Adv Physiol Educ 47: 282–295, 2023.
First published February 2, 2023; doi:10.1152/advan.00155.2022

american physiological society | ADVANCES IN PHYSIOLOGY EDUCATION

EDUCATION RESEARCH

Physiology Core Concepts

Oaks to arteries: the Physiology Core Concept of flow down gradients supports transfer of student reasoning

Jennifer H. Doherty,[1,2]* Jack A. Cerchiara,[3]* Emily E. Scott,[3] Lauren N. Jescovitch,[4] Jenny L. McFarland,[5] Kevin C. Haudek,[4,6] and Mary Pat Wenderoth[3]

[1]*Department of Physiology, Michigan State University, East Lansing, Michigan, United States;* [2]*Lyman Briggs College, Michigan State University, East Lansing, Michigan, United States;* [3]*Department of Biology, University of Washington, Seattle, Washington, United States;* [4]*CREATE for STEM Institute, Michigan State University, East Lansing, Michigan, United States;* [5]*Edmonds College, Lynnwood, Washington, United States; and* [6]*Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan, United States*

Abstract

The Physiology Core Concept of flow down gradients is a major concept in physiology, as pressure gradients are the key driving force for the bulk flow of fluids in biology. However, students struggle to understand that this principle is foundational to the mechanisms governing bulk flow across diverse physiological systems (e.g., blood flow, phloem sap flow). Our objective was to investigate whether bulk flow items that differ in scenario context (i.e., taxa, amount of scientific terminology, living or nonliving system) or in which aspect of the pressure gradient is kept constant (i.e., starting pressure or pressure gradient) influence undergraduate students' reasoning. Item scenario context did not impact the type of reasoning students used. However, students were more likely to use the Physiology Core Concept of "flow down [pressure] gradients" when the pressure gradient was kept constant and less likely to use this concept when the starting pressure was kept constant. We also investigated whether item scenario context or which aspect of the pressure gradient is kept constant impacted how consistent students were in the type of reasoning they used across two bulk flow items on the same homework. Most students were consistent across item scenario contexts (76%) and aspects of the pressure gradient kept constant (70%). Students who reasoned using "flow down gradients" on the first item were the most consistent (86, 89%), whereas students using "pressures indicate (but don't cause) flow" were the least consistent (43, 34%). Students who are less consistent know that pressure is somehow involved or indicates fluid flow but do not have a firm grasp of the concept of a pressure gradient as the driving force for fluid flow. These findings are the first empirical evidence to support the claim that using Physiology Core Concept reasoning supports transfer of knowledge across different physiological systems.

NEW & NOTEWORTHY These findings are the first empirical evidence to support the claim that using Physiology Core Concept reasoning supports transfer of knowledge across different physiological systems.

bulk flow; cardiovascular; flow down gradients; Physiology Core Concepts; transfer

INTRODUCTION

The National Research Council has identified "twenty-first century skills" that are necessary for current and future individuals to successfully contribute to the global economy (1). In response, educational goals have reemphasized the need to shift away from memorization of content to fostering the higher-cognitive level skill of critical thinking (2). Scientific reasoning is a domain-specific form of critical thinking. To develop scientific reasoning, students must have a deep understanding of the principles and core concepts that are foundational to a given discipline (3, 4). It has been suggested, across many disciplines, that being able to reason with scientific principles can be the basis for individuals to successfully transfer their understanding to a novel situation (5–9). The ability to transfer scientific principles across contexts is an important index of adaptive, flexible learning that enables an individual to apply what they have learned in one setting to a new problem or setting (10).

