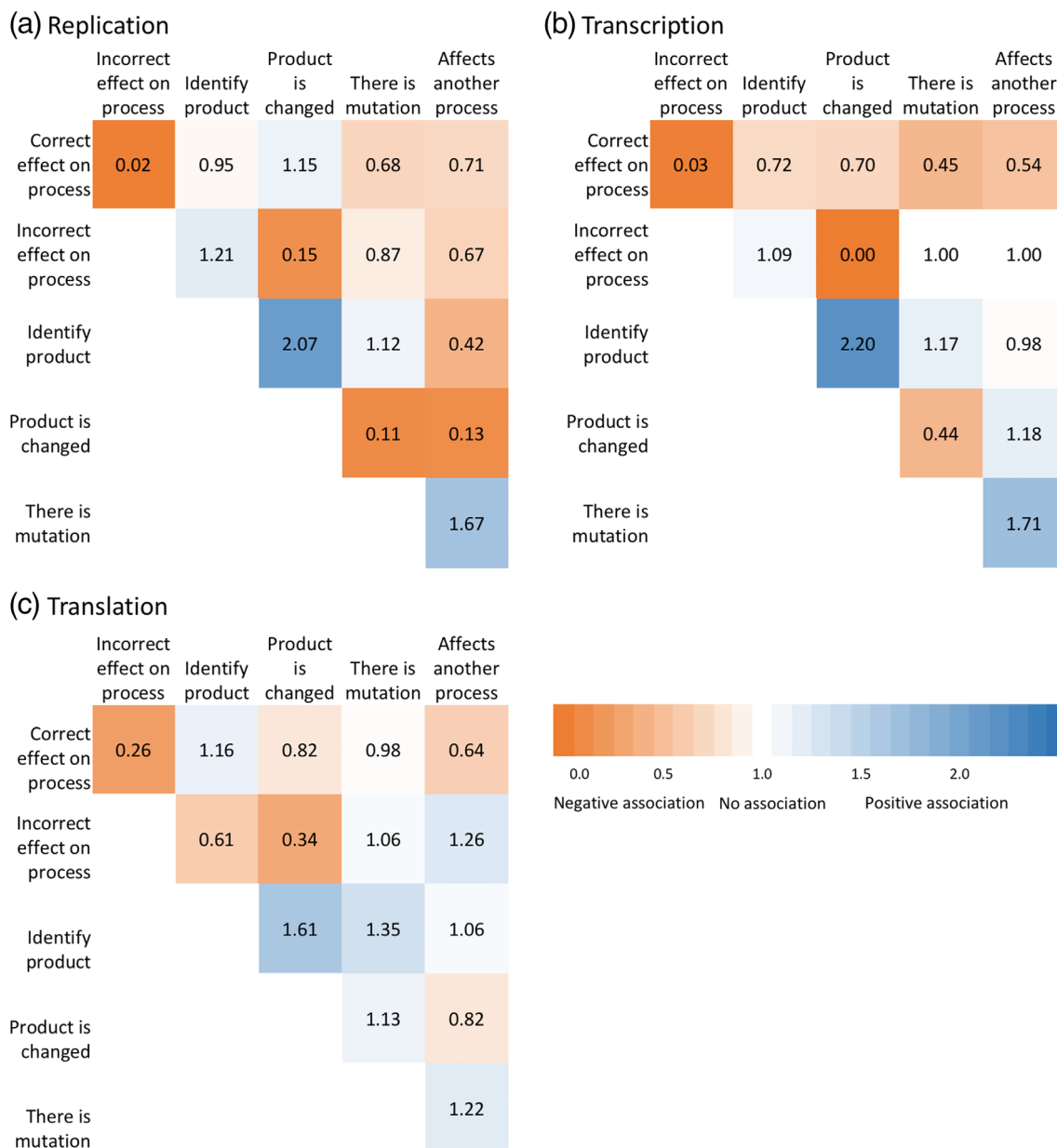


FIGURE 2 Pairwise conditional frequencies among rubric categories identified in student responses across the three Central Dogma (CD) processes. For each of the three CD processes, students tend to pair the ideas *There is mutation* and *Affects another process*. Any pair of analytic codes can co-occur for any of the three CD processes

of analytic categories in each holistic code. The number of categories in each holistic *Correct* response ranges between zero and five with an average of 1.7 replication, 1.6 transcription, and 1.9 translation (Table S5). The next most frequent analytic categories in holistically *Correct* responses are *Identify product* (41% replication, 37% transcription, and 50% translation) and *Product is changed* (24% replication, 35% transcription, and 25% translation).

