Article

DNA → RNA: What Do Students Think the Arrow Means?

L. Kate Wright, J. Nick Fisk, and Dina L. Newman

Rochester Institute of Technology, Rochester, NY 14623

The central dogma of molecular biology, a model that has remained intact for decades, describes the transfer of genetic information from DNA to protein though an RNA intermediate. While recent work has illustrated many exceptions to the central dogma, it is still a common model used to describe and study the relationship between genes and protein products. We investigated understanding of central dogma concepts and found that students are not primed to think about information when presented with the canonical figure of the central dogma. We also uncovered conceptual errors in student interpretation of the meaning of the transcription arrow in the central dogma representation; 36% of students ($n = 128$; all undergraduate levels) described transcription as a chemical conversion of DNA into RNA or suggested that RNA existed before the process of transcription began. Interviews confirm that students with weak conceptual understanding of information flow find inappropriate meaning in the canonical representation of central dogma. Therefore, we suggest that use of this representation during instruction can be counterproductive unless educators are explicit about the underlying meaning.

INTRODUCTION

"Information flow, exchange and storage," arguably the basis for all modern genetics and genomics research, was named one of the five core concepts necessary for biological literacy by the National Science Foundation's (NSF) high-profile report, *Vision and Change: A Call to Action* (American Association for the Advancement of Science, 2009). Furthermore, this report describes the importance of integrating topics related to information flow in an in-depth way throughout the biology curriculum. The 2012–2013 Advanced Placement Biology Curriculum Framework echoes the NSF in their "Big Idea 3: Living systems store, retrieve, transmit and respond to information essential to life processes" (www.collegeboard.org). Indeed, the concept of information being stored in DNA molecules, copied into RNA intermediates, and expressed in proteins that carry out cellular functions, is known as the "central dogma" of molecular biology.
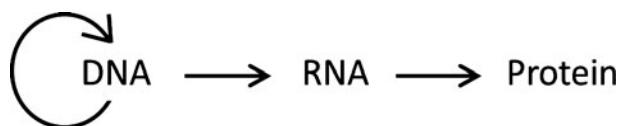
Figure 1. Typical depiction of the central dogma concept.

Francis Crick first described the central dogma as "the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred from protein to either protein or nucleic acid" (Crick, 1970). The canonical interpretation of the central dogma is that genetic information (DNA) is used to generate transient messenger molecules (RNA) that are themselves used to direct synthesis of particular protein products and that the proteins are responsible for most cellular functions.

In a recent review "revisiting" the central dogma, Shapiro (2009) wrote, "The idea of a 'dogma' in science has always struck me as inherently self-contradictory. The scientific method is based upon continual challenges to accepted ideas and the recognition that new information inevitably leads to new conceptual formulations." The discoveries of reverse transcription, catalytically active RNA molecules, and posttranscriptional modification of RNA molecules are examples of a few of the many exceptions to the central dogma. And while the traditional representation of the central dogma (Figure 1) suggests a very simplistic mechanism for information flow, in reality, the processes that direct the synthesis

of each macromolecule are numerous and complex. Most postsecondary educators would likely agree that concepts related to the central dogma are crucial to their curricula in courses such as those in introductory biology, cell and molecular biology, genetics, and developmental biology, as well as in many others. As the fields of genomics and bioinformatics continue to be revolutionized by advancing technologies such as whole-genome sequencing (Ng and Kirkness, 2010; Mavromatis *et al.*, 2012), it is critically important that we prepare our biology students to think deeply and carefully about information flow.

The molecular basis of inheritance and information exchange, though, is a difficult topic for biology students (Pashley, 1985; Stewart *et al.*, 1990; Allchin, 2000; Lewis *et al.*, 2000; Wood-Robinson *et al.*, 2000; Marbach-Ad, 2001), and many students have weak mental models of the relationships among genes, alleles, and chromosomes (Newman *et al.*, 2012). Almost every biology textbook (both secondary and postsecondary level) describes the work of Austrian monk Gregor Mendel and his famous pea experiments as an introduction to genetics. Almost none of these books, however, makes an explicit connection between the phenotype of the pea (e.g., round versus wrinkled) and the expression of a protein (e.g., a starch branching enzyme; Smith, 1988). Many students who struggle with concepts related to gene structure and expression do not appreciate the significance of the central dogma or information flow in the context of gene regulation (Khodor *et al.*, 2004), which is an essential concept in molecular biology.

Much work also demonstrates that high school curricula do not provide adequate scaffolding for students to deeply learn central dogma concepts (Lewis *et al.*, 2000; Shaw *et al.*, 2008) For example, Lewis *et al.* (2000) reported that, of 482 surveyed high school students, none knew that a final product was connected with a gene. Data obtained through the National Center for Education Statistics (http://nces.ed.gov) demonstrate that only 21% of tested 12th-grade high school students could correctly answer the multipart question "What is a gene? (genetic instructions), What is it made of? (DNA), What is the major function of a gene? (to code for protein product)." When asked to explain what a "broken gene" is, only 3% correctly described a broken gene in terms of a change in the DNA sequence and/or a change in genetic instructions. This knowledge gap may be due, in part, to the fact that few high school students know what a gene is—a concept that scientists still struggle to define in the midst of our bioinformatics and genomics explosion (Pearson, 2006; Pennisi, 2007; Pray, 2008). In fact, the definition of a gene has changed dramatically since Mendel's original conception of a unit of heredity, including the post–Watson and Crick definition of a section of DNA that encodes a protein and a recent computational characterization as a subroutine within the genomic operating system (Gerstein *et al.*, 2007). For experts, this multitude of definitions is not generally problematic, because the context is immediately recognized and applied. However, having multiple definitions that are context dependent may be very confusing to students, particularly considering that most high school and early college students are "dualists" according to Perry's stages of intellectual development (Perry, 1970). In other words, they see things in black and white and have trouble with multiple "right" answers.

There are ample data to show that a large percentage of first-year biology students enter with weak or missing mental models about genes and gene expression (e.g., O'Sullivan *et al.*, 2003; Shaw *et al.*, 2008; http://nces.ed.gov). It is little wonder, then, that learning about the complex molecular processes that maintain genomic integrity, drive gene expression, and control protein synthesis is problematic for students. As Kozma *et al.* (2000) stated, "Understanding molecular properties and processes has been a challenge, in large part because molecules and their properties are not available to direct perception." This challenge is most likely confounded by the fact that many students entering college are poor problem solvers with underdeveloped formal reasoning skills (Gilbert *et al.*, 1982; Lawson *et al.*, 2000; Wilson *et al.*, 2006). Thus, biology instruction must be designed to provide practice with formal reasoning in order to help students develop deep, transferable learning of biological concepts like gene expression. College students, of course, are not blank slates; they bring diverse experiences, background, and prior knowledge into the classroom. Issues can arise, however, when these prior conceptions are incorrect or incomplete; especially when an instructor is not aware of these shortcomings.