Physiology is a discipline with many diverse phenomena that can be challenging for students to understand (11, 12). Modell (5) identified a set of conceptual models that he

*J. H. Doherty and J. A. Cerchiara contributed equally to this work.
Correspondence: J. H. Doherty (dohert59@msu.edu).
Submitted 3 August 2022 / Revised 27 January 2023 / Accepted 27 January 2023

proposed could help students better understand physiology, as students could use these models to reason mechanistically across seemingly unrelated physiological systems. These conceptual models, many of which are Physiology Core Concepts, may help students transfer their understanding across physiological systems that appear quite different at an anatomical level (5, 9). One conceptual model is the principle of "mass and heat flow," which is the Physiology Core Concept of "flow down gradients" (13, 14). Flow down gradients states that the rate of movement of a substance is directly proportional to the magnitude of the driving force, or gradient, and inversely proportional to the magnitude of the resistance that impedes movement. Flow down gradients governs many different types of flow, such as diffusion of glucose or ions across membranes, osmosis of water between anatomical compartments, or bulk flow of fluids over longer distances within an organism. In our present work, we focus on bulk flow, which includes chyme moving through the gastrointestinal (GI) tract, blood flowing through the cardiovascular system, lymph moving through the lymphatic system, and sap flowing through the phloem and xylem vascular tissues of plants. The gradient driving bulk flow is the difference in hydrostatic or atmospheric pressure between two places, whereas the sources of resistance are the length and diameter of the tube through which the fluid is flowing and the viscosity of the fluid.

There have been previous investigations of students' understanding of bulk flow in organisms and engineering contexts (15–20). From our previous work, we learned that about half of students we surveyed use "flow down gradients" reasoning (Table 1, Level 3; Ref. 17). However, we also found that students express a variety of ideas about how bulk flow works, and many students did not use the scientific principle of "flow down gradients" in their reasoning even in simplified scenarios (17). Some students explained that the magnitude of the pressure at one point (not the pressure gradient) caused the flow rate (Table 1, Sublevel 2.1). Other students explained that pressure values indicated what flow had occurred but did not use pressure as a driving force for flow (Table 1, Sublevel 2.2). Yet another group of students explained fluid flow as a consequence of the needs or behavior of organisms (Table 1, Level 1). Reasoning with "flow down gradients" may help students better transfer their understanding across physiological systems that are quite different at an anatomical level (5, 9). For instance, the lungs and airways look quite different from a plant's phloem sieve tubes, yet the mechanism for air flow in the respiratory system is governed by the same scientific principle of "mass and heat flow" as sap flow through phloem.

Transfer of understanding is very challenging (10, 21). It is difficult for students to recognize the fundamental connections across systems that appear distinct at observable scales (22–26). Students often focus on the surface features of a phenomenon as key to explaining how it occurred and overlook the underlying principles (6, 27). It has been proposed that students struggle to apply principles across disparate contexts because students activate different cognitive resources based on the perceived context of the situation (28). Cognitive resources are ideas, pieces of knowledge, or ways of reasoning that students have accumulated from their lived experience or informal and formal education and are often situated within specific contexts (29, 30). Numerous articles have documented the impact of context on student reasoning in biological topics such as evolution (25, 31), photosynthesis (32), homeostasis (33), interpretation of histograms (34), and genetic information flow (35).

Even with nearly identical questions (i.e., isomorphic questions), the surface context of an assessment item may strongly influence how a student frames the task (8). For example, in the realm of evolutionary biology, Federer and colleagues (31) found that natural selection items that varied only in whether they asked the student to reason about a plant versus an animal taxon, or familiar versus unfamiliar taxa, elicited different constellations of resources from the students. Our previous work identified a multitude of ways in which students reason about bulk flow of fluids through a tube (17). Now, we explore how students use these types of reasoning across disparate physiological scenarios.

Given the known impact of context on students' reasoning, we were interested in whether assessment item characteristics influence how students reason about bulk flow of fluids down pressure gradients. Specifically, would students use the same type of reasoning across two similar assessment items that differ only in the context presented? In

Table 1. *Student explanations collected in response to interview tasks and short-answer questions to create a three-level bulk flow pressure gradient reasoning framework*

Level	Description
Level 3 "flow down gradients"	The magnitude of the pressure difference is proportional to the rate of fluid flow (i.e., Poiseuille's law).
Level 2	Emerging mechanistic ideas about pressure and flow
Sublevel 2.1 "pressure causes"	Pressures at a single location along the tube, not the pressure gradient, impact fluid flow. • High pressure values cause a large force "pushing" on the fluid. • High pressure values at the end of a tube push back, causing a high resistance to flow.
Sublevel 2.2 "pressures indicate"	The magnitude of pressures are only a result, not the cause, of fluid flow. • A small difference between pressure values at the start and end of a tube indicates that flow is maintained or that the tube has a low resistance and higher flow. • Pressure magnitude indicates the volume of blood that has flowed or is flowing (e.g., high pressures indicates high volumes are flowing, low pressure indicate a high volume of fluid has flowed).
Level 1 "nonmechanistic ideas"	Ideas about characteristics and behaviors of organisms

