

**Developing assessments to elicit and characterize undergraduate
mechanistic explanations about information flow in biology**

Juli D. Uhl^{†*}, Megan Shiroda[†], Kevin C. Haudek^{†‡}

[†]*CREATE for STEM Institute, Michigan State University, East Lansing, Michigan, USA*

[‡]*Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan, USA*

*Corresponding Author: Juli D. Uhl (uhljuli@msu.edu)

ORCID

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

Megan Shiroda <https://orcid.org/0000-0002-2504-7564>

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

Developing assessments to elicit and characterize undergraduate mechanistic explanations about information flow in biology

Abstract: National calls to improve science education include focusing on scientific practices coupled with learning disciplinary core ideas. Among the practices is constructing explanations. In the field of cellular and molecular biology explanations typically include a mechanism and can be used to make predictions about phenomena. In this work, we developed an assessment item about transcription, a key process in the biology core concept of genetic information flow. We used a mechanistic framework to develop a rubric that identifies undergraduate explanations that leverage molecular or sub-molecular mechanisms, descriptions, or use unlinked ideas. We applied this rubric to categorize 346 undergraduate written explanations and compare five item versions. We found that one version elicited sub-molecular mechanistic explanations from 20% of students, compared to between 2 and 13% from other versions. This version included the element of time, by indicating that a new RNA was formed as part of transcription. We also developed and applied a conceptual rubric to capture the context students used in their explanations and found a median of two context ideas in student explanations of transcription. Our work demonstrates that with careful item wording, undergraduates can explain molecular processes like transcription by leveraging sub-molecular mechanisms.

Keywords: Mechanistic explanation; undergraduate; genetic information flow; prompt development; constructed response

Introduction

National calls to improve science education include a focus on scientific practices linked to disciplinary core ideas (e.g., American Association for the Advancement of Science, 2011; National Research Council, 2012). In the disciplines of molecular and cellular biology, the scientific practice of explaining core ideas of genetic information flow should include a mechanism using entities at lower scalar levels, for example from molecular to sub-molecular, to describe the phenomenon (van Mil et al., 2013). Assessing the complex ideas and connections in mechanistic explanations requires prompts that allow students to respond in their own words, such as constructed response (CR) items. In this paper, we describe the development of a CR item with associated rubric capable of eliciting and capturing undergraduate mechanistic explanations of the process of transcription, a process within the core concept of genetic information flow.

Background

For undergraduate biology the American Association for the Advancement of Science document, *Vision and Change*, sets a goal for students to be capable of interpreting and communicating about science within the disciplinary practices, as well as the need for students to learn the core biological concepts (American Association for the Advancement of Science, 2011). The three-dimensional learning framework (National Research Council (NRC), 2012) was developed to guide science education at the K-12 level and is applicable to college-level introductory course and assessment design (e.g., Roche Allred et al., 2021; Cooper et al., 2015; Matz et al., 2018). The NRC framework includes three linked learning dimensions: scientific and engineering practices (SEPs), disciplinary core ideas (DCIs), and cross-cutting concepts (CCCs). The NRC framework includes among the SEPs,

“Constructing Explanations and Designing Solutions,” and describes scientific explanations as supporting cause and effect relationships (NGSS Lead States, 2013). Vision and Change implies inclusion of explanations within the practice of applying the scientific process (Clemmons et al., 2020). Based on these frameworks, students should learn to produce explanations about phenomena linked to core concepts. Further, explanations which draw upon scientific concepts provide students a way to apply their conceptual knowledge and thus represent a potential method for conceptual change (NRC, 2012).

Scientific explanations can draw from multiple philosophies (Braaten & Windschitl, 2011), and experts often incorporate analogies, methods, and context as part of their explanation of a molecular phenomenon (Trujillo et al., 2015). However, it is commonly accepted that an explanation of a scientific phenomenon should be a causal account of how and why the phenomenon occurs (NRC, 2012; Sandoval & Millwood, 2005). Explanations should be more than a definition or description and should include a causal chain linking reasoning that fits the available evidence to a claim (Reiser et al., 2012). Furthermore, scientific explanations should apply scientific models or theory to explain the cause of the phenomenon and be consistent with available evidence and student understanding of science (NRC, 2012, Chapter 3). When experts make explanations, they often use them for the purpose of making predictions and generating new hypotheses, but for students an educational goal might be to make an explanation that accounts for the cause(s) of the phenomenon (NRC, 2012, Chapter 3).

In molecular and cellular biology, scientific explanations include mechanisms as a central feature (Machamer et al., 2000; van Mil et al., 2013). In a mechanistic explanation, one pays attention to entities and their spatial and temporal organization as well as multiple scalar levels of organization. For example, entities interact by binding, which leads to a

change in state and a change in activity (van Mil et al., 2013). Furthermore, biological explanations often require reasoning across scalar levels, from the molecular to organismal level for example (Parker et al., 2012). In particular, these types of explanations need to identify the scalar level of the phenomenon, reason at one or more scalar levels below the phenomenon, or to an abstract level (such as energy) and leverage characteristics and activities of entities to describe the phenomenon (Krist et al., 2018).

Work on molecular genetic mechanisms indicates that in addition to physical levels of entities, the level of information must be considered. For example, a DNA sequence has an information level in that it contains genetic information, which must be leveraged alongside a physical level such as the protein for which the DNA codes (Duncan & Reiser, 2007). Thus, in addition to considering a lower scalar level (e.g., to the molecular or sub-molecular level) in their explanations of genetic information flow, students also need to link the information level to the ultimate output of the transcriptional process, which is commonly linked to the translation of protein. Previous work suggests that students can identify molecular mechanisms involved in biological functions but may have challenges linking across scalar levels (e.g., from molecular to cellular; see Marbach-Ad & Stavy, 2000; Southard et al., 2016; Southard et al., 2017).

Genetic mechanisms and concepts - when taught together - have the potential to improve learning. Learning about genetic information flow is challenging for both secondary and undergraduate students (see for example, Briggs et al., 2017; Gericke & Wahlberg, 2013; Marbach-Ad & Stavy, 2000), and some misconceptions persist even after instruction (Smith & Knight, 2012). At the college level, introductory biology students have misconceptions about the processes of replication, transcription, and translation (Prevost et al., 2016; Smith et al., 2008). Furthermore, undergraduates often include

additional, sometimes irrelevant concepts in their descriptions of each process, and transcription can be especially difficult for students to learn and may be considered a threshold concept (Moscarella et al., 2016).

Students can have several misconceptions about transcription, including that the DNA undergoes a chemical conversion to become RNA or that the RNA exists before transcription (Wright et al., 2014). Even when students attain correct understandings of genetics, retaining that learning can be limited, which poses challenges for students when they need to apply introductory genetics concepts during upper-level courses. After completion of introductory biology, students retain information about the relationships of subunits to DNA and protein and less so on the relationships between the DNA and protein and enzymes with the processes of gene expression (Briggs et al., 2016). Memorizing patterns instead of learning mechanisms can hinder learning (Castro-Faix et al., 2021). It has been proposed that learning genetic mechanisms leads to longer retention of knowledge than rote memorization (Todd & Romine, 2018) and that non-mechanistic conceptions of phenomena can be a barrier to learning (Haskel-Ittah & Yarden, 2018). Thus, some learning challenges may be mitigated by teaching students the skills necessary to mechanistically reason about genetics.

Study Context

We wanted to know if students in introductory biology classes were connecting the process of transcription to the core concept of information flow from Vision and Change (American Association for the Advancement of Science; 2011). To analyze student understanding of transcription, we asked students to write explanations to the item, “When an organism needs a particular gene product (protein), gene expression is activated, resulting in the

production of RNA. How is the information contained in a gene transferred from DNA to RNA during transcription?” We found that many explanations described the process of RNA polymerase synthesizing a new RNA, and/or described that complementary base-pairing occurred between the RNA and DNA molecule. However, few students linked the idea of complementary base-pairing with the idea that the information transferred is in the form of a sequence transferred from the DNA to RNA during transcription (unpublished results).

In focus group interviews related to this item we found that when pressed, students were able to state that the nature of the genetic information is the sequence of DNA and RNA, and that the sequence is transferred during transcription (unpublished results). We concluded that written responses to the original item to be potentially unaligned with student ability to explain and link ideas about transcription. Based on this, we sought to develop a new item to elicit student mechanistic understanding of transcription and an aligned coding rubric that could identify mechanistic explanations.

Research Questions

We reviewed the item described above and noted that there was nothing that indicated to students that they need to describe how the information is transferred, or to make any links between the information transferred and the entities, characteristics, and activities involved in transcription. Thus, we sought to develop a new assessment item and rubric that could elicit and evaluate student explanations about transcription. We then applied this rubric and asked the following research questions.