In *Science Teaching Reconsidered: A Handbook*, conceptual misunderstandings are described as phenomena that occur when students are not forced to confront the discrepancies between their own preconceived ideas and real-world observations. The failure to critically examine one's own mental models leads to weak, faulty models, with little confidence behind them (National Research Council [NRC], 1997). Conceptual misunderstandings are problematic, because their existence often interferes with learning new concepts (McDermott, 1991), especially if the misconceptions seem rational and useful to the learner at the time (Fay and Mayer, 1987). Several decades ago, Fisher (1985) explored what we believe is the first published account of a central dogma misconception held by college biology students: the idea that amino acids are synthesized during the process of genetic translation. This is an important conceptual misunderstanding, as students with this belief develop incorrect ideas about molecular processes associated with protein translation and biomolecule synthesis.

Such foundational weakness can be detected in student populations with careful assessment of conceptual understanding. A number of instruments, such as the Genetics Concept Assessment (Smith *et al.*, 2008), the Genetics Literacy Assessment (Bowling *et al.*, 2008), the Biological Concepts Instrument (Garvin-Doxas and Klymkowsky, 2008), the Introductory Molecular and Cell Biology Assessment (Shi *et al.*, 2010), and the Meiosis Concept Inventory (Kalas *et al.*, 2013), have been developed in an effort to help postsecondary instructors identify gaps in knowledge about essential biological processes and evaluate learning after instruction. Many of these concept assessment tools include questions that rely on deep conceptual understanding of information flow. While the published inventories cover many topics related to information flow, they are not exhaustive, and our own experiences with first- and second-year biology majors suggested there are some foundational concepts that need further exploration.

While carefully designed assessments such as concept inventories help instructors determine understanding of

concepts, they may not be as useful for assessing the differences between expert–novice reasoning. Cognitive and developmental psychologists agree that expertise in a discipline involves not only the collection of knowledge but also the ability to organize, adapt, retrieve, and apply knowledge in a dynamic way; such as when experts are presented with a novel research finding in their disciplinary field (Newell and Simon, 1972; Bédard and Chi, 1992; Chi, 2006). The difference between novice and expert reasoning has been explored most fully in physics education research. For example, Chi *et al.* (1981) elegantly described several key differences in how novices (physics undergraduate students), midlevel (physics graduate students), and experts (physics faculty) solve physics problems. Experts thought about unifying principles, such as Newton's laws, while students focused on superficial aspects of the problem, such as whether a ramp or inclined plane was in use. More recently, the novice–expert continuum was illustrated in a biology context by Smith *et al.* (2013). In this study, biology students and biology faculty were asked to sort a number of biology-based problems printed on cards. Students sorted the cards based on surface features (e.g., what type of organism was used in the problem), while faculty sorted the cards by the underlying biological concept (e.g., evolution).

Differences in how experts and novices interpret and use scientific representation has also been investigated by the discipline-based education research (DBER) community. Formulas, structural diagrams, cartoon images, and models are tools created by scientists to explain, demonstrate, or quantify their research. When these representations make their way into high school and college textbooks, slides, and work sheets, experts have little problem interpreting and using these representations. Students, though, often cannot make connections between complex processes and scientific representations and probably do not appreciate their importance or complexity. Kozma *et al.* (2000), for example, described the importance of representations such as structural diagrams and chemical equations in the daily lives of typical chemists in either an academic laboratory setting or a pharmaceutical laboratory setting. These visual representations for compounds and processes invisible to the human eye were the common language spoken by true academic and professional chemists. Observational studies of paired college chemistry students,

on the other hand, demonstrate only limited use of visualization during actual laboratory experimentation (Kozma, 2003), demonstrating that novices and experts use scientific representations differently.

In this paper, we focus on the canonical representation of central dogma, and transcription in particular, to gain insight into student conceptions of these processes. We report two novel and interesting misinterpretations of the central dogma that persist among different populations of postsecondary biology students: 1) DNA is transformed into RNA, and 2) the mRNA molecule exists before transcription takes place. We classify these findings as *misconceptions*, as they can be described as "scientifically inaccurate understanding that students have developed about natural phenomena" (Andrews *et al.*, 2012). The incorrect ideas we have uncovered would likely interfere with student learning rather than move students toward deeper understanding of information flow. We support our findings through analysis of open-ended responses, student interviews, and results of new conceptual assessments based on research findings.

METHODS

All student data presented were gathered with institutional review board approval. Students from four 4-yr institutions in the northern United States were recruited for these studies (see Table 1): a large private university (institution A: 15,000 undergraduates, 12% minority), two small private universities (institution B: 2000 students, 12% minority; institution C: 2100 total, 14.7% students of color), and a medium-sized public university (institution D: 7100 undergraduates, 12.5% minority).

Analysis of Concept Maps

Data were gathered using artifacts generated from students ($n = 86$) in a sophomore-level molecular biology course at institution A. As described by Adams and Wieman (2011), an important first step for determining areas of student confusion can be identified by observing students participating in group homework assignments, problem solving, or even help sessions. In our case, three- to four-member student

Table 1. Student populations studied[a]

Data gathered	Course	Level	Institution	Number of students
C-maps: "What is molecular biology?"	Molecular Biology	Sophomore majors	A (large, private)	86 (24 groups)
Open-ended assessment: "What is the meaning of this representation?"	Introduction to Biology I	Freshman majors	A (large, private)	53
Open-ended assessment: "What is happening at the arrow?"	Biological Systems I	Freshman majors	B (small, private)	12
	Introduction to Biology II	Sophomore majors	C (small, private)	26
	Molecular Biology	Sophomore majors	A (large, private)	101
	Upper-level elective	Junior/senior majors	D (medium-sized, public)	20
	Upper-level elective	Junior/senior majors	D (medium-sized, public)	15
Interviews: Explain the diagram, including meaning of arrows	Multiple	Multiple	A (large, private)	20

[a]Undergraduates from seven classes at four institutions were included in various parts of the project. No individuals were included in more than one group listed.

groups (n = 24 groups total) were asked to create concept maps (C-maps) in order to answer the focus question "What is molecular biology?" C-maps, initially developed to evaluate children's conceptual frameworks, are based on a model of meaningful learning in relation to science knowledge (Novak and Musonda, 1991). They can illustrate what a student knows about a particular topic and how those concepts are utilized, integrated, and linked in students' minds (Novak and Gowin, 1984; Novak, 1990; Mintzes *et al.*, 2005). Identifying and understanding differences between expert–novice thinking has been an important area of research in the field of cognitive psychology (Ericsson, 2006). We found that analysis of our student-generated C-maps gave us initial insights into their mental models of the central dogma and helped us develop focused research questions for the studies that followed. We defined a depiction of the central dogma as "correct" when students used appropriate terminology (e.g., "transcription") and linked concepts or processes together in a scientifically accurate way (e.g., "tRNAs *transport* amino acids"). We did not, however, adhere to a rigorous definition of what a C-map should look like (e.g., correct propositional structure), since most participants had little prior experience with this tool. Our goal was merely to gain insight into student thinking, not to compare this information with other groups of students or make generalizable claims.