In replication and transcription responses, the most common analytic category within the holistic *Incorrect* code was *Incorrect effect on process* (92% replication & 81% transcription). However, no single analytic category

occurs in a majority of holistically *Incorrect* translation responses, including *Incorrect effect on process* (16%). The number of categories in a holistic *Incorrect* response range from 0 to 5 with averages of 1.6 replication, 1.4 transcription, and 1.5 translation (Table S5). In replication and transcription responses holistically coded *Incorrect*, additional frequent analytic categories include *Identify product* (42% replication and 32% transcription) and *There is mutation* (18% replication and 10% transcription). In translation responses holistically coded *Incorrect*, the most frequent analytic category included is *Affects another process* (42% of *Incorrect* responses).

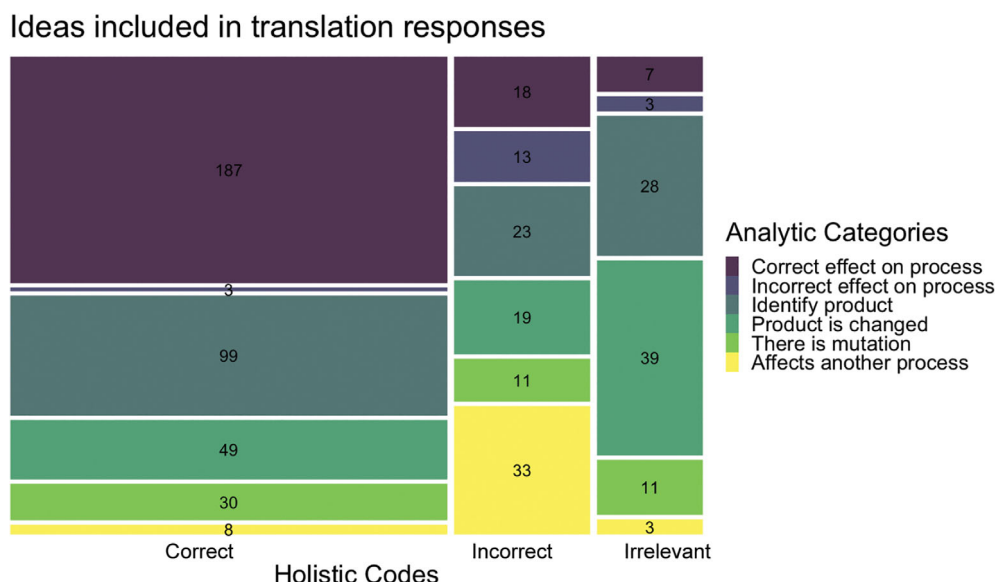


FIGURE 3 A mosaic plot showing the categorization of student responses as holistically and analytically coded for translation. Holistic codes include ideas from multiple analytic categories. Column widths are proportional to the number of student responses categorized as *Correct* (199 responses), *Incorrect* (79 responses), or *Irrelevant* (72 responses). Number of student responses categorized in a given holistic and analytic category are printed on graph. Note that analytic codes can occur in zero or more categories for each response so the total number of codes within a holistic category is the sum of the analytic category codes and may be greater than 350. Therefore, the number of analytic category codes does not equal to the number of holistically coded responses

TABLE 2 Percent of responses for each question and holistic code that are categorized as containing each of six analytic ideas

Category	Replication			Transcription			Translation		
	Correct (%)	Incorrect (%)	Irrelevant (%)	Correct (%)	Incorrect (%)	Irrelevant (%)	Correct (%)	Incorrect (%)	Irrelevant (%)
Correct effect on process	76	1	2	71	2	12	94	23	10
Incorrect effect on process	0	92	5	0	81	2	2	16	4
Identify product	41	42	8	37	32	19	50	29	39
Product is changed	24	1	2	35	2	26	25	24	54
There is mutation	16	18	35	7	10	14	15	14	15
Affects another process	11	11	47	6	10	16	4	42	4

Note: Note that analytic ideas can co-occur and thus may not equal 100%.