In previous work, we used student explanations collected in response to interview tasks and short-answer questions to create a three-level bulk flow pressure gradient reasoning framework describing common conceptual patterns in students' reasoning about bulk flow of fluid through a tube in an organism (17).

other words, we wondered how consistent students' reasoning would be across diverse contexts. We were particularly interested in this last question, as consistency of reasoning across different contexts would provide evidence to support the claim that teaching students to reason using Physiology Core Concepts supports transfer of knowledge. In this study we investigated the impact of four different types of item contexts on consistency of reasoning: taxon of the organism, the amount of scientific terminology, whether the bulk flow happened in a living or nonliving system, and what key aspects of the pressure gradient were kept constant. In the paragraphs that follow, we summarize the literature, propose hypotheses, and state our research questions for the different item contexts we investigated.

Given that taxon context influences student reasoning about evolution for natural selection (31, 36), we investigated the impact of organism taxon on students' bulk flow pressure gradient reasoning (i.e., blood flow in animals or phloem sap flow in plants). We hypothesized that animal contexts would be more likely to cue students to use "flow down gradients" pressure gradient reasoning compared with plant contexts because of students' greater familiarity with blood vessels, blood pressure, and blood flow compared with phloem sap and internal plant anatomy.

The amount of contextual detail can also impact student reasoning. For example, when reasoning about complex problems, abstract representations with little contextual detail can elicit higher student performance and more readily facilitate transfer (6, 37, 38). In other examples, concrete representations with greater contextual detail encourage students to use more familiar, intuitive reasoning strategies that may not be helpful or appropriate (39–41). Heckler (37) hypothesized that when students' knowledge is in conflict with scientific understanding, items with contextual detail may trigger application of inaccurate mental models. We investigated the impact of contextual detail (i.e., few or many scientific terms) on students' bulk flow pressure gradient reasoning. We hypothesized that items with few scientific terms (i.e., a more abstract representation) would allow students to focus on relevant aspects of the problem (i.e., the pressure gradient) and not be distracted by scientific terminology or contextual detail that might trigger students to use alternative reasoning.

The tendency to focus on surface features rather than fundamental principles may be exacerbated when students must consider scenarios that are situated in living organisms (18). When considering questions that deal with a specific physiological system (i.e., blood flow in arteries), students may be cued to call on the facts they have learned in class or their everyday knowledge of that system. Often when reasoning about blood flow, students invoke their knowledge of blood pressure being a value of 120/80 or of the cardiovascular system's purpose of delivering oxygen to keep the body alive (17). Accessing either of these resources to answer a question about what causes blood to flow could be counterproductive. We hypothesized that if the same question is now situated in a nonliving system (i.e., fluid flowing through a tube or hose) students may not invoke the needs or behaviors of an organism to explain fluid movement.

The first set of research questions below (RQ1A and RQ1B) investigated the impact of three item contexts that vary in the scenario in which the fluid is flowing: taxa (i.e., blood flow in animals vs. phloem sap flow in plants), the amount of scientific terminology (i.e., few or many specific terms), and living versus nonliving systems (i.e., blood/sap flow in an organism vs. fluid flow in tubes).

> RQ1A: Does assessment item scenario context influence the *level or sublevel* of bulk flow pressure gradient reasoning students use (Table 1)?
>
> RQ1B: To what extent are undergraduate students *consistent* in the level or sublevel of bulk flow pressure gradient reasoning they use across assessment item scenario contexts?

While developing the bulk flow pressure gradient reasoning framework (i.e., Ref. 17), we observed that many students used reasoning that focused on the magnitude of the starting pressure. Their reasoning would sometimes indicate that they thought pressure difference was relevant but the more important feature was the magnitude of the starting pressure. In this study, we more deeply investigated this line of student reasoning by varying which aspect of the pressure gradient was kept constant (i.e., the starting pressure or the pressure gradient). We hypothesized that students who provided a "flow down gradient" explanation when starting pressures were the same would focus on the magnitude of the starting pressure when confronted with different starting pressures but the same pressure gradient.