Open-Ended Assessments

Analysis of student-generated C-maps informed us that many biology students had a poor understanding of the central dogma but did not reveal specific details of their mental models. To gain deeper understanding of how students think about the central dogma, we conducted a formative assessment with first-year biology majors (n = 53) at institution A by showing them the representation in Figure 1 with the following prompt:

> The central dogma is usually represented by the diagram above. What is the significance or meaning of this representation? Please explain.

We used inductive strategies to analyze the open-ended descriptions generated by first-year biology students (Otero and Harlow, 2009). This unconstrained coding strategy allowed us to develop a schema that was grounded in the data and to generate an emergent coding scheme. It also allowed us to derive explanations from student language rather than from the researchers' preliminary ideas. In the first rounds of coding, original student language was used to create a number of individual categories. After discussion and analysis, categories were further collapsed, and all responses were recoded using the new scheme (an example is shown in Figure 2). This procedure was repeated until both coders were satisfied all responses were accurately represented with no redundancy.

In the next round of data collection, new open-ended assessments were designed to specifically capture student thinking about the process of transcription, represented by the linear arrow between DNA and RNA. By expanding our study population, we could find out whether these incorrect ideas were unique to inexperienced first-year students or were prevalent in other populations. Thus, students from all four institutions (A–D) were recruited for this study (n = 172, ranging from first to fourth year, see Table 1). As part of precourse assessments, students were given Figure 1 with the following prompt:
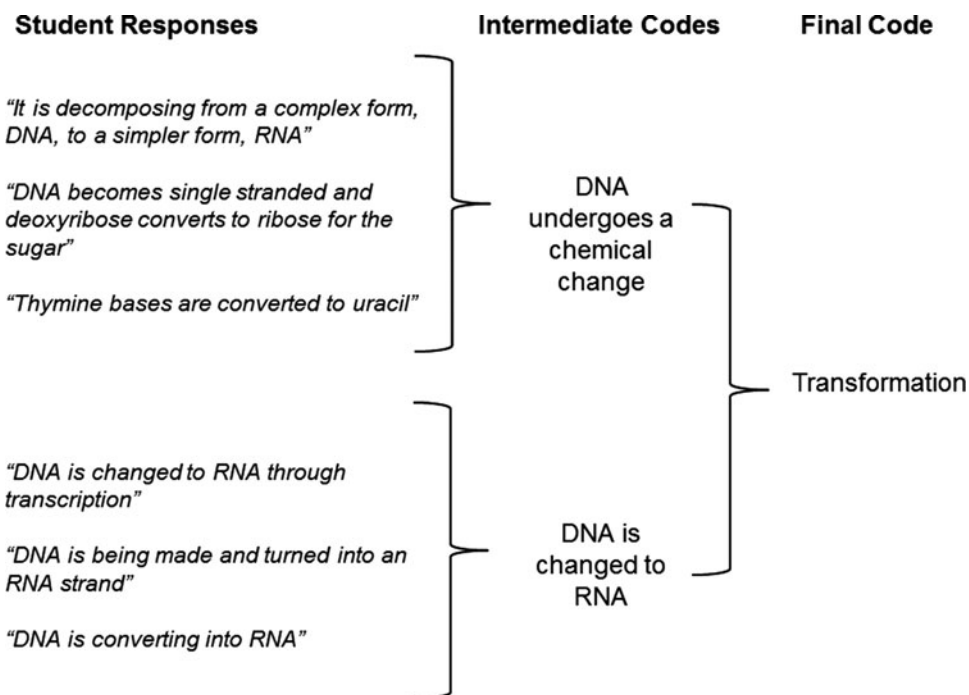

Figure 2. Example of grouping multiple student responses into a single category. Original quotes from students were initially coded into preliminary categories, and then those codes were further collapsed into final categories.

This is a representation of the central dogma. Please describe, as fully as you can, what is happening at the arrow between "DNA" and "RNA."

We were able to apply deductive strategies (Otero and Harlow, 2009) in the analysis of the 174 new responses to determine whether our previously identified themes were consistent in the new data set. Asking students to describe the molecular processes implied by the arrows in the central dogma representation allowed us to probe more deeply into student thinking and define new conceptual misunderstandings. All open-ended responses were examined through the lens of phenomenography, a qualitative research method that allowed us to investigate how students perceive a given phenomenon (Bodner and Orgill, 2007). Phenomenography is a tool particularly suited to answering educational research questions about thinking and learning (Marton, 1986) and also allowed us to describe the variation of ideas, both correct and incorrect, held by students in our test population. Phenomenography falls under the theoretical framework of hermeneutics, the science of interpretation or understanding (Weinsheimer, 1988; Henriksson, 2012), which often follows a circular approach. This qualitative method has been used by others in the DBER field to uncover student conceptions about a particular phenomenon. For example, Carlsson (2002) used a phenomenographic approach to describe the different ways teachers in training understood the biological process of photosynthesis. In the context of chemistry learning, Ebenezer and Erickson (1996) used a phenomenographic approach to describe high school students' conceptions about solubility and discovered that students inappropriately apply knowledge about how materials behave at the macroscopic level to the microscopic level.

For example, interview passages involving student descriptions of the molecular process of transcription were analyzed in the context of the entire interview and in the context of all other responses (written or verbal) describing "transcription." As noted by Dahlin (1999), the validity of a phenomenographic approach is achieved through categories that are distinct and exclusive, recognizable as realistic perceptions, and, finally, connected with previously described results. In the fourth round of open-ended responses, two researchers independently coded the data set and reached a Cohen's kappa coefficient for interrater agreement of 0.807 (Cohen, 1960; Carletta, 1996). After that, the investigators debated each mismatch until agreement was reached on the final coding.

Student Interviews

Written results were replicated during think-aloud interviews with 20 students (first-year through third-year students) from institution A. Subjects were shown the representation of the central dogma (Figure 1) and asked to explain their interpretation of the diagram as a whole, as well as any individual processes it included. During the interviews, the terms central dogma, transcription, and translation were not introduced by the interviewer. Students would only be asked to further explain the process of transcription, for example, if the term was initiated by the student, in order to find out what that term meant to the student. If students did not use technical vocabulary, the interviewer did not interject any terminology. Students were also asked to explain their understanding of

the process that connects the DNA to the RNA terms in the representation.