- (1) What types of explanations do undergraduates write about transcription?

- (2) What item characteristics are more likely to elicit a mechanistic explanation about transcription?

Methods

Item Development

We based our assessment design on the NRC assessment triangle (NRC, 2001, 44-51), which includes three vertices of Cognition, Observation, and Interpretation. In our assessment, Cognition is represented by mechanistic links across scalar levels to describe how and why a phenomenon occurs (Machamer et al., 2000). We used student CRs as our Observations, which have been shown to closely resemble student verbal responses during interviews (Nehm & Schonfeld, 2008; Weston et al., 2015). Student responses were Interpreted with a mechanistic coding rubric based on mechanistic frameworks capturing the target phenomenon, identifying activities of entities at scalar level below the phenomenon, and connecting to target phenomenon in the context of genetic information flow during transcription (Krist et al., 2018; Russ et al., 2008).

All items developed for this study included the context of information as transferred from DNA to RNA during transcription, plus one or more prompts for students to describe the process used in information transfer. Items were developed with input from a group of biochemistry, molecular biology, and microbiology experts with the intent that they be useful as formative assessment for college biology instructors who wish to understand and respond to students' thinking (Pellegrino et al., 2016). We used written scaffolding in our items (McNeill et al., 2006), which involved using keywords in prompts such as *describe*, *explain*, and *why* to aid students in writing explanations. We also scaffolded content to encourage respondents to think about the levels below the phenomena (molecular and sub-

molecular), and to include the element of time as a comparison point (as recommended by Krist et al., 2018). Item versions 1, 2, and 3 ask students to use the characteristics of DNA and RNA in their explanations, and item versions 4 and 5 include three prompts to guide students to think about the molecular and sub-molecular level of the process (Table 1). The two multi-part prompts were intended to help students separately focus on characteristics of entities at multiple scalar levels.

Data Collection

We collected 346 undergraduate responses from introductory biology and upper division biochemistry courses at a large research-intensive university in Fall 2019 and Spring 2020 (Table 2). Students randomly received an item version, and responses were collected electronically through a course learning management system as post-instructional homework. Spring 2020 responses were collected after Covid-19 forced a shift to online learning. Responses were de-identified and those with multiple text box entries were concatenated into a single block of text for coding. Typical student responses were a paragraph in length. Coders were unaware of the item version and course level to avoid bias. Responses presented as examples in this paper were spell-checked for clarity.

Theoretical Framework

We drew upon the Essential Epistemic Heuristics for Mechanistic Reasoning (EEHMR) framework proposed by Krist et al. (2018), to develop a rubric to capture student explanations about genetic information flow during transcription (Krist et al., 2018). The EEHMR framework consists of four codes (italicized) for mechanistic reasoning: *Multiple Levels* which refers to considering levels above or below the phenomenon, *Identifying* and

Unpacking Factors at a lower scalar level, and *Linking* lower-level interactions and behaviors to the target phenomenon. In *Multiple Levels*, the EEHMR framework indicates that explanations must include a cause at least one scalar level below the phenomenon or consider abstract factors. Krist et al. (2018) also suggest that non-material factors such as energy or forces like gravity, represent a move from the abstract to the concrete and thus can also be categorized as moving a level below. To help identify abstract factors that relate to the phenomenon of sequence transfer during transcription, we drew upon the ideas presented by Duncan and Reiser (2007), who state that students need to move across ontological levels when reasoning about genetics. They described genetic phenomena such as genes affecting traits as a hybrid hierarchical phenomenon that includes an information level (e.g., genes code for proteins) in addition to a physical level (e.g., the molecular processes of replication).

In *Identifying and Unpacking factors*, Krist et al. (2018) indicate that students must consider the entities present at the lower scalar level or consider abstract level factors. To *Unpack factors*, explanations also should consider the behaviors and interactions causal to the phenomenon. Using factors at the molecular level is a part of a mechanistic explanation in molecular biology according to van Mil et al. (2013), who state that molecules and their interactions lead to molecular activities (van Mil et al., 2013). In *Linking*, Krist et al., (2018) indicate that student explanations need to ‘check’ that the factors they unpacked are relevant to the target phenomenon in that they can be used to explain the cause of the phenomenon.

Mechanistic Rubric

Informed by Jescovitch et al.'s (2019) work on rubrics, which demonstrated that deconstructing a holistic rubric into analytic components can improve both reliability and validity (Jescovitch et al., 2019), we chose to develop an analytic coding scheme to code the responses and then align combined analytic codes to holistic categories. To ensure our analytic codes (underlined) captured pertinent ideas for mechanistic explanations, each code was associated with the elements the EEHMR framework (italicized). Detailed descriptions and example responses of the analytic codes can be found in Supplemental Table 1.

The phenomenon is captured by the analytic code 1. Transfer of Base Sequence from DNA to RNA. The phenomenon is partially included in prompt text for all five items, though the wording differs slightly, all in some manner indicate that information from DNA is transferred to RNA. We recognize that this raises a potential issue, in that students might not recognize a need for an explanation of sequence transfer, and thus will not include any description of the sequence or its transfer in their explanations. Despite this potential limitation, we considered it important for students to recognize that information transfer must include the base sequence and thus, required that explanations either state or imply that the sequence is transferred for coding in this category. We chose to capture the EEHMR *Multiple Levels* code implicitly in our coding of *Identification and Unpacking* of entities a scalar level below. Because the phenomenon is at the macromolecular level (nucleic acids are polymers), the scalar levels below include entities and interactions at the molecular level (nucleotide base-pairing) and the sub-molecular level (non-covalent interactions between bases). We also argue that RNA polymerase activity is at the molecular level on the basis that it catalyzes covalent bond formation between the molecular entity of a nucleotide with the RNA strand. Thus, while RNA polymerase itself

is a macromolecule at the same scalar level as DNA and RNA, we considered its activity for identifying a scalar level.

We used a set of analytic codes to capture Krist et al.'s (2018) *Identifying and Unpacking* with the following categories. We considered the RNA polymerase behavior in synthesizing a new RNA to be similar to Krist's description of structure-behavior-function and thus refers to a factor that can be unpacked in explanations about information flow during transcription. We captured the unpacked factor RNA polymerase with the analytic code 2. RNA Polymerase Activity (RNA Polymerase synthesizes a new RNA). We capture responses that describe the nature of RNA and DNA nucleotide bases as complementary with category 3. Complementary Base-pairing. We found that students often used this idea in two distinct ways, and thus further divided this code into two sub-categories: 3.1. Present and Unlinked for responses that describe RNA and DNA nucleotides as complementary without making connections to sequence transfer or 3.2 Linked to Sequence Transfer for responses that describe that the newly built RNA is complementary to DNA and contains the sequence information from DNA. We capture responses that describe the nature of the base complementarity during transcription with category 4. Non-Covalent Interactions (e.g., hydrogen bonds or molecular shape). Similar to our Complementary Base-pairing code, we further separated this code by whether students linked interactions to other ideas, coding as 4.1. Present and Unlinked for responses that used non-covalent interactions to describe something other than the reason for base-pairing between RNA and DNA including statements that were too vague to accurately categorize, or as 4.2. Linked to Complementarity when the student response used non-covalent interactions as the cause of base pairing specificity between RNA and DNA during transcription, or directly as the cause of sequence transfer.

The final code from the EEHMR framework is *Connect to Target Phenomenon* (Krist et al., 2018). This code requires causally linking one or more unpacked factors to the target phenomenon. Our analytic codes do not capture linkages of multiple unpacked factors to the target phenomenon, thus we used combinations of analytic codes to categorize student responses into holistic categories representing the type of explanation (Table 3). We captured two types of mechanistic explanations. First, responses that include the concept of 3.2 Complementary Base-Pairing linked to 1. Transfer of Base Sequence and an accurate account of 2. RNA Polymerase Activity were considered a Molecular Explanation. Second, we considered Molecular Explanations that also included 4.2 Non-covalent Interactions linked to 3.2 Complementary Base-pairing as Sub-molecular Explanations. Sub-molecular Explanations also include responses in which students directly link 4.2 Non-covalent Interactions to 1. Transfer of Base Sequence plus an accurate account of 2. RNA Polymerase Activity. Responses that linked the ideas of 1. Transfer of Base Sequence with 2. RNA Polymerase Activity were categorized as Descriptions. Responses that contained two or three analytic mechanism codes but lacked a link to 1. Transfer of Base Sequence and/or to 2. RNA Polymerase Activity were categorized as Unlinked, and responses that contained one or fewer analytic mechanism codes were categorized as Other.