The *Irrelevant* holistic codes for all three questions were not associated with a particular analytic category and were more likely to include a single category than either of the other holistic codes (range 0–3; averages of 1.0 replication, 0.9 transcription, 1.3 translation; Table S5). The analytic categories *There is mutation* (35%) and *Affects another process* (47%) are most

common in replication responses holistically coded *Irrelevant*. In transcription responses, the most common analytic category (*Product is changed*) only occurs in 26% of holistic *Irrelevant* responses. The analytic categories *Product is changed* (54%) and *Identify product* (39%) are most common in holistic *Irrelevant* translation responses.

3.3 | Research Question 3: How does student thinking change after instruction?

To answer our third research question, we analytically coded 350 pre instruction responses and 350 paired post instruction responses from a dataset previously collected to assess the effects of an instructional intervention. The instructional intervention included a clicker case study designed to address common student misconceptions about the CD.¹⁹ Student responses were categorized to reflect whether the student used an idea, as captured by the analytic rubrics, before and/or after the intervention as (1) *Never used*, (2) *Removed*, (3) *Maintained*, (4) *Added* (Figure 4).

Student responses to the replication and transcription questions have similar patterns of change after instruction, where students *Add* the idea *Correct effect on process* (152 replication and 169 transcription) and *Remove* the idea *Incorrect effect on process* (107 replication and 100 transcription). Many of the students who *Removed* the *Incorrect effect on the process* replaced it by *Adding* the *Correct effect on process* (94 replication and 86 transcription). Students removed two other ideas from their responses to the replication and transcription question: *Identify product* (109 replication and 83 transcription) and *Product is changed* (87 replication and 80 transcription). The ideas *There is mutation* and *Other process affected* did not substantially change in either the replication or transcription responses after instruction.

The patterns of change were slightly different for responses to the translation question. Inclusion of the *Incorrect effect on process* idea did not change while 125 students *Added* the *Correct effect on process* idea to their responses. Two other ideas were added after instruction:

Identify product (90 students) and *There is mutation* (67 students). The ideas captured by the categories *Product is changed* and *Affects another process* did not substantially change after instruction. For co-occurrence of analytic categories pre and post instruction, see Figures S3–S5.

4 | DISCUSSION

The previously applied holistic rubric¹⁴ produced a rapid tally of the number of students who could differentiate between the three processes by correctly describing that the stop codon only affects translation. Because the holistic rubric places student responses into one of three mutually exclusive scores, *Correct*, *Incorrect*, or *Irrelevant*, additional ideas included in student responses may be masked from instructors. We developed and applied an analytic rubric that reveals a nuanced picture of student understanding that can help reveal the range and mix of ideas in student written explanations. We found that students mix ideas in their written responses pre and post instruction, across all three CD processes, and regardless of whether their answers were *Correct*, *Incorrect*, or *Irrelevant*. This suggests that no individual analytic category can fully explain a holistic code and underscores the importance of analytic categorization of concepts in student writing.