RESULTS

Construction of C-Maps Revealed Weak Mental Models of the Central Dogma

After student-generated C-maps answering the focus question "What is molecular biology?" were analyzed, we found that 50% of the C-maps correctly depicted the idea of information flow, with propositions such as "transcription[s'] purpose is to make RNA products," "transcription requires nucleotides," or "RNA is transcribed [*sic*] into proteins by ribosomes." The students in this group integrated processes related to DNA replication, transcription, and protein synthesis and used correct terminology and linkages, as seen in the map shown in Figure 3A. Almost half of the groups, though, made multiple mistakes, such as not connecting RNA with the process of protein translation, depicting transcription and translation as parallel instead of sequential events (maps in Figure 3, B and C), and making incorrect connections between RNA and DNA (e.g., DNA converts to RNA) or C-maps with DNA and RNA nodes never connecting. Table 2 presents the most common mistakes and their frequencies (a single map could fall into more than one category). These findings were surprising, given the C-map activity occurred on the last day of a molecular biology course. The question of why so many students failed to connect these ideas appropriately was intriguing. We were also curious to understand how deeply students understood terminology such as "transcription" and "translation": Could students use the correct words without really understanding the underlying processes?

Open-Ended Assessments Confirm Errors of Interpretation

First-year biology majors, given the canonical central dogma representation and asked to explain its meaning, provided further insight into novice mental models. After all open-ended responses were coded, 12 distinct themes of student responses emerged, three of which were considered interesting and therefore are described here. Many students used a correct technical term, such as "transcription," and then described the process incorrectly. Thus, we felt that we could not assess the meaning of the word "transcription" when it was not explained further. In later work, we were more careful to elicit student interpretation of terminology and asked students specifically *not* to use these technical terms.

Given the lack of molecular understanding by high school students described in the literature, it is not surprising that less than 10% of the students described the central dogma in the context of genetic information or information flow (see Table 3). Two additional themes suggest conceptual misunderstandings not yet described by the literature. In the first theme, students described RNA as if it were a participant in rather than the product of transcription. In the second theme, students described a transformation of one molecule to another, suggesting that DNA was converted, changed, or transformed into RNA during the transcription process. The language choices made by these students demonstrate very poor understanding of the processes involved in genetic
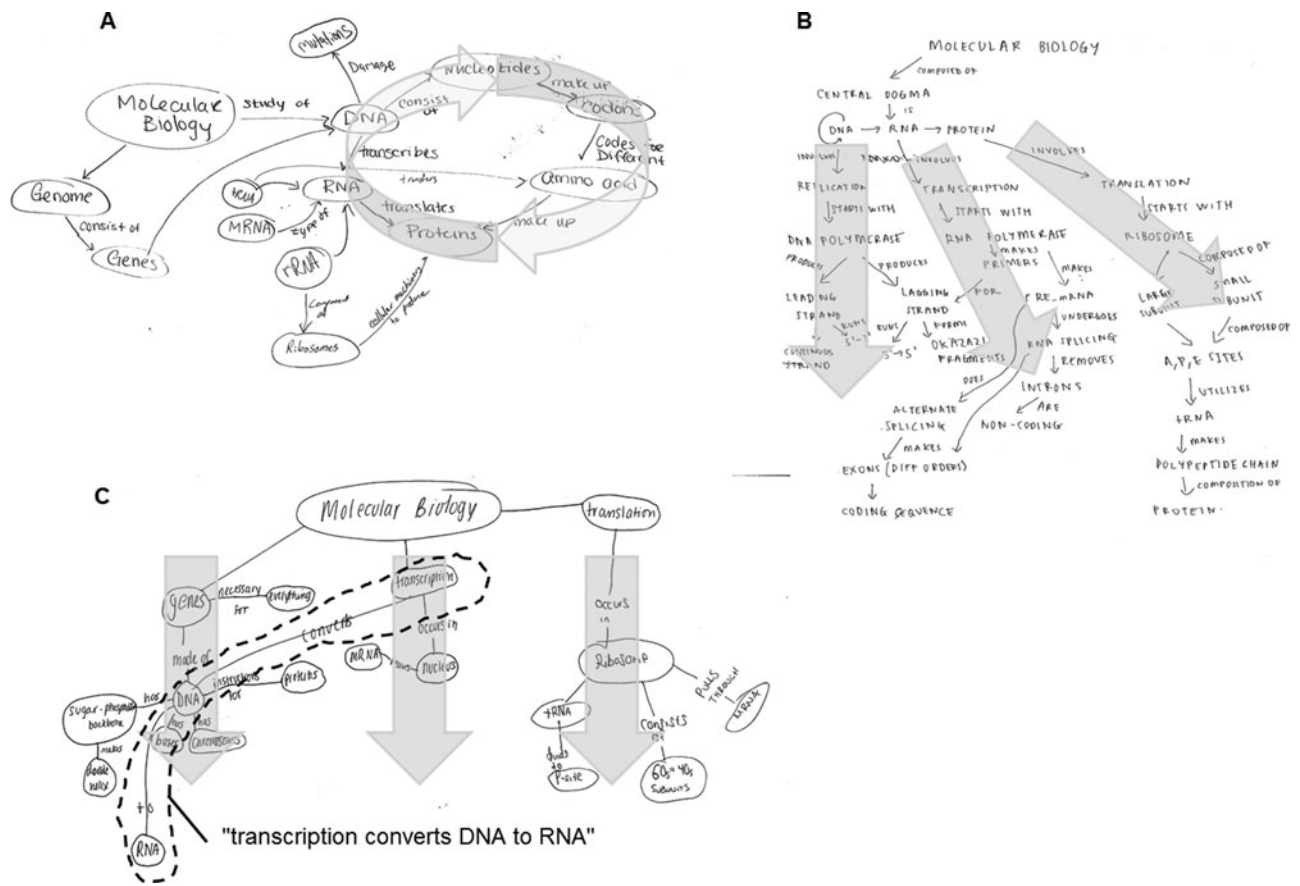
Figure 3. Representative C-maps of central dogma drawn by molecular biology students at the end of the course. (A) Correct linking of transcription and translation. (B and C) Transcription and translation drawn as parallel, disconnected processes; (C) also shows the misconception that transcription involves conversion of DNA to RNA. (Thick gray arrows superimposed on original C-maps for emphasis.)

information flow. Our results also suggest that the canonical representation of the central dogma adds confusion, not clarity, in understanding these processes (see excerpts below).