Jescovitch and colleagues indicated that analytic codes when recombined could lose capacity to accurately categorize responses along a holistic scale (Jescovitch et al., 2019). Thus, we performed a holistic coding check. Two coders, experts in biology and chemistry, who were not familiar with the analytic coding rubric were introduced to the EEHMR framework, and mechanistic coding rubric with holistic categories from Table 3. The coders independently coded 20 responses holistically as Sub-molecular Explanation,

Molecular Explanation, or Description, then met with one of the analytic coders to discuss and to resolve discrepancies. Following this process, the holistic codes were compared to analytically determined explanation categories and this comparison was used to further refine the analytic rubric such that the codes from the analytic categories accurately reflected the holistic classifications. For details on analytic rubric categories and final coding criteria, see Supplemental Table 1.

Conceptual Rubric

To capture ideas and concepts students used to contextualize their responses, we used qualitative analysis software (QDA Miner Version 5.0.31 Copyright 2004-2016) combined with an emergent coding technique similar to that used by Sripathi et al., (2019). We identified three themes in student responses representing processes involved in gene expression. These are captured in the codes Translation, in which students describe that protein is produced through translation of RNA; Gene Regulation, in which students describe the location of transcriptional initiation as the promoter and/or describe the role of transcription factors in controlling transcription; and RNA Processing, in which students describe one or more mRNA post-processing steps such as poly adenylation or splicing. We also identified three themes that did not represent an additional process in gene expression. These were captured in the codes Phase-Naming, in which students name two or more of the phases of transcription (Initiation, Elongation, and Termination); Cellular Structure, in which students describe DNA and/or RNA in relation to their location in the cell; and General Nucleic Acid Characteristics, in which students describe one or more structural or chemical characteristic of DNA, RNA, or nucleotides. A final category, Nature of Genetic Information captures when students state or imply that the sequence of RNA or

DNA hold information or that RNA or DNA code for proteins. This category captures a key aspect of student understanding of genetic information flow, that students understand what is meant by the term genetic information, and how nucleic acids can store information. For detailed descriptions of the conceptual rubric, including coding rules and example responses see Supplemental Table 2.

Coding and Inter-Rater Reliability

To code student responses, two individuals with Ph.Ds. in Biology independently coded 20 responses using the analytic rubric, met to discuss discrepancies and to revise and refine rubric criteria. This process was repeated three more times with new responses (for a total of 80 responses) until the inter-rater reliability (IRR) of each analytic category measured by Cohen's Kappa (Cohen, 1960) exceeded 0.6 (considered substantial agreement per Landis & Koch, 1977) and coding rubric rules were considered complete (Supplemental Table 3). Following this, the two coders independently coded 40 responses and met to resolve discrepancies. This process was repeated until all 346 responses were consensus coded. It is documented that students often confuse the names of the entities and processes involved in gene expression (Fisher, 1985; Southard et al., 2016; Zukswert et al., 2019). Thus, in cases of uncertainty about the named entities or processes, coders used the entirety of the response when coding to assign the code. That is, when a student used unconventional language to describe a process as well as more typical language, coders used the more typical language in assigning codes. In cases where consensus could not be reached, those responses were brought to a third coder with a Ph.D. in Biology and discussed among all three coders until consensus was reached. All responses were reviewed by the primary coders to ensure final codes were consistent with the final rubric rules. The final codes

presented here have all been agreed upon by two coders, subject to a holistic code check, and a consistency check.

Data and Statistical Analysis

We performed a Kruskal-Wallis test to compare numbers of context ideas based on explanation type and Mann-Whitney tests for post-hoc pairwise analysis. Effect sizes for Mann-Whitney test were estimated according to Fritz et al., (2012). For Mann-Whitney tests, p values are reported in article main text and other statistics (i.e., test statistic and effect sizes) are reported in full in the Supplemental Materials. Phi correlation coefficient was calculated to compare the degree to which individual ideas co-occurred with each explanation type. A Chi-square test was performed to compare the proportions of explanation type based on prompt version. All statistical tests were performed in SPSS, Version 24.

Results

We found that over half of undergraduate explanations in our sample described information transfer during transcription by linking RNA Polymerase Activity synthesizing an RNA with the idea Transfer of Base Sequence. Many of those students also leveraged Complementary Base-pairing or Non-covalent Interactions mechanistically. The remaining students either included these ideas but failed to link these ideas with the target phenomenon or focused on concepts that did not address the phenomenon of information transfer during transcription. We also identified ideas that students included in their explanations and found differences in quantity and content of context in student mechanistic explanations compared to non-mechanistic.

Mechanistic responses

We categorized responses that included the concept of 3. Complementary Base-pairing in addition to 1. Transfer of Base Sequence and 2. RNA Polymerase Activity in student responses as Molecular mechanisms. For example, upper-level student 4256 wrote in response to version 5 (underlining added by authors to emphasize coded phrases),

During transcription, RNA polymerase is used to copy the information from the DNA sequence and create an appropriate mRNA template. The RNA polymerase goes from the 5' end to the 3' end and builds RNA in an antiparallel fashion. DNA and RNA hold information in their nitrogen bases, Adenine, Thymine, Guanine, and Cytosine. Each of these bases has a matching pair that they bond with, which creates the genetic code. The RNA polymerase is able to create the RNA by copying the genetic code and the base pairs match with each other.

This student describes that RNA polymerase copies the DNA sequence to the mRNA. They go on to describe complementary base-pairing by stating that each base has a matching pair they bond with, and that this is the reason the code is copied. Other students list the base-pairs (A with T/U and C with G) and/or give a more detailed account of base-pairing during RNA synthesis. For example, introductory student 4302 wrote in response to item version 1,

During the process of transcription, the DNA double helix is unwound, separating the DNA into a template strand and a noncoding strand for a short series of nucleotides in a sort of transcription bubble. This process starts at a promoter region, where proteins, including RNA polymerase bind. The RNA polymerase takes nucleoside triphosphates (NTPs) and matches them with the complementary nucleotides in DNA. In RNA, A matches with T, G with C, C with G and U with A. As each new nucleotide is added, a phosphodiester bond will link the nucleotides. The base pairing continues along the

DNA molecule until the terminator region is reached. At this point, transcription is over, and the new RNA molecule is released from the RNA polymerase complex and the DNA molecule is reconnected back to its double helix shape. The sequence of nucleotides in the RNA molecule will go on to be the code for protein during the process of translation.

This student also wrote a more detailed explanation of the process of RNA synthesis by RNA polymerase, stating that RNA nucleoside triphosphates are matched with complementary nucleotides in DNA, and that phosphodiester bonds form between the nucleotides. Each of these ideas builds toward the idea that a sequence in DNA is transferred to the RNA, and the student makes that explicit at the end by stating that the sequence of nucleotides in the RNA will go on for translation. While the level of detail a student uses to describe RNA synthesis varies, we did not view a highly detailed account of RNA polymerization as necessary to adequately respond to these items. However, students who provided a highly detailed account of RNA polymerization also tended to be capable of writing a mechanistic explanation.

Molecular responses that also described hydrogen bonding, purine and pyrimidine interactions, or other non-covalent interactions as the reason that complementary base pairing occurred were categorized as a Sub-molecular. For example, introductory student 2005 wrote in response to item version 4,

An RNA polymerase enzyme attaches to a DNA strand that is targeted for its code for a cell's desired protein. The enzyme "unzips" the DNA double helix, matches up complementary nucleotides to form an mRNA strand, then re-attaches the two strands to DNA. DNA and RNA are able to hold information because their nucleic bases form unique codes for the proper order of amino acids added to a peptide chain during protein synthesis. The information *is* the order of the nucleotides. Each nucleotide forms a hydrogen-bonded base with exactly one other nucleic base... or two bases in

the case of adenine. The specific combination requirements (mostly) guarantee that the correct information is passed along the transcription and translation process chains to create properly constructed, structured, and functioning proteins.

This student begins by stating that RNA polymerase forms an mRNA strand and is clear that complementary nucleotides are matched during formation. The student then gives evidence that they understand the concept of sequence transfer from DNA to RNA by defining the information in two ways; that bases form codes for amino acid order as well as stating that the order of the nucleotides is the information. They then complete their description of why base-pairing occurs by explaining that hydrogen bonding is specific to base pairs, and that is the reason for information transfer during transcription. This response may seem a bit out of order, but that is likely a result of the format of this question version, which asked in the final part of the item how the characteristics of nucleotides allows information transfer. A more detailed description of the items follows; however, here we wish to point out that we allowed in our coding for students to link entity behavior and characteristics in a non-linear fashion.