A second set of research questions (RQ2A and RQ2B) investigated how students reason when different aspects of the pressure gradient were kept constant.

> RQ2A: Do different aspects of the pressure gradient being kept constant influence the *level or sublevel* of bulk flow pressure gradient reasoning students use?
>
> RQ2B: To what extent are undergraduate students *consistent* in the level or sublevel of bulk flow pressure gradient reasoning they use between items that differ in which aspect of the pressure gradient is kept constant?

METHODS

Course and Student Population

We conducted investigations for all research questions in the "Introductory Biology III: Plant & Animal Physiology" course at a 4-yr R1 institution with 614 students enrolled. This course was the third in the major's introductory biology series and surveyed multiple organ systems and physiological processes in animals and plants. The course was taught with high-structure active learning pedagogical practices (42–45). The curriculum was designed using Modell's general models/Physiology Core Concepts (specifically flow down gradients, mass balance, and control systems) as an organizing framework to emphasize the broad utility of reasoning with general models and to align all instructional tools (5, 13, 46–48).

In addition to completing Introductory Biology II with a 2.0 or higher grade point average (GPA) on a 4.0 scale, students must have taken the first and second quarters of general chemistry, but there are no physics or math prerequisites. We obtained student demographic information from the registrar to investigate how student characteristics intersected with their level of reasoning. Our student population was 69% female (binary gender), 24% eligible for the university's Educational Opportunity Program (EOP) for economically or educationally disadvantaged students, 33% who identified with a race/ethnicity of persons historically excluded from science because of their ethnicity or race (PEER; i.e., African American, Hispanic, Native American, or Hawaiian-Pacific Islander) (59), and 14% first-generation college (FGN) students. Additionally, we obtained students' grade point average (GPA) at the start of the term to account for differences in student academic performance before entering this biology course.

All procedures were conducted in accordance with approval from the Institutional Review Board at The University of Washington (STUDY00001316). As is typical for studies that use data collected as a part of normal class events (e.g., homework), a waiver of consent was obtained, and students were able to opt-out of sharing their data with researchers.

RQ1: Impact of Item Scenario Context

To investigate whether assessment item scenario context influenced the level or sublevel of bulk flow pressure gradient reasoning a student used, we developed five bulk flow items that varied in taxa, amount of scientific terminology, or living versus nonliving system (Fig. 1). These items were adapted from Doherty et al. (17). Each item provided students with a prompt describing a scenario for an individual context and a table listing the pressures at the beginning and end of a series of identical tubes. We asked students to identify which tube had the highest flow rate and to explain the reasoning for their choice in writing. To investigate RQ1A (impact of item scenario context on level or sublevel), we developed multiple versions of an online homework assignment to facilitate assessment of all five contexts (Table 2). To investigate RQ1B (impact of item scenario context on consistency across items), we created six different item pairings (Table 2). We separated each item pair by six unrelated assessment items to avoid the possibility that any consistency of reasoning we observed between the two bulk flow items was an artifact of their similar formats. To ensure there was no bias based on the order in which students were presented with a particular item within a pair, we used two homework versions per pair, alternating which item came first. In total, we implemented 12 versions of the homework (Table 2).

We collected data before instruction in week 1 of the 10-wk quarter via an online homework assignment that was graded only for completion rather than correctness (n = 603). As we were investigating consistency (or lack of consistency) in reasoning across questions, it was important that the homework assignment did not allow students to go back and change their answers after seeing subsequent questions. An equal number of each homework version was assigned to

Animal with few terms

The same kind of blood flows through five very long identical blood vessels. Each vessel has the same diameter but different flow rates of blood. A scientist measures the pressure of the blood near the beginning and near the end of each blood vessel.

Blood vessel	Start pressure	End pressure
1	210	170
2	220	160
3	110	20
4	160	140
5	210	190

Which blood vessel has the greatest flow rate (L/min) of blood?

a. Blood vessel 1
b. Blood vessel 2
c. Blood vessel 3
d. Blood vessel 4
e. Blood vessel 5

Explain why the blood vessel you selected has the greatest flow rate (L/min) of blood.