The Transcription Arrow Is a Persistent Source of Confusion

Because much of the student confusion seemed to center around the process of transcription, which is represented by a single arrow in the central dogma representation (Figure 1),

Table 2. Themes of incorrect C-maps[a]

Category	Percent of total C-maps
RNA is not connected to the process of translation	29.2
Transcription and translation are parallel (not sequential)	21
Incorrect or missing connection between RNA and DNA	29.2
Unclear C-map	8.3

[a]$n = 24$; 12 maps were coded as incorrect. A single map could fall into more than one category.

we refined our research question to "What do students think the transcription arrow represents?" and expanded our study population to more advanced students at several institutions (Table 1). We assumed that an expert would view the arrow between DNA and RNA in the diagram as a representation for the process of transcription—the synthesis of a new RNA (or mRNA) molecule that is complementary to one strand of DNA. Unfortunately, about one-third of students responded with only a single term (usually "transcription" but occasionally an incorrect term such as "translation"), even though they were explicitly asked to explain any technical terms used. In those cases, it was not possible to determine what that term meant to the student. Of the remaining 128 responses, only 30% percent of students stated that a new molecule of RNA was being created. A few students crafted in-depth explanations that demonstrated a clear mental model of transcription, but others used such ambiguous language that it was difficult to interpret their meaning. In those cases, a student may have understood the underlying concept or may just have known the jargon associated with this context.

As shown in Table 4, 20% of students indicated that the arrow was representative of a transformation from one molecule to another; just as we observed in the first-year population, in all the classes that were queried, at all institutions and levels. The other major category from the previous work,

Table 3. Freshman responses to open-ended assessment question[a]

Category	Sample quotes	Frequency
Flow of genetic information	This represents the path of the info stored in molecules like DNA . . . The info in DNA can be passed on to RNA molecules, which in turn actually execute the code for proteins.	9.4% (correct)
Transformation of one molecule to another	DNA makes DNA which becomes RNA which becomes protein	20.8% (incorrect)
	DNA is in the nucleus, it gets converted into RNA, base pairs are changed because the protein that makes amino acid can only use RNA.	
	This means that DNA replication happens and the DNA is then turned into RNA and the RNA is then used in proteins.	
RNA exists before the process and/or drives it	RNA can read the DNA and code for amino acids	20.8% (incorrect)
	RNA is a messenger that copies parts of DNA and codes for proteins.	

[a]During the analysis of responses to a query about the meaning of the central dogma, three important categories were noted (one correct and two incorrect). Although various other responses were given, these three types were striking for their frequency.

RNA exists before transcription and/or drives it, was also present in the new data set. We found that 16% of students described a scenario in which RNA existed before the process of transcription began and thus was able to be actively involved in the molecular process. These types of statements were given by students in all classes except the upper-level electives, where numbers were small ($n = 20$ and 15).

Interviews Confirm Existence of Student Difficulties with the Central Dogma

Results from the open-ended responses were confirmed during think-aloud interviews as subjects were asked to explain their understanding of the process(es) that connect the DNA to the RNA and RNA to the protein terms in the central dogma representation (Figure 1). Examples of students describing the "transformation" model are given below.

Sarah: The result is an RNA strand.
Interviewer: What happens to the DNA strand?
Sarah: The DNA strand is split into RNA
Interviewer: Can you try to explain the entire figure [Figure 1] from start to finish without using the words "transcribed" or "translated"?
Steve: Ok. Um. DNA has to be taken from its double-stranded form to—and converted to RNA, which is converted to proteins, which encode for different functions in the body.

Interviewer: Describe to me what this figure means to you . . . if you've seen it before, what do you think the big picture is here?
Cory: I mean it seems like since you made the arrow (points to the first circular arrow) like that it feels like DNA is a replicating process. Um—and then I'm guessing it's more of a— to me it's a reverse process because protein is made up of, I mean RNA is made up of protein and then RNA is what makes up DNA. I think.

Cory repeated his idea that "RNA is made up of protein" and "RNA is what makes up DNA" several more times during the interview. To clarify his thinking about the arrows in the figure, the interviewer sketched a generic diagram of a metabolic pathway "A→B→C" and asked Cory whether he thought it was a comparable diagram. He agreed this new diagram was the same as Figure 1, suggesting that he interpreted the central dogma representation as a series of chemical reactions.

The model of RNA existing before transcription was espoused by another subject in an interview.

Libby: Yeah, oh yes, I have [seen a diagram like Figure 1]. It's transcription or something with a "trans" in the beginning. There's, like, three of them, and I think it's transcription where you have the DNA and then the RNA comes and it transcribes so the RNA knows what protein to make, from the DNA.
[. . .]

Table 4. Student explanations of the central dogma from students at four different institutions, across all levels of undergraduate study[a]

Category	Example	Frequency
DNA is copied or carries information	The DNA is providing a template for the RNA to form off of.	24% (correct)
A new molecule of RNA is being created	Transcription. DNA is used as a template to create an RNA molecule, which is single-stranded, to be used in translation	30% (correct)
Transformation of one molecule to another	Transcription: the process of turning a section of the double-stranded DNA into a piece of single-stranded RNA using certain proteins.	20% (incorrect)
RNA exists before the process and/or drives it	transcription occurs at #2 as the DNA transfers information to mRNA	16% (incorrect)
DNA is a "template" (used the term only)	DNA acts as a template for RNA.	6% (ambiguous)
DNA is "transcribed to RNA."	DNA is transcribed to RNA.	9% (ambiguous)
DNA is "transcribed into RNA."	DNA is transcribed into RNA.	23% (ambiguous)

[a]$n = 128$. Some responses fell into more than one category. Some responses were considered ambiguous because the student did not explain the meaning of a key technical term.

Interviewer: Okay, I labeled the arrows 1, 2, and 3. If we just focus on arrow 1, can you tell me what you think is being represented by that arrow or what is going on in that process?

Libby: Okay. I think it means, like, well, DNA in the beginning has to unzip because otherwise RNA can't get in. So I think I'm guessing that what's [*sic*] the little circle arrow thing means because it has to unzip so the RNA can get in there and then attach to one side of the DNA and then transcribe from that.

Interviewer: Okay. What about arrow 2?

Libby: I'm just assuming that arrow 2 is DNA allows RNA to get in, like, that's the process that it has to go through. DNA has to unzip first and then the RNA comes in and then the protein comes in so, it's just, like, step by step by step.

[….]

Libby: I think arrow 3 represents the whole process of the RNA, like, transcribing all the amino acids together and making it into a protein so, like, the arrow 3 just represents that process. RNA makes the amino acids and then it makes the protein.

Libby consistently described RNA as an entity that already existed in the system and an active participant in various processes through her language usage such as, "RNA gets in there" or "RNA comes in" or "RNA makes …" throughout the interview. At another point, she was asked to explain the origin of the RNA in her model. The student responded with surprise to this question.

Interviewer: And how does, um, where does that RNA strand come from? You said the DNA unzips and RNA strand can match up with it. Where does that RNA strand come from?

Libby: That's a very good question! Um, I'm assuming it just comes from somewhere in the cell. Maybe leftover parts? I—I have actually never thought about that.