Descriptive responses

Some students wrote responses we categorized as Descriptions, which we did not consider mechanistic. These students described that the sequence was transferred by the RNA polymerase synthesizing a new RNA without leveraging the properties or activities of the nucleotide bases in their response. For example, introductory student 1204 in response to version 5 wrote,

Transcription is the first step during gene expression. By copying a gene's DNA sequence you can make a RNA molecule. This process is performed using the enzyme

RNA polymerase which is used to link the nucleotide sequences to form an RNA strand from a DNA template. DNA and RNA hold information due to the nucleotide sequences they possess which correctly links base pairs together (amino acids) which can then be used to make proteins. Because DNA and RNA are made of long chains of nucleic acids, which carry out information that can be passed down from generation to generation through the series of base pairs present that provide a way for copying genetic information from an existing nucleotide to form a new one.

In the third sentence, the student clearly states that the enzyme RNA polymerase links nucleotide sequences to form an RNA. This student also wrote earlier in the response that the DNA sequence was copied to make an RNA molecule. Later in the response the student appears to indicate that base pairs linked together are amino acids and used to create proteins; we interpreted this portion of the response to be related to the student putting the entire process of gene expression into the context of translation. Other students implied that the sequence was transferred from DNA to RNA by stating that the sequence of DNA held information and that information was transferred during transcription. For example, introductory biology student 7793 indicated that the order of bases holds information (the genetic code, further defined as genetic information), and that the RNA polymerase uses the DNA template to form an RNA in their response to version 2,

During transcription, the information stored in a gene's DNA is transferred to RNA. DNA is made of molecules called nucleotides. The order of the bases is the thing that determines the genetic code. DNA is better at holding genetic information because it is a stable storage form. RNA contains a ribose sugar that makes it more reactive. Transcription uses DNA as a template to make an RNA molecule. An RNA polymerase enzyme strand comes in to separate the template and its nontemplate strand. then forms an mRNA that is later used in translation to a protein.

These responses demonstrate these students understand that the DNA has a sequence, which is transferred to RNA during transcription. However, they either believe complementary base-pairing is implied in the text of their response, is obvious to the instructor and thus does not need to be written, do not think it is important to include, or do not recognize or understand the importance of complementary base-pairing in transcription.

Unlinked explanations

Many students were capable of incorporating some, but not all, of the ideas related to explaining sequence transfer during transcription. Some students provided a reasonable account of complementary base-pairing during RNA polymerase activity without attending to the relationship between RNA and DNA sequence. For example, introductory student 8443 wrote in response to version 3,

Initially DNA is split by the RNA polymerase and is copied into mRNA. The RNA polymerase reads the DNA starting from promoters and moves from 3' to 5' copying the genetic code with bases except replacing thymine with uracil. Only certain bases are joined together and this ensures that the genetic code is copied correctly although there can be possible mutation which can affect how the RNA is transcribed.

This student describes base-pairing as only certain bases are joined together as well as RNA polymerase 'reads' and copies bases. We consider this type of response similar to a descriptive response, in that students are describing RNA polymerase activity without linking to the full phenomenon of sequence transfer or leveraging molecular or sub-molecular entities properties or activities. We were cautious when coding responses that included statements that DNA is copied into mRNA or used terms like 'genetic code' without further elaboration. We wanted to identify only responses that were less likely to

have been learned by rote memory and thus chose not to code DNA copied to mRNA or genetic code copied or similar as the idea of Transfer of Base Sequence.

Another type of response that we encountered included Complementary Base pairing without linking to RNA polymerase activity or to Transfer of Base Sequence. For example, introductory student 5143 wrote in response to version 2,

The double helix two-stranded structure of DNA allows the information within DNA to be accurately transmitted. Each nitrogen base within the helix is uniformly paired to a complementary base through hydrogen bonding interactions. The purine-pyrimidine pairs are specifically bonded to one another so that when the DNA strands unwind, the template strand that be transcribed to create a complementary pre-mRNA strand. RNA is a single stranded molecule composed of nucleotides lined by phosphodiester bonds. There are also complementary bases but instead of thymine, there is the pyrimidine base uracil. Though RNA is single stranded, it can fold on itself which can be stabilized by small areas of base pairing. This allows the RNA molecule to stabilize throughout the translation process.

This student uses both hydrogen bonds and purine-pyrimidine interactions to explain complementary base pairing in the context of DNA then extends those characteristics and activities to RNA. This student included a thorough description of base pairing in DNA, RNA structure, and during transcription. This type of explanation in which students give descriptions of the mechanism of base-pairing without linking back the phenomenon or to RNA polymerase activity is rare and may be representative of responses from students who felt they did not need to describe sequence transfer or RNA polymerase.

Other Explanations

The remaining students included one or fewer ideas categorized in our mechanistic rubric and were categorized holistically as Other. Some of these responses included ideas that

further described the phenomenon of information transfer during transcription by defining the information as being held in a sequence of nucleotides that is then transferred such as introductory student 2475 who wrote in response to version 5, “The central dogma describes the flow of information from DNA to RNA to Protein. The specific nucleotide sequence allows DNA and RNA to hold specific information. Nucleotides hold genetic information and can use it to make specific proteins.” This response was categorized as containing the Transfer of Base Sequence idea. Some Other responses included no mechanistic ideas, such as introductory student 8267 who wrote in response to version 4, “Strand of DNA gets copied into mRNA, storing the DNA in the nucleus. The double helix in DNA and RNA uses the ribosomes in the cytoplasm to carry it. They are acidic and found in all living things.” This student focused on ideas related to the structure (double helix) of DNA and the cellular locations of DNA and RNA (nucleus, ribosomes, and cytoplasm).

Context in Student Explanations

We were interested in other ideas or related processes student drew upon in their responses and thus applied the conceptual rubric to categorize and quantify ideas related to context.

We captured seven other concepts that students used in their explanations: Gene Regulation, which includes the role of the promoter or transcription factors; mRNA Processing, which includes post-transcriptional modifications to RNA such as splicing; Translation, where students state the RNA will be translated to a protein; Phase Naming, in which students say there are or name the three phases of transcription; Cellular Structure, in which students state where in the cell a process or molecule occurs; General Nucleic Acid Characteristics, in which students explain DNA and/or RNA structure such as describing the base, sugar, and phosphate components of a nucleotide or the double-helical nature of

DNA; and Nature of Genetic Information, in which responses state genetic information is present in the sequence or that nucleic acids code for proteins. We found that the most frequent context idea was Nature of Genetic Information, which occurred in 63% of responses and the least common context idea was mRNA Processing, which occurred in 10% of responses. (See Supplemental Table 2 for details and examples).

We examined whether student responses categorized as Mechanistic (either Molecular or Sub-molecular) would include similar numbers of context ideas as student responses categorized as Descriptions, Unlinked, or Other. We found that students included a median of one context idea in responses categorized as Unlinked, two context ideas in responses categorized as Other or Descriptions, and a median of three context ideas in responses categorized as Molecular or Sub-molecular (Figure 1). We performed a Kruskal-Wallis test and found a difference in the numbers of ideas included based on explanation type (Kruskal-Wallis $H = 44.448$, $df = 4$, $p < 0.005$). To more closely examine the differences in number of context ideas included based on response type, we performed pairwise Mann-Whitney tests and found a significant difference between the ideas in responses categorized as Other and Molecular ($p < 0.005$) and Sub-molecular ($p < 0.005$). Descriptions differed slightly from Molecular ($p < 0.065$) and significantly from Sub-molecular ($p = 0.013$; see Supplemental Table 4 for U statistic and effect sizes). We found no evidence of a difference in context ideas included in responses categorized as Molecular compared to Sub-Molecular ($p = 0.359$). Response lengths differed between different explanation types; the median word count of a Sub-molecular response was 171.5 words, Molecular was 140 words, Descriptions 89 words, Unlinked 109 words, and Other 79.5 words (Supplemental Figure 1). Thus, most students who wrote a Molecular or Sub-molecular explanation included more content as measured by word count and more context

as measured by number of ideas than students who wrote Description, Unlinked, or Other responses.

To explore the context that students included alongside their explanations, we identified the most frequent context idea in student responses categorized as a Description, Molecular, or Sub-molecular, which was the idea Nature of Genetic Information (86%, 78%, 89% respectively; Supplemental Table 5). This idea is significantly positively correlated to each of the three response types based on the Phi Coefficient (Supplemental Table 6). The frequency of Nature of Genetic Information in these categories is partially an effect of the coding rubric and rules as the rubric allowed combining Nature of Genetic Information ideas with the concept that RNA is made by sequential addition of bases into the phenomenon of Transfer of Base Sequence. However, presence of the Nature of Genetic Information idea was insufficient for serving as the phenomenon in coding rules, and 43% and 42% of responses categorized as Other or Unlinked also included this idea.