Animal with many terms

A scientist is studying blood flow in the aorta of five different animal species of similar size and age. She found that the composition of the blood was identical in each animal as well as the diameter of their aortas, but the rate of blood flow through the aorta was different. The scientist measured the following pressures at the beginning (i.e., ascending aorta) and near the end (i.e., abdominal aorta) of the aorta.

Blood vessel	Start pressure	End pressure
Zebra	210	150
Camel	200	180
Elk	200	160
Water Buffalo	150	130
Sitka Deer	100	30

Which animal has the greatest flow rate (L/min) of blood through the aorta?

a. Zebra
b. Camel
c. Elk
d. Water Buffalo
e. Sitka Deer

Explain why the animal you selected has the greatest flow rate (L/min) of blood through the aorta.

Plant with few terms

The same kind of sap flows through three very long, phloem tubes. Each tube has the same diameter but different flow rates of sap. A scientist measures the pressure of the sap near the beginning and near the end of each phloem tube.

Phloem tube	Start pressure	End pressure
1	0.85	0.65
2	1.30	1.05
3	0.95	0.85
4	0.65	0.35
5	1.25	1.15

Which phloem tube has the greatest flow rate (L/hour) of sap?

a. Phloem tube 1
b. Phloem tube 2
c. Phloem tube 3
d. Phloem tube 4
e. Phloem tube 5

Explain why the phloem tube you selected has the greatest flow rate (L/hour) of sap.

Plant with many terms

A scientist is studying sap flow in phloem vessels in five different trees of similar size and age. She found that the composition of the sap was identical among the trees as well as the diameter of their phloem vessels, but the rate of sap flow through the phloem vessels was different. The scientist measured the following pressures at the top (i.e., in the tree crown) and near the bottom (i.e., the base of the trunk) of each phloem vessel.

Phloem tube	Start pressure	End pressure
Beech	0.60	0.30
Oak	1.20	1.10
Chestnut	0.90	0.80
Maple	1.25	1.00
Hickory	0.80	0.60

Which tree has the greatest flow rate (L/hour) of sap through the phloem tube?

a. Beech
b. Oak
c. Chestnut
d. Maple
e. Hickory

Explain why the tree you selected has the greatest flow rate (L/hour) of sap through the phloem tube.

Fluid with few terms

The same kind of fluid flows through three very long identical tubes. Each tube has the same diameter but different flow rates of fluid. A scientist measures the pressure of the fluid near the beginning and near the end of each tube.

Tube	Start pressure	End pressure
1	210	150
2	200	180
3	200	160
4	150	130
5	100	10

Which tube has the greatest flow rate (L/min) of fluid?

a. Tube 1
b. Tube 2
c. Tube 3
d. Tube 4
e. Tube 5

Explain why the tube you selected has the greatest flow rate (L/min) of liquid.

Figure 1. Items for RQ1 investigating the impact of item scenario context on students' bulk flow pressure gradient reasoning. Scenarios vary by taxa (i.e., blood flow in animals vs. phloem sap flow in plants), the amount of scientific terminology (i.e., few or many specific terms), and living vs. nonliving systems (i.e., blood/sap flow in an organism vs. fluid flow in tubes).