Analysis of the responses in this data set revealed that using correct terminology was not necessarily linked with correct understanding of the molecular process of transcription. Forty-four percent of students (20 out of 45) who answered the open-ended survey question with one of the two conceptual misunderstandings also used the term "transcribe" or "transcription" in their explanations. Although the technical language was correct, the underlying mental models of the molecular process were not. We highlight this point, as 23% of students who responded to the open-ended question described the transcription arrow by writing the phrase "DNA is transcribed *into* RNA," while 9% of students answered the question by writing "DNA is transcribed *to* RNA." Although neither one of these statements is technically incorrect, we cannot know for certain what these student answers mean. When a student writes the phrase, "DNA is transcribed into RNA," does he or she mean that a new molecule of RNA is synthesized using a DNA template or does he or she mean that DNA is converted into RNA? This cannot be determined, because correct terminology does not necessarily correlate with understanding.

DISCUSSION

Through this work, we have uncovered novel misconceptions concerning how students think about information flow as it pertains to the central dogma of molecular biology.

We feel that it is important to disseminate these findings to instructors and textbook authors, because they must be aware of these ideas in order to help students confront them. Research in cognitive science demonstrates that deep learning requires dynamic reorganization of mental structures; misconceptions must be replaced with true conceptions (Lawson *et al.*, 2000; O'Donnell *et al.*, 2012; Jensen *et al.*, 2013). Instructors cannot simply add more information to their students' minds, especially if they are not aware of what their students are thinking about. Redish and Steinberg (1999) remind us that effective instruction happens only when educators gain insight into student thinking, insight we attempt to provide in this work. Our analysis revealed several interesting findings: 1) Students are not primed to think about "information" when presented with the canonical figure of the central dogma, and 2) misconceptions about the central dogma were exposed when students had to describe the meaning of an arrow in the central dogma representation.

Our work demonstrates that the arrow representing transcription in the typical central dogma figure seems to miscue many biology students: in cognitive terms, the arrow does not prime them to think about molecular mechanisms or even information flow. When asked to describe the significance of the arrow, some biology students think of the arrow as a chemical conversion similar to the arrow they might see when studying a metabolic pathway such as glycolysis (Figure 4A). This is the wrong cognitive resource, which could interfere with meaningful learning. In other words, students who inappropriately apply the representation of a chemical reaction to the central dogma might conclude that DNA gets transformed through chemical processes into RNA during transcription. Because we found evidence of this misconception in all levels of students (first year to advanced), we conclude that this incorrect idea is robust and not just something that inexperienced students say.

Other students seem to interpret transcription as a transfer of material between two existing entities, DNA and RNA—an incorrect model of information movement analogous to the movement of electrons by respiratory chain proteins during aerobic respiration (Figure 4B). One of the major flaws of this model is that RNA must already exist before the process of transcription begins so that "information" or "material" can be transferred to and from DNA and RNA. These students completely miss the idea of new RNA synthesis and seem to be thinking about information as an entity or particle, similar to what Lewis and Kattmann (2004) observed among high school students trying to explain the nature of genes.

In the case of the central dogma, we suggest that the arrow representation is too vague and can mean too many things to students. We found that students who do not have a solid understanding of the molecular mechanisms used during information flow see the arrow mainly as a sign for a chemical conversion or as a signal for a physical transfer of something. Sometimes students in our study population misinterpreted the symbolism of the arrow itself rather than the mechanism of transcription. For example:
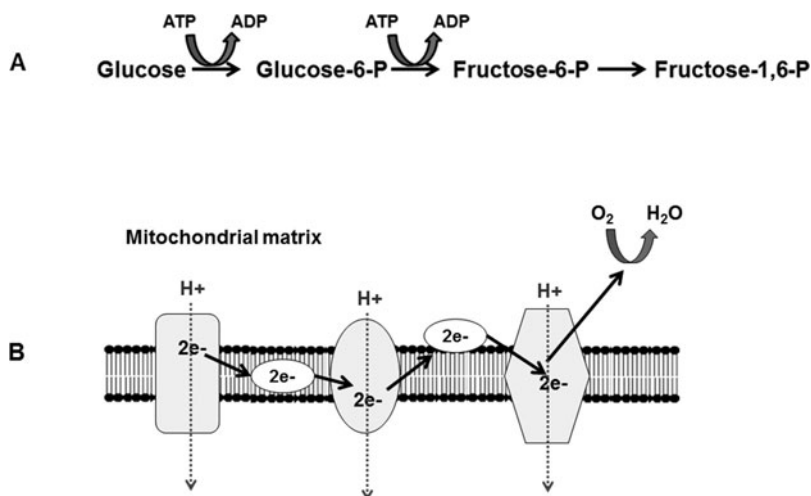
DNA acts a template as production of RNA. The *arrows* show how the genes are actually used.

DNA is a template for RNA so the *arrow* between then is representative of base-pairing with thymine being substituted with Uracil to create RNA

The Genes from DNA is what defines the structure of RNA. What the *arrow* could mean is a specific strand of DNA that is used in RNA.

Arrows are not found only in biology representations, they are also used extensively in the field of organic chemistry to demonstrate a variety of concepts. Unlike traditional college biology courses, though, students in an organic chemistry course are given explicit directions on how to use and interpret arrow symbols that will be encountered during the course. A typical organic chemistry textbook explains that chemical diagrams are good for showing the positions of atoms but not electrons, because electrons are not solid particles with a singular position in place and time (Klein, 2011). Students are then presented with explanations and diagrams of a correctly drawn curved arrow (Figure 5A) that is used in the field of organic chemistry to represent the movement of electrons in different contexts. Later, students are introduced to full-headed arrows that represent actual mechanisms (Figure 5B), while partial-headed or "fishhook" arrows (Figure 5C) indicate the shift of a single electron. The double-arrow structure in Figure 5D is the symbol for equilibrium. Organic chemists are able to convey scientific concepts to one another and to their students using arrow symbolism, because there is an accepted and universal language that everyone in the field follows. Biologists, on the other hand, do not seem to subscribe to a given set of scientific representations—likely confusing for the nonexpert!
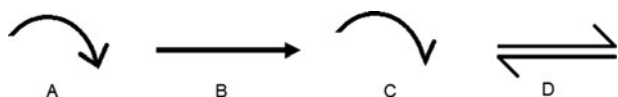

Figure 5. Different arrow representations used in organic chemistry. Arrows are used to represent (A) movement of electrons, (B) mechanisms, (C) single electron shift, and (D) equilibrium.

While representations are crucial for communicating science, language is just as important. Students will, unfortunately, encounter imprecise language in textbooks, tutorials, and classrooms that may confirm, rather than refute, their misconceptions. For example, in an online tutorial written about the central dogma (https://cm.jefferson.edu/learn/dna_and_rna.html), the author writes:

> The central dogma of molecular biology is "DNA makes RNA makes protein." This general rule emphasizes the order of events from transcription through translation and provides the basis for much of the genetic code research in the post double helix 1950s. The central dogma is often expressed as the following: "DNA makes RNA, RNA makes proteins, proteins make us."