Analytic rubrics allow us to identify ideas and capture the context in which students use them in responses. For example, the word DNA in reference to replication could refer to the template, the new daughter strand, a change in sequence, or the DNA sequence provided in the item. Our coding scheme captures the ideas *Product is DNA* or *The DNA is different* as a way to differentiate two of these potential uses. This approach helps us better identify

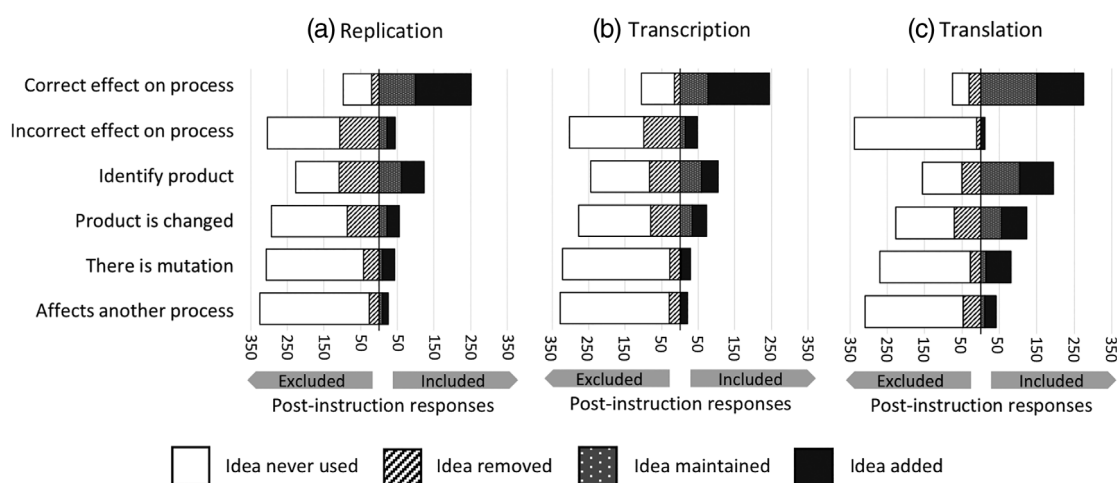


FIGURE 4 Butterfly plot of analytic categories in paired student responses preintervention and postintervention; 350 preinstruction responses and 350 postinstructional intervention. Students often add correct ideas about the three Central Dogma (CD) processes. *Never used*, absent pre and post. *Removed*, present pre and absent post, *Maintained*, present pre and post. *Added*, absent pre and present post

potential issues in students' explanations and understanding.

The analytic categorization helps uncover ideas students include in holistically *Irrelevant* responses.¹⁴ Identifying the ideas students include in an *Irrelevant* response is key for targeting instruction. *Irrelevant* responses included ideas in all analytic categories and a significant portion of student responses included relevant ideas about the CD processes. *Irrelevant* responses to the transcription and translation questions often include the ideas *Identify product* or *Product is changed*, and responses to the replication question often include the *Mutation* idea. These ideas alone do not necessarily represent misconceptions, though without connecting to other ideas, students cannot exhibit complete understanding of the effects on each process. Two interventions specifically address effects of mutations in the CD.^{19,24} To address issues with CD products, an instructor might consider in-class discussions to help students focus on relating the CD processes to the production of the relevant macromolecules.

We found that the holistic *Incorrect* code for the translation question was not strongly associated with the analytic *Incorrect effect on process* category. *Incorrect* responses include a wide variety of analytic categories; representing the variety of ways students think about and describe translation. Analytic categorization can help the instructor to pinpoint and address misconceptions. A misconception that replication is stopped requires additional work on the (lack of) effect of a stop codon on replication, while a misconception that translation continues requires additional work on the process of translation. The ideas *Affects another process* and *Product is changed* often co-occur in translation responses, which indicates how the student was incorrect. Thus, by employing a combination of scores that provide a rapid overview (holistic) of student responses as well as a more detailed (analytic) report of student thinking, it becomes possible to more clearly interpret some holistic codes, including when students mix ideas.

We found that students add the *Correct effect on process* to responses about all three processes after instruction. This is consistent with the instructional intervention targeted at dispelling the misconceptions that a stop codon would affect replication or transcription.¹⁹ Students also added to their translation responses the ideas *Product is changed* and *There is mutation*. Our analysis on ideas prone to change after instruction could be extended to the assessment of other targeted instructional interventions, including the *Product is changed* idea as we found that this idea was associated with other normative ideas such as *Correct effect on product* and with holistic *Correct* codes.