The second most frequent context idea was General Nucleic Acid Characteristics, which occurred in 40% of Descriptions, 60% of Molecular, and 64% of Sub-molecular responses. General Nucleic Acid Characteristics was also the most frequent context idea occurring in Other and Unlinked responses (50% and 59%). However, there was a negative and significant correlation between General Nucleic Acid Characteristics with Description, and no significant correlation between any other response types. The context of Gene Regulation occurred frequently in, and was positively correlated with, both Molecular and Sub-molecular explanations and negatively correlated with Other responses. Finally, we highlight that no context ideas are positively correlated with responses categorized as Other, all are either negatively or uncorrelated. It appears that students using Molecular or

Sub-molecular explanations include more context ideas related to ideas about gene regulation and genetic information than students producing non-mechanistic explanations.

Items that separately prompt for DNA, RNA, and nucleotide characteristics elicit mechanistic explanations more often than other prompts

We next asked whether one or more of the item versions were more likely to elicit mechanistic explanations than others. We found that the lowest percent of mechanistic (Molecular and Sub-molecular) explanations arose from versions 2 and 3 (29% each) and the highest from version 5 (50%; see Figure 2). Version 1 resulted in 40% mechanistic explanations and version 4 in 33%. We performed a Chi-square test and found that the differences in explanation types by prompt version were significant ($X^2 = 34.5$, $df = 16$, $p < 0.005$), with an effect size of 0.158. We found that version 5 elicited a larger percentage of Sub-molecular explanations than other versions.

All five versions included a prompt for students to explain how information is transferred from DNA to RNA during transcription plus a prompt to describe or use the characteristics of DNA and RNA to help explain information storage. Versions 4 and 5 included three text entry boxes, one for the explanation, a second for DNA and RNA characteristics and information storage and a third in which they were prompted to describe or use the characteristics of the nucleotide to explain information transfer. Many students responded to parts B and C of versions 4 and 5 by naming structural components of a nucleotide: nitrogenous base, sugar, and phosphate. While this was not strictly necessary for a student to describe information transfer, it did seem to prompt students to think about the molecular level; that is the role of the base and subsequently about the role of non-covalent interactions.

For example, in response to part A, introductory student 2534 wrote in response to version 5, “Copying part of a nucleotide sequence of DNA into a complementary sequence in RNA. This process is called process of transcription. It includes enzymes like DNA polymerase.” In response to part B, they wrote “Two nucleotides are bound together by phosphate bond. Two strands of DNA and RNA are held together by hydrogen bonds between two nitrogen bases.” In response to part C they wrote, “Nucleotides have nitrogen bases on one of their ends and those nitrogen bases keep the encoded information in them. These Nucleotides are held together by phosphate bond which do not break easily but two nitrogen bases are held together by hydrogen bond which are easy to break, thus can help in uncoiling of DNA while no breaking the whole structure.” This student described hydrogen bonding as related to complementary base pairing and sequence transfer in response to part B and went on to explain the role of the base vs. the role of the nucleotide in part C.

Versions 1, 2, and 3 were also developed with the intent to elicit sub-molecular and molecular ideas. Specifically, version 3 prompted students to think about the ‘molecular events,’ which was intended to prompt a molecular level explanation, but student responses to version 3 had similar proportion of Molecular explanations to versions 1 and 2, and very few (2%) students were categorized as Sub-molecular (Figure 2). Thus, simply calling attention to the molecular level isn’t enough to elicit molecular thinking. It appears that the scaffolding in the form of multiple prompts and entry boxes, which draw attention first to the process, then names and asks for explanations about the macromolecular (DNA and RNA) and molecular (nucleotides) levels are more productive ways to prompt students to provide a mechanistic explanation. The wording of item version 5 indicated that a new RNA was formed, which was intended to provide the element of time as a point of comparison for students (as suggested by Krist et al., 2018). As version 5 elicited the most

Sub-molecular explanations, the addition of the time element may have enhanced the effect of scaffolding.

Discussion

In this study, we found that undergraduates are capable of writing a mechanistic explanation of sequence transfer during transcription and are more successful at doing so when items separately prompted students to respond about the characteristics of DNA, RNA, and nucleotides in holding and transferring information. Further, we found that students who provided a mechanistic explanation included more context than students who wrote a non-mechanistic explanation and were more likely to include ideas about gene regulation and genetic information.

In student CRs to the “DNA to RNA Information Transfer” prompts, we found that most students (66%) included the phenomenon 1. Transfer of Base Sequence from DNA to RNA in their responses, which was a part of multiple types of mechanistic and non-mechanistic explanations. The EEHMR framework describes the aspects of a mechanistic explanation with the codes, *Multiple Levels*, *Identifying and Unpacking Factors*, and *Linking* (Krist et al., 2018). Our coding scheme did not separately capture *Multiple Levels*, instead this code is captured as part of *Identifying and Unpacking Factors*. We captured *Identifying and Unpacking Factors* with the following codes: 2. RNA Polymerase Activity (64%); 3.1 Complementary Base-pairing, Present and Unlinked (18%); 3.2 Complementary Base-pairing, Linked to Sequence Transfer (49%); 4.1 Non-covalent Interactions, Present and Unlinked (1%); and 4.2 Non-covalent Interactions, Linked to Complementarity or Sequence Transfer (22%). In our coding scheme, 2, 3.2, and 4.2 both *Identify and Unpack Factors* and 3.1 and 4.1 either only *Identify Factors* or incorrectly link or describe the

activities of the factors. We found that 86% of students identified one or more factor, and 78% of students unpacked one or more factors. Undergraduates in this study were largely able to identify and unpack one or more factors as well as explicate the phenomenon beyond what was wrote in the prompt.

For this work, our interest was primarily whether undergraduates could write CRs categorized as *Linking* per the EEHMR framework. That is, we were interested in student ability to link the phenomenon of 1. Transfer of Base Sequence, which was commonly included in student CRs, with the underlying mechanism of 2. RNA Polymerase Activity, 3. Complementary Base-pairing and 4. Non-covalent Interactions. We found that 38% of students *linked* the above ideas as captured by the holistic codes Molecular Explanation (25%) and Sub-molecular Explanation (13%).

Educational Implications

We found many students *Identified or Unpacked Factors* without *Linking*. For example, some students described aspects of transcription (e.g., RNA polymerase transfers the sequence) without linking to ideas about 3. Complementary Base-pairing or 4. Non-covalent Interactions. Reasons these students may not include the full mechanism include they do not recognize which additional entities interact to produce an effect. We suggest that instruction focus on helping students build on the ideas they hold by encouraging discussion about the application of ideas they may have memorized such as applying the idea of base-pairing, often learned about in relation to the DNA double helix, to the phenomenon of sequence transfer. Some students CRs were categorized as Unlinked because they had mis-assigned the role of RNA polymerase or wrote that the mRNA itself transcribes RNA or claimed that another entity (e.g., the ribosome or helicase) synthesizes

RNA. These students might be considered to have mechanistic conceptions of how transcription works and have simply mistaken terms, which is a documented learning challenge (Zukswert et al., 2019). In these cases, instructors might develop or utilize tools to help clear up confusion between the three Central Dogma processes (e.g. Pelletreau et al., 2016).

Chemical concepts underlie many molecular biological phenomena, but students struggle to make connections between the two fields (e.g. Kohn et al., 2018; Loertscher et al., 2014; Roche Allred et al., 2021). We found this in our work about information transfer as well; many students could provide an explanation that included complementary base-pairing as a mechanism, but they less frequently included the non-covalent interactions between bases. This is a concern for instruction, as memorizing a set of base pairs (A-T or A-U and G-C) makes it difficult for an instructor to know whether the student considers these ‘letters’ to represent molecules with interactions or if the student only uses them procedurally. That is, a student can use the base pairs to produce a DNA-RNA sequence or even a protein sequence without understanding the cause or effect behind the rules. The underlying cause of base-pairing is non-covalent interactions, which underlie many other molecular and cellular phenomena, such as protein-protein interactions. Loertscher et al. (2014) identified that students struggle with understanding non-covalent interactions, rather they memorized types of interactions. The authors also found that when students understand these interactions, they can begin to fully understand and make connections between structure and function, a core biological and scientific concept (Loertscher et al., 2014). We suggest that specifically prompting for nucleotide characteristics, as in item versions 4 and 5, may promote students making connections between chemical interactions

and their relationship to biological phenomena like base-pairing and even key biological core ideas like information transfer and structure-function relationships.