We suggest the author of this tutorial really believes that "DNA makes RNA," but the choice of language is potentially confusing and may also solidify, not refute, certain weak mental models of the central dogma. An expert reading the passage would be able to look beyond the literal meaning of the word "makes" to envision the complex processes that underlie the synthesis of RNA, but a novice might interpret the word "makes" in a very literal sense and infer that DNA is the driver, not the template, in the process. Therefore, it is crucial that we, as educators, are careful about our language with students and demand that students be equally precise. Without such rigor, we cannot evaluate their mental models or prevent the creation of new misconceptions.

As Wiggins and McTighe (2005) pointed out, "knowing the facts and doing well on tests of knowledge do not mean that we understand." Many biology education research (BER) practitioners have echoed this observation in their own work and have made a concerted effort to avoid questions and responses based solely on student recognition of correct vocabulary terms in the development of biology concept assessment tools (Bowling *et al.*, 2008; Smith *et al.*, 2008; Klymkowsky *et al.*, 2010; Shi *et al.*, 2010). Our findings from open-ended responses and interviews revealed that students can recognize and use correct terminology while still employing faulty mental models of information flow. This suggests that classroom assessment strategies must probe students to do more than come up with correct vocabulary terms.

Students could be asked, for example, to explain information flow or create C-maps in the context of the central dogma *without* using technical terms such as *transcription/transcribe* or *translation/translate*. We have begun to use this strategy with our own classes, and preliminary results suggest that taking away the scientific jargon is a good way to elicit true mental models of students and begin discussion of scientific language and underlying concepts. Instructors might also work with their students to trace matter (including nucleotides, polymerases, etc.) through each major step of the central dogma, similar to what Clark *et al.* (2009) demonstrated with geoscience students trying to understand movement and transformation of water in dynamic Earth systems.

The student-generated C-maps (Figure 3), which were the catalyst for this study, actually revealed very little of what students know about information flow. What did the students mean, for example, when they linked nodes containing the terms "DNA" and "RNA" with the linking word "transcription"? Without further probing, it is not possible to know. Thus, future work will incorporate the findings from this study into a new concept assessment that specifically focuses on information flow. As part of that project, we will also investigate student understanding of the process of protein translation.

The development and dissemination of more concept assessment instruments focusing on central dogma topics may help the BER community identify learning progressions for a number of related topics. The NRC (2007) describes learning progressions as descriptions of increasingly sophisticated ways to think about or understand a topic. The coordination of assessments, classroom activities, and instructor preparation on any number of topics could potentially be guided by these learning progressions (Alonzo and Gotwals, 2012). For example, The Environmental Literacy Project at Michigan State University (http://edr1.educ.msu.edu/environmentallit) has made great strides in the development and dissemination of learning progressions in the context of environmental science. Because human endeavors greatly impact natural and environmental systems of our planet, there is a dire need to educate all citizens, so they may take part in decisions about environmental issues and policies (Gunckel *et al.*, 2012). The BER community might look to the Environmental Literacy Project for guidance in the development of learning progressions for understanding concepts related to the central dogma. As advances in biomedical knowledge and technology continue to outpace our current education system, and direct-to-consumer genetic testing companies make genetic analysis affordable and accessible, it is essential for all citizens to be genetically literate and make informed decisions concerning healthcare and privacy issues. The growing body of literature on student understanding of information flow suggests the BER community could be poised to undertake such an endeavor.

ACKNOWLEDGMENTS

REFERENCES

Adams WK, Wieman CE (2011). Development and validation of instruments to measure learning of expert-like thinking. Int J Sci Educ *33*, 1289–1312.

Allchin D (2000). Mending Mendelism. Am Biol Teach *62*, 632–639.

Alonzo AC, Gotwals AW (2012). Learning Progressions in Science: Current Challenges and Future Directions, Rotterdam, The Netherlands: Sense Publishers.

American Association for the Advancement of Science (2009). Vision and Change in Undergraduate Biology Education: A Call to Action, Washington, DC.

Andrews TM, Price RM, Mead LS, McElhinny TL, Thanukos A, Perez KE, Herreid CF, Terry DR, Lemons PP (2012). Biology undergraduates' misconceptions about genetic drift. CBE Life Sci Educ *11*, 248–259.

Bédard J, Chi MTH (1992). Expertise. Curr Dir Psychol Sci *1*, 135–139.

Bodner GM, Orgill M (2007). Theoretical Frameworks for Research in Chemistry/Science Education, Upper Saddle River, NJ: Prentice Hall.

Bowling BV, Acra EE, Wang L, Myers MF, Dean GE, Markle GC, Moskalik CL, Huether CA (2008). Development and evaluation of a genetics literacy assessment instrument for undergraduates. Genetics *178*, 15–22.

Carletta J (1996). Assessing agreement on classification tasks: the kappa statistic. Comput Linguist *22*, 249–254.

Carlsson B (2002). Ecological understanding 1: ways of experiencing photosynthesis. Int J Sci Educ *24*, 681–699.

Chi MTH (2006). Two approaches to the study of experts' characteristics. In: The Cambridge Handbook of Expertise and Expert Performance, ed. KA Ericsson, New York: Cambridge University Press, 21–30.

Chi MTH, Feltovich PJ, Glaser R (1981). Categorization and representation of physics problems by experts and novices. Cogn Sci *5*, 121–152.

Clark SK, Sibley DF, Libarkin JC, Heidemann M (2009). A novel approach to teaching and understanding transformations of matter in dynamic earth systems. J Geosci Educ *57*, 233–241.

Cohen J (1960). A coefficient of agreement for nominal scales. Educ Psychol Meas *20*, 37–46.

Crick F (1970). central dogma of molecular biology. Nature *227*, 561–563.

Dahlin B (1999). Ways of coming to understand: metacognitive awareness among first-year university students. Scand J Educ Res *43*, 191–207.

Ebenezer JV, Erickson GL (1996). Chemistry students' conceptions of solubility: a phenomenography. Sci Educ *80*, 181–201.

Ericsson KA (2006). The Cambridge Handbook of Expertise and Expert Performance, New York: Cambridge University Press.

Fay AL, Mayer RE (1987). Children's naive conceptions and confusions about logo graphics commands. J Educ Psychol *79*, 254–268.

Fisher K (1985). A misconception in biology: amino acids and translation. J Res Sci Teach *22*, 53–62.

Garvin-Doxas K, Klymkowsky MW (2008). Understanding randomness and its impact on student learning: lessons learned from building the Biology Concept Inventory (BCI). CBE Life Sci Educ *7*, 227–233.

Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, Emanuelsson O, Zhang ZD, Weissman S, Snyder M (2007). What is a gene, post-ENCODE? History and updated definition. Genome Res *17*, 669–681.

Gilbert JK, Osborne RJ, Fensham PJ (1982). Children's science and its consequences for teaching. Sci Educ *66*, 623–633.