The combination of holistic and analytic rubrics provides a useful formative assessment tool for instructors

to monitor student ideas and potentially tailor instruction. In the goal of improving student learning, formative assessment involves identifying where the learner is and how to get the learner to the desired learning goal.^{25,53} The instructor's role is to elicit student understanding to make informed instructional decisions.⁵³ When students use ideas in a nonexpert-like manner, for example, using an idea about mutation without referencing the effect on the relevant CD process, instructors can help students understand the connections between a mutation and the processes of gene expression.^{15,19,24} Furthermore, when students mix ideas³² for example the protein will be shorter because replication stops, instructors can help students examine the validity of each idea. Knowing what students are thinking at a conceptual level can help instructors make informed decisions about instruction, a critical piece of active-learning classrooms.⁵⁴


Leveraging the work from this project, we developed automated scoring models for instructor use that allows new student responses to be coded at the analytic level (this work) or the holistic level.¹⁴ The "stop codon" item and computer automated categorization models, including a variety of ways for instructors to analyze student responses are available for instructor use at <https://beyondmultiplechoice.org>.

ACKNOWLEDGMENTS

The authors thank the Automated Analysis of Constructed Response (AACR) collaboration for helpful conversations while preparing this manuscript, especially Drs. Jennifer Kaplan, Alex Lyford, Lauren Jescovitch, and Michael Fleming. Andrea Bierama, Anne-Marie Hoskinson, and Alexandria Mazur participated in creation of early versions of the rubric. Additionally, we would like to thank collaborating instructors for feedback that informed our rubric revisions. This material is based upon work supported by the National Science Foundation (DUE grant 1323162).