In our work, we found that students were more likely to elaborate on the idea that the information transfer between DNA and RNA is in the form of a nucleotide sequence when we prompted separately for how DNA and RNA hold information. Duncan and Reiser (2007) described genetic information flow between genes and proteins as a genetic hybrid hierarchical phenomenon, which requires understanding and connecting between two organizational levels: information level (genes) and the physical level (hierarchical entities like proteins, cells, tissues). The authors suggested that this is one of the main challenges of learning about molecular genetics (Duncan & Reiser, 2007). In our assessment items with multiple parts, getting students to think about the nature of information by prompting in a second part was related to students providing a mechanistic explanation. Thus, we suggest that using this item as a formative assessment could aid in driving student thinking about the information level in addition to the physical/mechanical aspects of transcription and improve student ability to construct a mechanistic explanation.

Limitations

Responses collected for versions 4 and 5 were collected shortly after COVID-19 forced instruction to move online, thus we hesitate to draw broad conclusions about student ability to write a mechanistic explanation of transcription, as this rapid shift to remote learning may have affected student ability to participate in class, access materials, or focus given potential personal concerns. We are currently working on a study to characterize student explanations under the current circumstance of planned remote learning (rather than emergency remote learning).

While we collected data from introductory and upper-level students, who have different levels of chemistry background, we cannot draw conclusions regarding abilities of these students to write mechanistic responses. This is in part because we did not collect sufficient data from all question versions to make a fair comparison, while accounting for the potential impacts of COVID. Ongoing work aims to use the question developed in this work to compare student responses across educational backgrounds.

Because this study took place at one large research-intensive university, student demographics may not be representative of the larger college student population. Ongoing work aims to include student responses from additional colleges and universities to represent a broader demographic.

Ethical Approval

The study was designated exempt by Michigan State University's Institutional Review Board (IRB x10-577). This project was done under a waiver of informed consent. All student responses collected were part of normal course assessment practice and the study presented minimal risk to participants. Responses were anonymized before being used by the research team.

Funding

This work was supported by the National Science Foundation under grant number DUE 1323162.

Acknowledgements

The authors thank the Automated Analysis of Constructed Response (AACR) collaboration for helpful conversations while developing the item and coding rubric, especially Drs. Jenifer Saldanha, Leonora Kaldaras, and John Merrill. Details about the prompt “DNA to RNA Information Transfer” can be found at beyondmultiplechoice.org.

References

- American Association for the Advancement of Science. (2011). Vision and change in undergraduate biology education: A call to action. Final report. Washington, DC.
- Braaten, M., & Windschitl, M. (2011). Working toward a stronger conceptualization of scientific explanation for science education. *Science Education*, 95(4), 639–669. <https://doi.org/10.1002/sce.20449>
- Briggs, A. G., Hughes, L. E., Brennan, R. E., Buchner, J., Horak, R. E. A., Amburn, D. S. K., McDonald, A. H., Primm, T. P., Smith, A. C., Stevens, A. M., Yung, S. B., & Paustian, T. D. (2017). Concept Inventory Development Reveals Common Student Misconceptions about Microbiology. *Journal of Microbiology & Biology Education*, 18(3). <https://doi.org/10.1128/jmbe.v18i3.1319>
- Briggs, A. G., Morgan, S. K., Sanderson, S. K., Schulting, M. C., & Wieseman, L. J. (2016). Tracking the Resolution of Student Misconceptions about the Central Dogma of Molecular Biology. *Journal of Microbiology & Biology Education*, 17(3), 339–350. <https://doi.org/10.1128/jmbe.v17i3.1165>
- Castro-Faix, M., Duncan, R. G., & Choi, J. (2021). Data-driven refinements of a genetics learning progression. *Journal of Research in Science Teaching*, 58(1), 3–39. <https://doi.org/10.1002/tea.21631>
- Clemmons, A. W., Timbrook, J., Herron, J. C., & Crowe, A. J. (2020). BioSkills Guide: Development and National Validation of a Tool for Interpreting the Vision and Change Core Competencies. *CBE—Life Sciences Education*, 19(4), ar53. <https://doi.org/10.1187/cbe.19-11-0259>

- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46.
<https://doi.org/10.1177/001316446002000104>
- Cooper, M. M., Caballero, M. D., Ebert-May, D., Fata-Hartley, C. L., Jardeleza, S. E., Krajcik, J. S., Laverty, J. T., Matz, R. L., Posey, L. A., & Underwood, S. M. (2015). Challenge faculty to transform STEM learning. *Science*, 350(6258), 281–282.
<https://doi.org/10.1126/science.aab0933>
- Duncan, R. G., & Reiser, B. J. (2007). Reasoning across ontologically distinct levels: Students' understandings of molecular genetics. *Journal of Research in Science Teaching*, 44(7), 938–959. <https://doi.org/10.1002/tea.20186>
- Fisher, K. M. (1985). A misconception in biology: Amino acids and translation. *Journal of Research in Science Teaching*, 22(1), 53–62. <https://doi.org/10.1002/tea.3660220105>
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2–18. <https://doi.org/10.1037/a0024338>
- Gericke, N., & Wahlberg, S. (2013). Clusters of concepts in molecular genetics: A study of Swedish upper secondary science students understanding. *Journal of Biological Education*, 47(2), 73–83. <https://doi.org/10.1080/00219266.2012.716785>
- Haskel-Ittah, M., & Yarden, A. (2018). Students' Conception of Genetic Phenomena and Its Effect on Their Ability to Understand the Underlying Mechanism. *CBE—Life Sciences Education*, 17(3), ar36. <https://doi.org/10.1187/cbe.18-01-0014>
- Jescovitch, L. N., Scott, E. E., Cerchiara, J. A., Doherty, J. H., Wenderoth, M. P., Merrill, J. E., Urban-Lurain, M., & Haudek, K. C. (2019). Deconstruction of Holistic Rubrics into Analytic Rubrics for Large-Scale Assessments of Students' Reasoning of

Complex Science Concepts. *Practical Assessment, Research & Evaluation*, 24(7).

<https://par.nsf.gov/biblio/10112967>

Kohn, K. P., Underwood, S. M., & Cooper, M. M. (2018). Energy Connections and Misconnections across Chemistry and Biology. *CBE—Life Sciences Education*, 17(1), ar3. <https://doi.org/10.1187/cbe.17-08-0169>

Krist, C., Schwarz, C. V., & Reiser, B. J. (2018). Identifying Essential Epistemic Heuristics for Guiding Mechanistic Reasoning in Science Learning. *Journal of the Learning Sciences*, 1–46. <https://doi.org/10.1080/10508406.2018.1510404>

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174. JSTOR. <https://doi.org/10.2307/2529310>

Loertscher, J., Green, D., Lewis, J. E., Lin, S., & Minderhout, V. (2014). Identification of Threshold Concepts for Biochemistry. *CBE—Life Sciences Education*, 13(3), 516–528. <https://doi.org/10.1187/cbe.14-04-0066>

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about Mechanisms. *Philosophy of Science*, 67(1), 1–25. <https://doi.org/10.1086/392759>

Marbach-Ad, G., & Stavy, R. (2000). Students' cellular and molecular explanations of genetic phenomena. *Journal of Biological Education*, 34(4), 200–205. <https://doi.org/10.1080/00219266.2000.9655718>

Matz, R. L., Fata-Hartley, C. L., Posey, L. A., Laverty, J. T., Underwood, S. M., Carmel, J. H., Herrington, D. G., Stowe, R. L., Caballero, M. D., Ebert-May, D., & Cooper, M. M. (2018). Evaluating the extent of a large-scale transformation in gateway science courses. *Science Advances*, 4(10), eaau0554. <https://doi.org/10.1126/sciadv.aau0554>

- McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting Students' Construction of Scientific Explanations by Fading Scaffolds in Instructional Materials. *Journal of the Learning Sciences*, 15(2), 153–191.
https://doi.org/10.1207/s15327809jls1502_1
- Moscarella, R. A., Haudek, K. C., Knight, J. K., Mazur, A., Pelletreau, K. N., Prevost, L. B., Smith, M. K., Steele, M., Urban-Lurain, M., & Merrill, J. E. (2016, April). Automated Analysis Provides Insights into Students' Challenges Understanding the Processes Underlying the Flow of Genetic Information. National Association for Research in Science Teaching Annual International Conference, Baltimore, MD. National Research Council. 2001. *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, D.C.: National Academies Press. doi:10.17226/10019
- National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. National Academies Press.
<https://doi.org/10.17226/13165>
- Nehm, R. H., & Schonfeld, I. S. (2008). Measuring knowledge of natural selection: A comparison of the CINS, an open-response instrument, and an oral interview. *Journal of Research in Science Teaching*, 45(10), 1131–1160.
<https://doi.org/10.1002/tea.20251>
- NGSS Lead States. (2013). *Next Generation Science Standards; For States, By States*. The National Academies Press. <https://www.nextgenscience.org/>