Gunckel KL, Mohan L, Covitt BA, Anderson CW (2012). Addressing challenges in developing learning progressions for environmental science literacy. In: Learning Progressions in Science: Current Challenges and Future Directions, ed. A Alonzo and AW Gotwals, Sense Publishers, Rotterdam, The Netherlands, 39–75.

Henriksson C (2012). Hermeneutic phenomenology and pedagogical practice. In: Hermeneutic Phenomenology in Education: Method and Practice, ed. N Friesen, C Henriksson, and T Saevi, Sense Publishers, Rotterdam, The Netherlands, 119–140.

Jensen JL, Kummer TA, Banjoko A (2013). Research and teaching: assessing the effects of prior conceptions on learning gene expression. J Coll Sci Teach *42*, 82–91.

Kalas P, O'Neill A, Pollack C, Birol G (2013). Development of a meiosis concept inventory. CBE Life Sci Educ *12*, 655–664.

Khodor J, Halme DG, Walker GC (2004). A hierarchical biology concept framework: a tool for course design. Cell Biol Educ *3*, 111–121.

Klein D (2011). Organic Chemistry as a Second Language, First Semester Topics, Hoboken, NJ: Wiley.

Klymkowsky MW, Underwood SM, Garvin-Doxas K (2010). Biological Concepts Instrument (BCI): a diagnostic tool for revealing student thinking. arXiv:1012.4501v1.

Kozma R (2003). The material features of multiple representations and their cognitive and social affordances for science understanding. Learn Instr *13*, 205–226.

Kozma R, Chin E, Russell J, Marx N (2000). The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. J Learn Sci *9*, 105–143.

Lawson AE, Clark B, Cramer-Meldrum E, Falconer KA, Sequist JM, Kwon Y-J (2000). Development of scientific reasoning in college biology: do two levels of general hypothesis-testing skills exist? J Res Sci Teach *37*, 81–101.

Lewis J, Kattmann U (2004). Traits, genes, particles and information: re-visiting students' understandings of genetics. Int J Sci Educ *26*, 195–206.

Lewis J, Leach J, Wood-Robinson C (2000). All in the genes?—young people's understanding of the nature of genes. J Biol Educ *34*, 74–79.

Marbach-Ad G (2001). Attempting to break the code in student comprehension of genetic concepts. J Biol Educ *35*, 183–189.

Marton F (1986). Phenomenography—a research approach to investigating different understandings of reality. J Thought *21*, 28–49.

Mavromatis K *et al.* (2012). The fast changing landscape of sequencing technologies and their impact on microbial genome assemblies and annotation. PLoS One *7*, e48837.

McDermott LC (1991). Millikan Lecture 1990: What we teach and what is learned—closing the gap. Am J Phys *59*, 301.

Mintzes JJ, Wandersee JH, Novak JD (2005). Teaching Science for Understanding: A Human Constructivist View, Burlington, MA: Elsevier Academic.

National Research Council (NRC) (1997). Science Teaching Reconsidered: A Handbook, Washington, DC: National Academies Press.

NRC (2007). Taking Science to School: Learning and Teaching Science in Grades K–8, Washington, DC: National Academies Press.

Newell A, Simon HA (1972). Human problem solving, Englewood Cliffs, NJ: Prentice-Hall.

Newman DL, Catavero C, Wright LK (2012). Students fail to transfer knowledge of chromosome structure to topics pertaining to cell division. CBE Life Sci Educ *11*, 425–436.

Ng PC, Kirkness EF (2010). Whole genome sequencing. Methods Mol Biol *628*, 215–226.

Novak JD (1990). Concept maps and Vee diagrams: two metacognitive tools to facilitate meaningful learning. Instr Sci *19*, 29–52.

Novak JD, Gowin DB (1984). Learning How to Learn, Cambridge, UK: Cambridge University Press.

Novak JD, Musonda D (1991). A twelve-year longitudinal study of science concept learning. Am Educ Res J *28*, 117–153.

O'Donnell AM, Hmelo-Silver CE, Erkens G (2012). Collaborative Learning, Reasoning, and Technology, New York: Routledge.

O'Sullivan CY, Grigg WS, Qian J, Zhang J, Lauko MA (2003). The Nation's Report Card: Science 2000, Washington, DC: National Center for Education Statistics.

Otero VK, Harlow DB (2009). Getting started in qualitative physics education research. In: Getting Started in PER, College Park, MD: American Association of Physics Teachers, 1–66.

Pashley M (1985). A level students: their problems with genes and alleles. J Biol Educ *28*, 120–127.

Pearson H (2006). What is a gene? Nature *441*, 398–401.

Pennisi E (2007). DNA study forces rethink of what it means to be a gene. Science *316*, 1556–1557.

Perry WG, Jr. (1970). Forms of Intellectual and Ethical Development in the College Years: A Scheme, New York: Holt, Rinehart, and Winston.

Pray L (2008). What is a gene? Colinearity and transcription units. Nat Educ *1*, 97.

Redish EF, Steinberg RN (1999). Teaching physics: figuring out what works. Phys Today *52*, 24–30.

Shapiro JA (2009). Revisiting the central dogma in the 21st century. Ann New York Acad Sci *1178*, 6–28.

Shaw KRM, Van Horne K, Zhang H, Boughman J (2008). Essay contest reveals misconceptions of high school students in genetics content. Genetics *178*, 1157–1168.

Shi J, Wood WB, Martin JM, Guild NA, Vicens Q, Knight JK (2010). A diagnostic assessment for introductory molecular and cell biology. CBE Life Sci Educ *9*, 453–461.

Smith AM (1988). Major differences in isoforms of starch-branching enzyme between developing embryos of round- and wrinkled-seeded peas (Pisum sativum L.). Planta *175*, 270–279.

Smith JI *et al.* (2013). Development of the biology card sorting task to measure conceptual expertise in biology. CBE Life Sci Educ *12*, 628–644.

Smith MK, Wood WB, Knight JK (2008). The Genetics Concept Assessment: a new concept inventory for gauging student understanding of genetics. CBE Life Sci Educ *7*, 422–430.

Stewart J, Hafner B, Dale M (1990). Students' alternate views of meiosis. Am Biol Teach *52*, 228–232.

Weinsheimer PJ (1988). Gadamer's Hermeneutics: A Reading of Truth and Method, New Haven, CT: Yale University Press.

Wiggins GP, McTighe JA (2005). Understanding by Design, Alexandria, VA: Association for Supervision & Curriculum Development, 39.

Wilson CD, Anderson CW, Heidemann M, Merrill JE, Merritt BW, Richmond G, Sibley DF, Parker JM (2006). Assessing students' ability to trace matter in dynamic systems in cell biology. Cell Biol Educ *5*, 323–331.

Wood-Robinson C, Lewis J, Leach J (2000). Young people's understanding of the nature of genetic information in the cells of an organism. J Biol Educ *35*, 29–36.