ORCID

Juli D. Uhl  <https://orcid.org/0000-0003-1569-7122>

Kamali N. Sripathi  <https://orcid.org/0000-0002-2002-8540>

Rosa A. Moscarella  <https://orcid.org/0000-0002-6263-7555>

Mark Urban-Lurain  <https://orcid.org/0000-0002-2243-8252>

Kevin C. Haudek  <https://orcid.org/0000-0003-1422-6038>

REFERENCES

1. NGSS. Lead States Next Generation Science Standards; For States, By States; 2013. <https://www.nextgenscience.org/>.

2. National Research Council. A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, DC: The National Academies Press; 2012.
3. American Association for the Advancement of Science (AAAS). Vision and change in undergraduate biology education: a call to action. Final report. AAAS: Washington, DC; 2011.
4. Brownell SE, Freeman S, Wenderoth MP, Crowe AJ. BioCore Guide: a tool for interpreting the core concepts of vision and change for biology majors. *CBE Life Sci Educ*. 2014;13:200–11.
5. Cary T, Branchaw J, Shuster M. Conceptual elements: a detailed framework to support and assess student learning of biology Core concepts. *CBE Life Sci Educ*. 2017;16:ar24.
6. Kalinowski ST, Leonard MJ, Andrews TM. Nothing in evolution makes sense except in the light of DNA. *CBE Life Sci Educ*. 2010;9:87–97.
7. Mead R, Hejmadi M, Hurst LD. Teaching genetics prior to teaching evolution improves evolution understanding but not acceptance. *PLoS Biol*. 2017;15:e2002255.
8. Dougherty MJ, Lontok KS, Donigan K, McInerney JD. The critical challenge of educating the public about genetics. *Curr Genet Med Rep*. 2014;2:48–55.
9. Guttmacher AE, Porteous ME, McInerney JD. Educating health-care professionals about genetics and genomics. *Nat Rev Genet*. 2007;8:151–7.
10. Redfield RJ. “Why do we have to learn this stuff?”—a new genetics for 21st century students. *PLoS Biol*. 2012;10:e1001356.
11. Briggs AG, Morgan SK, Sanderson SK, Schulting MC, Wieseman LJ. Tracking the resolution of student misconceptions about the central dogma of molecular biology. *J Microbiol Biol Educ*. 2016;17:339–50.
12. Fisher KM. A misconception in biology: amino acids and translation. *J Res Sci Teach*. 1985;22:53–62.
13. Marbach-Ad G, Stavy R. Students' cellular and molecular explanations of genetic phenomena. *J Biol Educ*. 2000;34:200–5.
14. Prevost LB, Smith MK, Knight JK. Using student writing and lexical analysis to reveal student thinking about the role of stop codons in the Central Dogma. *CBE Life Sci Educ*. 2016;15:ar65.
15. Southard K, Wince T, Meddleton S, Bolger MS. Features of knowledge building in biology: understanding undergraduate students' ideas about molecular mechanisms. *CBE Life Sci Educ*. 2016;15:ar7.
16. Wood-Robinson C, Lewis J, Leach J. Young people's understanding of the nature of genetic information in the cells of an organism. *J Biol Educ*. 2000;35:29–36.
17. Wright LK, Fisk JN, Newman DL. DNA → RNA: what do students think the arrow means? *CBE Life Sci Educ*. 2014;13:338–48.
18. Smith MK, Knight JK. Using the genetics concept assessment to document persistent Conceptual difficulties in undergraduate genetics courses. *Genetics*. 2012;191:21–32.
19. Pelletreau KN, Andrews T, Armstrong N, Bedell MA, Dastoor F, Dean N, et al. A clicker-based case study that untangles student thinking about the processes in the central dogma. *CourseSource*. 2016;3.
20. Zhao F, Schuchardt A. Exploring students' descriptions of mutation from a cognitive perspective suggests how to modify instructional approaches. *CBE Life Sci Educ*. 2019;18:ar45.
21. Zukswert JM, Barker MK, McDonnell L. Identifying troublesome jargon in biology: discrepancies between student performance and perceived understanding. *CBE Life Sci Educ*. 2019;18:ar6.
22. Jensen JL, Kummer TA, Banjoko A. Assessing the effects of prior conceptions on learning gene expression. *J Coll Sci Teach*. 2013;42:82–91.
23. Moscarella R, Haudek K, Knight J, Mazur A, Pelletreau K, Prevost L, et al. Automated analysis provides insights into students' challenges understanding the processes underlying the flow of genetic information; 2016.
24. Sieke SA, McIntosh BB, Steele MM, Knight JK. Characterizing students' ideas about the effects of a mutation in a noncoding region of DNA. *CBE Life Sci Educ*. 2019;18:ar18.
25. Black P, Wiliam D. Developing the theory of formative assessment. *Educ Assess Eval Acc*. 2009;21:5–31.
26. Birenbaum M, Tatsuoaka KK. Open-ended versus multiple-choice response formats—it does make a difference for diagnostic purposes. *Appl Psychol Measur*. 1987;11:385–95.
27. Nehm RH, Schonfeld IS. Measuring knowledge of natural selection: a comparison of the CINS, an open-response instrument, and an oral interview. *J Res Sci Teach*. 2008;45:1131–60.
28. Stanger-Hall KF. Multiple-choice exams: an obstacle for higher-level thinking in introductory science classes. *CBE Life Sci Educ*. 2012;11:294–306.
29. Hubbard JK, Potts MA, Couch BA. How question types reveal student thinking: an experimental comparison of multiple-true-false and free-response formats. *CBE Life Sci Educ*. 2017;16:ar26.
30. Sripathi KN, Moscarella RA, Yoho R, You HS, Urban-Lurain M, Merrill J, et al. Mixed student ideas about mechanisms of human weight loss. *CBE Life Sci Educ*. 2019;18:ar37.
31. Weston M, Haudek KC, Prevost L, Urban-Lurain M, Merrill J. Examining the impact of question surface features on Students' answers to constructed-response questions on photosynthesis. *CBE Life Sci Educ*. 2015;14:ar19.
32. Vosniadou S. In: Fraser BJ, Tobin K, McRobbie CJ, editors. Second international handbook of science education. Dordrecht: Springer Netherlands; 2012.
33. Tanner K, Allen D. Approaches to biology teaching and learning: understanding the wrong answers—teaching toward Conceptual change. *Cell Biol Educ*. 2005;4:112–7.
34. Nehm RH, Ha M. Item feature effects in evolution assessment. *J Res Sci Teach*. 2011;48:237–56.
35. Nehm RH, Beggrow EP, Opfer JE, Ha M. Reasoning about natural selection: diagnosing contextual competency using the ACORNS instrument. *Am Biol Teach*. 2012;74:92–8.
36. Ha M, Nehm RH, Urban-Lurain M, Merrill JE. Applying computerized-scoring models of written biological explanations across courses and colleges: prospects and limitations. *CBE Life Sci Educ*. 2011;10:379–93.
37. Haudek KC, Prevost LB, Moscarella RA, Merrill J, Urban-Lurain M. What are they thinking? Automated analysis of student writing about acid–base chemistry in introductory biology. *CBE Life Sci Educ*. 2012;11:283–93.
38. M. Urban-Lurain, R. A. Moscarella, K. C. Haudek, E. Giese, D. F. Sibley, J. E. Merrill 2009 39th IEEE Frontiers in Education Conference: FIE 2009, IEEE, San Antonio, TX, USA; 2009, pp. 1–6.