- Parker, J. M., Anderson, C. W., Heidemann, M., Merrill, J., Merritt, B., Richmond, G., & Urban-Lurain, M. (2012). Exploring Undergraduates' Understanding of Photosynthesis Using Diagnostic Question Clusters. *CBE—Life Sciences Education*, 11(1), 47–57. <https://doi.org/10.1187/cbe.11-07-0054>
- Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A Framework for Conceptualizing and Evaluating the Validity of Instructionally Relevant Assessments. *Educational Psychologist*, 51(1), 59–81. <https://doi.org/10.1080/00461520.2016.1145550>
- Pelletreau, K. N., Andrews, T., Armstrong, N., Bedell, M. A., Dastoor, F., Dean, N., Erster, S., Fata-Hartley, C., Guild, N., Greig, H., Hall, D., Knight, J. K., Koslowsky, D., Lemons, P., Martin, J., McCourt, J., Merrill, J., Moscarella, R., Nehm, R., ... Smith, M. K. (2016). A clicker-based case study that untangles student thinking about the processes in the central dogma. *CourseSource*, 3. <https://doi.org/10.24918/cs.2016.15>
- Prevost, L. B., Smith, M. K., & Knight, J. K. (2016). Using Student Writing and Lexical Analysis to Reveal Student Thinking about the Role of Stop Codons in the Central Dogma. *CBE—Life Sciences Education*, 15(4), ar65. <https://doi.org/10.1187/cbe.15-12-0267>
- Reiser, B. J., Berland, L. K., & Kenyon, L. (2012). Engaging Students in the Scientific Practices of Explanation and Argumentation. *Science and Children*, 49(8), 8–13.
- Roche Allred, Z. D., Farias, A. J., Kararo, A. T., Parent, K. N., Matz, R. L., & Underwood, S. M. (2021). Students' use of chemistry core ideas to explain the structure and stability of DNA. *Biochemistry and Molecular Biology Education*, 49, 55–68. <https://doi.org/10.1002/bmb.21391>

- Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, 92(3), 499–525.
<https://doi.org/10.1002/sce.20264>
- Sandoval, W. A., & Millwood, K. A. (2005). The Quality of Students' Use of Evidence in Written Scientific Explanations. *Cognition and Instruction*, 23(1), 23–55.
https://doi.org/10.1207/s1532690xci2301_2
- Smith, M. K., & Knight, J. K. (2012). Using the Genetics Concept Assessment to Document Persistent Conceptual Difficulties in Undergraduate Genetics Courses. *Genetics*, 191(1), 21–32. <https://doi.org/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. <https://doi.org/10.1187/cbe.08-08-0045>
- Southard, K. M., Espindola, M. R., Zaepfel, S. D., & Bolger, M. S. (2017). Generative mechanistic explanation building in undergraduate molecular and cellular biology. *International Journal of Science Education*, 39(13), 1795–1829.
<https://doi.org/10.1080/09500693.2017.1353713>
- Southard, K., Wince, T., Meddleton, S., & Bolger, M. S. (2016). Features of Knowledge Building in Biology: Understanding Undergraduate Students' Ideas about Molecular Mechanisms. *CBE—Life Sciences Education*, 15(1), ar7.
<https://doi.org/10.1187/cbe.15-05-0114>
- Sripathi, K. N., Moscarella, R. A., Yoho, R., You, H. S., Urban-Lurain, M., Merrill, J., & Haudek, K. (2019). Mixed Student Ideas about Mechanisms of Human Weight Loss. *CBE—Life Sciences Education*, 18(3), ar37. <https://doi.org/10.1187/cbe.18-11-0227>

- Todd, A., & Romine, W. (2018). The Learning Loss Effect in Genetics: What Ideas Do Students Retain or Lose after Instruction? *CBE—Life Sciences Education*, 17(4), ar55. <https://doi.org/10.1187/cbe.16-10-0310>
- Trujillo, C. M., Anderson, T. R., & Pelaez, N. J. (2015). A Model of How Different Biology Experts Explain Molecular and Cellular Mechanisms. *CBE—Life Sciences Education*, 14(2), ar20. <https://doi.org/10.1187/cbe.14-12-0229>
- van Mil, M. H. W., Boerwinkel, D. J., & Waarlo, A. J. (2013). Modelling Molecular Mechanisms: A Framework of Scientific Reasoning to Construct Molecular-Level Explanations for Cellular Behaviour. *Science & Education*, 22(1), 93–118. <https://doi.org/10.1007/s11191-011-9379-7>
- Weston, M., Haudek, K. C., Prevost, L., Urban-Lurain, M., & Merrill, J. (2015). Examining the Impact of Question Surface Features on Students' Answers to Constructed-Response Questions on Photosynthesis. *CBE—Life Sciences Education*, 14(2), ar19. <https://doi.org/10.1187/cbe.14-07-0110>
- 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. <https://doi.org/10.1187/cbe.cbe-13-09-0188>
- Zukswert, J. M., Barker, M. K., & McDonnell, L. (2019). Identifying Troublesome Jargon in Biology: Discrepancies between Student Performance and Perceived Understanding. *CBE—Life Sciences Education*, 18(1), ar6. <https://doi.org/10.1187/cbe.17-07-0118>

Figures

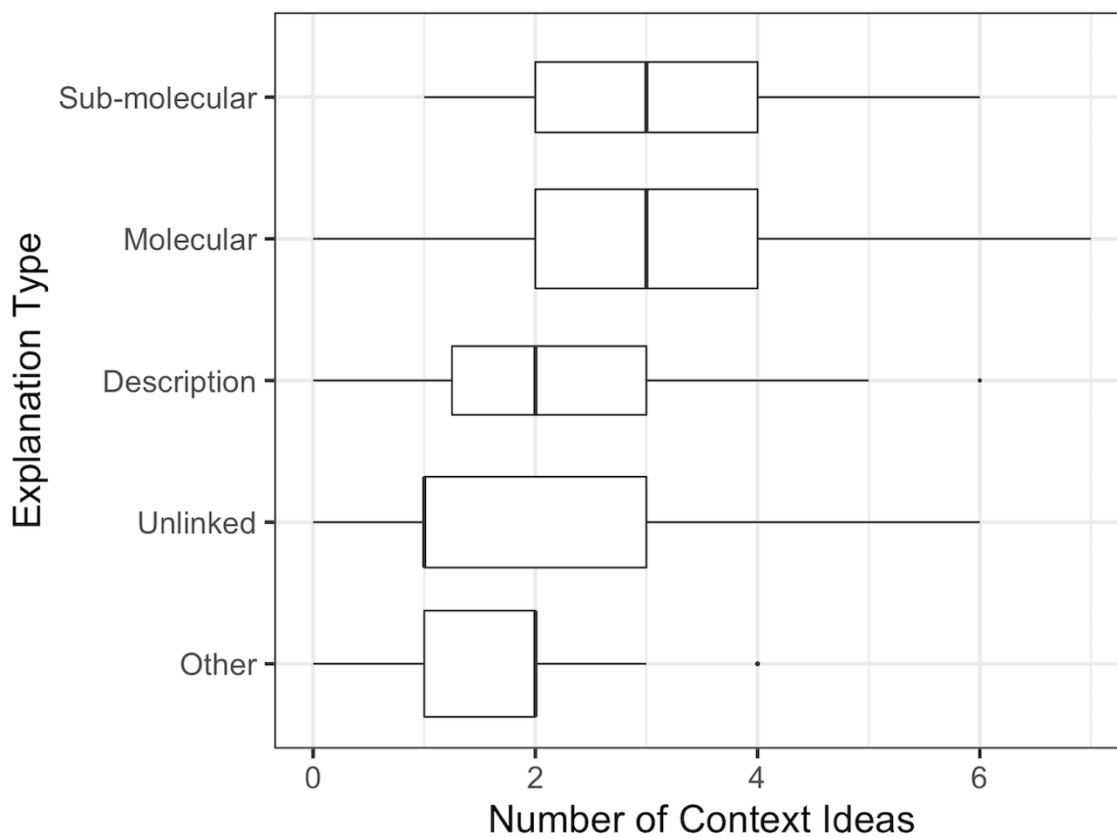


Figure 1. Students include different numbers of context ideas by response type: Other, $n = 100$; Unlinked, $n = 73$; Description, $n = 42$; Molecular, $n = 87$; Sub-molecular, $n = 44$. Vertical height of boxes is proportional to number of students in each group, box shows median and quartiles, whiskers 1.5 times the interquartile range. Outliers shown as dots.