39. Brookhart SM. Appropriate criteria: key to effective rubrics. *Front Educ*. 2018;3:22.
40. Tomas C, Whitt E, Lavelle-Hill R, Severn K. Modeling holistic marks with analytic rubrics. *Front Educ*. 2019;4.
41. Liu OL, Brew C, Blackmore J, Gerard L, Madhok J, Linn MC. Automated scoring of constructed-response science items: prospects and obstacles. *Educ Meas Issues Pract*. 2014;33:19–28.
42. Panadero E, Jonsson A. The use of scoring rubrics for formative assessment purposes revisited: a review. *Educ Res Rev*. 2013;9: 129–44.
43. Jonsson A, Svingby G. The use of scoring rubrics: reliability, validity and educational consequences. *Educ Res Rev*. 2007;2: 130–44.
44. Trevisan MS, Davis DC, Calkins DE, Gentili KL. Designing sound scoring criteria for assessing student performance. *J Eng Educ*. 1999;88:79–84.
45. Harsch C, Martin G. Comparing holistic and analytic scoring methods: issues of validity and reliability. *Assess Educ Princ Policy Pract*. 2013;20:281–307.
46. Hunter DM, Jones R, Randhawa B. The Use of Holistic versus Analytic Scoring for Large-Scale Assessment of Writing. *Can J Program Eval*. 1996;11:61–85.
47. Smith MK, Wood WB, Knight JK. The genetics concept assessment: a new concept inventory for gauging student understanding of genetics. *CBE Life Sci Educ*. 2008;7:422–30.
48. Peladeau N. Wordstat: content analysis and text mining software. Montreal, Quebec, Canada: Provalis; 2018.
49. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37–46.
50. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–74.
51. Jurka TP, Collingwood L, Boydstun A, Grossman E, van Atteveldt W. RTextTools: a supervised learning package for text classification. *R J*. 2013;5:6.
52. Adedokun OA, Burgess WD. Analysis of paired dichotomous data: a gentle introduction to the McNemar test in SPSS. *JMDE*. 2012;8:7.
53. Black P, Wiliam D. Classroom assessment and pedagogy. *Assess Educ Princ Policy Pract*. 2018;25:551–75.
54. Auerbach AJJ, Andrews TC. Pedagogical knowledge for active-learning instruction in large undergraduate biology courses: a large-scale qualitative investigation of instructor thinking. *Int J STEM Educ*. 2018;5:19.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Uhl JD, Sripathi KN, Saldanha JN, et al. Introductory biology undergraduate students' mixed ideas about genetic information flow. *Biochem Mol Biol Educ*. 2021;49: 372–382. <https://doi.org/10.1002/bmb.21483>