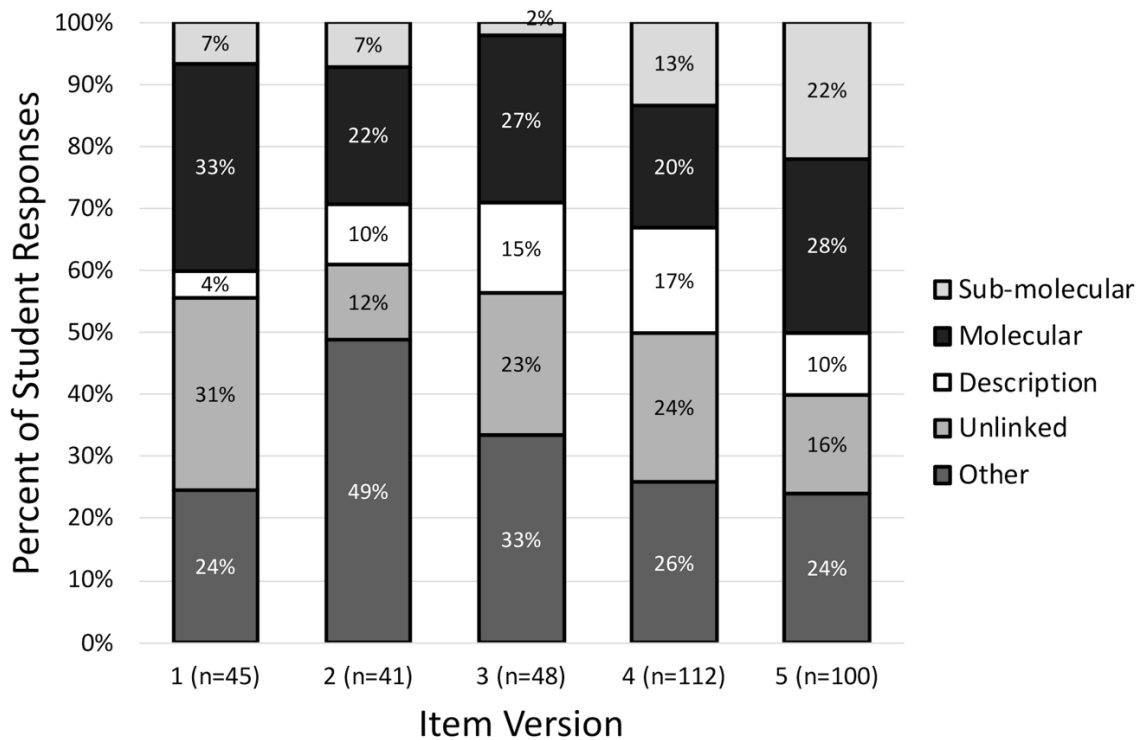


Figure 2. Students more often write Sub-molecular explanations in response to version 5 than other versions. Responses: Version 1, n = 45; Version 2, n = 41; Version 3, n = 48; Version 4, n = 112, Version 5, n = 100.

Tables

Table 1. Item versions with characteristics used to encourage students to explain. Keywords which differ between versions underlined in item text.

Version	Item	Biological Level	Comparison
1	During transcription, information contained within a gene is transferred from DNA to RNA. <u>Describe the process</u> used to transfer this information. Be sure to <u>identify and explain</u> the characteristics of DNA and RNA that allow it to hold and transfer information.	Macromolecular (DNA, RNA)	
2	During transcription, information contained within a gene is transferred from DNA to RNA. <u>Identify and explain</u> the characteristics of DNA and RNA that allow it to hold and transfer information. Then, using these characteristics, <u>describe the process</u> used to transfer this information.	Macromolecular (DNA, RNA)	
3	<u>Explain the molecular events</u> that transfer information from DNA to RNA during transcription. Be sure to discuss <u>why</u> the characteristics of DNA and RNA ensure this information is transcribed correctly.	Macromolecular (DNA, RNA)	
4	During transcription, the information from DNA is transferred to a new strand of RNA. <ol style="list-style-type: none"> <u>Describe the process</u> used to transfer this information. <u>Identify and explain</u> the characteristics of DNA and RNA that allow it to hold information. <u>Identify and explain</u> the characteristics of nucleotides that allow transfer of information. 	Macromolecular (DNA, RNA) Molecular (nucleotide)	
5	During transcription, a new strand of RNA is formed, which holds the information from the DNA. <ol style="list-style-type: none"> <u>Describe the process</u> used to transfer this information. <u>Identify and explain</u> the characteristics of DNA and RNA that allow it to hold information. <u>Identify and explain</u> the characteristics of nucleotides that allow transfer of information. 	Macromolecular (DNA, RNA) Molecular (nucleotide)	Time (a new RNA is formed)

Table 2. Student Responses Collected

<u>Number Responses by Course</u>		
<u>Version</u>	<u>Introductory Biology (Biochemistry)</u>	<u>Term</u>
1	45	Fall 2019
2	41	Fall 2019
3	48	Fall 2019
4	74 (38)	Spring 2019
5	68 (32)	Spring 2019

Table 3. Student explanation types and examples. Holistic categories cannot co-occur. All responses are spell-checked.

Holistic category (percent responses)	Analytic codes	Description	Example responses
Other (39%)	One or fewer analytic codes	No links between any analytic codes. Often describe other processes or includes errors	The RNA is transcribed from the DNA by ribosomes that read off the information necessary to create the new strand of RNA. The fact that DNA is a double stranded helix allows it to hold the bases within its structure to hold DNA. Nucleotides bind to the ribosomes and allow the ribosomes to copy the information. (introductory student 2624, version 5)
Unlinked (21%)	1 or 2 plus 3 and/or 4	Any pair or trio of analytic codes, lacking one or both of RNA polymerase and sequence transfer ideas	The DNA strand is unwound by RNA polymerase at a promoter. Then, the strand of DNA that is being used as the template is read by RNA polymerase (from 3' to 5'), which then begins to build an mRNA (5' to 3') with complementary bases (G to C, A to U (Uracil is used in RNA as opposed to Thymine in DNA)) (the bases are paired correctly through IMFs). When the RNA polymerase reaches a terminator, it is finished making the mRNA. (introductory student 4397, version 1)
Description (12%)	1 & 2	RNA polymerase synthesizes an RNA, which results in transfers of the DNA sequence to the RNA	The DNA strand is unwound and used as a template for RNA synthesis. RNA polymerase copies the DNA strand into the RNA strand. DNA is the template that holds information for coding an RNA sequence. The RNA sequence holds the information for coding proteins in terms of codons. The RNA sequence is translated into proteins via three base codons that align with an amino acid. Nucleotides are the foundation for DNA structure that holds information. The order of the nucleotides corresponds with the order of the DNA, RNA and protein sequence. (upper-level student 4254, version 4)

Table 3, continued. Student explanation types and examples. Holistic categories cannot co-occur. All responses are spell-checked.

Holistic category (percent responses)	Analytic codes	Description	Example responses
Molecular (25%)	1, 2, & 3.2	All elements of a description plus complementary base-pairing between DNA and RNA	Some characteristics of DNA that allow it to pass genetic information is base pairing. Base pairs in DNA include A to T, and C to G. These complimentary base pairs are also how the template strand of DNA is read to make RNA. Except for RNA it is not A to T, it is A to U. In transcription the RNA Polymerase starts transcribing at the promoter, then nucleotides are added with their complimentary bases to the growing RNA transcript, and then transcription ends when the RNA Polymerase comes across the terminator. At this point the RNA transcript is released and ready for translation. That is the process of how DNA's information is copied to RNA and to eventually create proteins. (introductory student 3149, version 2)

Table 3, continued. Student explanation types and examples. Holistic categories cannot co-occur. All responses are spell-checked.

Holistic category (percent responses)	Analytic codes	Description	Example responses
Sub-molecular (13%)	1, 2, 3.2, & 4.2	All elements of a molecular mechanism plus non-covalent interactions result in complementary base-pairing	During transcription, RNA polymerase reads a strand of DNA from the 3' end to the 5' and links nucleotides that are complementary to the ones in the template strand from the 5' end to the 3' end. The nucleotides of DNA and RNA are arranged in a certain order according to one's genetics, allowing them to act as a blueprint for translating into sequences of amino acids and eventually proteins. The nucleotides are complementary to each other. A is paired with T in DNA or U in RNA. C is paired with G in both. The nucleotides are purines and pyrimidines linked together by hydrogen bonds which can be broken for DNA replication or transcription. (introductory student 2172, version 5)
	or	Direct link between non-covalent interactions and sequence transfer	DNA's nucleotides attach to the nucleotides in RNA. Each type of nucleotide is able to attach to a specific nucleotide in RNA based on chemical interactions. DNA and RNA are made of nucleotides that hold an information sequence that tells the RNA which protein to make. Nucleotides are monomers that make up nucleic acid. Two types of nucleic acid are DNA and RNA. These nucleotides and interact and attach through chemical interactions that allow the RNA to read the information in the DNA (introductory student 4412, version 4)