Table 2. *Description of homework assignment used to collect data for RQ1A,B*

First Bulk Flow Pressure Gradient Item on HW (used in RQ1A)	Second Bulk Flow Pressure Gradient Item on HW	Comparison for RQ1B	Item Pair	HW Version	No. of Students Taking HW Version
Animal with few scientific terms	Plant with few scientific terms	Vary taxa, control sci. terms	1	1	48
Plant with few scientific terms	Animal with few scientific terms	Vary taxa, control sci. terms	1	2	54
Animal with many scientific terms	Plant with many scientific terms	Vary taxa, control sci. terms	2	3	34
Plant with many scientific terms	Animal with many scientific terms	Vary taxa, control sci. terms	2	4	51
Animal with few scientific terms	Animal with many scientific terms	Vary sci. terms, control taxa	3	5	36
Animal with many scientific terms	Animal with few scientific terms	Vary sci. terms, control taxa	3	6	44
Plant with few scientific terms	Plant with many scientific terms	Vary sci. terms, control taxa	4	7	34
Plant with many scientific terms	Plant with few scientific terms	Vary sci. terms, control taxa	4	8	57
Plant with few scientific terms	Fluid with few scientific terms	Living vs. nonliving system	5	9	84
Fluid with few scientific terms	Plant with few scientific terms	Living vs. nonliving system	5	10	71
Animal with few scientific terms	Fluid with few scientific terms	Living vs. nonliving system	6	11	35
Fluid with few scientific terms	Animal with few scientific terms	Living vs. nonliving system	6	12	55

RQ1A: Does assessment item scenario context influence *level or sublevel* of bulk flow pressure gradient reasoning students use? RQ1B: To what extent are undergraduate students *consistent* in the level or sublevel of bulk flow pressure gradient reasoning they use across assessment item scenario contexts? Each homework (HW) assignment contained 2 of the 5 possible bulk flow pressure gradient items: animal with few scientific terms, plant with few scientific terms, animal with many scientific terms, plant with many scientific terms, fluid with few scientific terms.

students in the course. However, some students selected a different version from the one they were assigned, causing our sample sizes for each homework version and item pair to be unequal (Table 2).

We used the same coding rubric and procedure as described in Doherty et al. (17) to code student responses (i.e., choice and written explanation) by level and sublevel of the bulk flow pressure gradient reasoning framework (i.e., Level 1 "nonmechanistic ideas," Level 2.1 "pressure causes," Level 2.2 "pressures indicate," Level 3 "flow down gradients"). See Table 1 and Doherty et al. (17) for a description of levels. We calibrated coding by having two researchers use the coding rubric to code 114 responses. Two codes from independent coders were considered a match if they coded the responses to the same level/sublevel. Interrater reliability for this calibration phase was >90%. After this calibration phase, one researcher coded the rest of the responses, with a second researcher coding 10% of those data. Final agreement was >90%.

Of the 1,206 responses analyzed, only 29 (2.4%) were Level 1. These Level 1 responses were spread roughly evenly across the five item scenario contexts. Given the small number of Level 1 responses, we omitted these explanations from our analysis, as we could not trust our interpretation of the impact of context on these explanations.

We used logistic regression to investigate whether assessment items with different contexts elicited different levels or sublevels of student reasoning (RQ1A). As all items were represented in Item 1 across homework pairs and again in Item 2 (Table 2), we took a more conservative approach to our analysis by limiting it to the level or sublevel of response that students provided for the first item. In this way, we avoided any impact answering the first item might have on the answer for the second item. We ran two models, one investigating level (Level 2 vs. Level 3) and one investigating sublevel (Sublevel 2.1 vs. Sublevel 2.2). The first model investigated whether item scenario context influenced the probability that a student provided a Level 3 or Level 2 explanation. This data set included 594 explanations (Level 1 explanations removed). The second model investigated whether item scenario context influenced the probability that students providing a Level 2 explanation also provided a Sublevel 2.1 or Sublevel 2.2 explanation. This data set included 307 explanations (Level 3 explanations removed). Each model included the fixed effects of item (Fig. 1), students' GPA at the beginning of the course (to account for differences in students' academic performance before entering the current course), and demographic variables: binary gender (men, women), EOP status, PEER status, and FGN status. For all statistical tests and graphing, we used R Statistical software (49). For statistical analysis we used packages *stats* and *sjplot* (50). To create figures we used both *ggplot2* and *ggalluvial* packages (51, 52).

To investigate to what extent undergraduate students used consistent reasoning in the level or sublevel of reasoning they used when item scenario context changed (RQ1B), we calculated the percentage of students who reasoned consistently. We coded each student's response as being consistent if they provided an explanation on the second item that was coded to the same level or sublevel of reasoning as the first item. To investigate whether item pairing or student's level/sublevel of reasoning influenced the probability that a student provided consistent reasoning, we used logistic regression. This data set included 578 students who answered both Items 1 and 2 at either Level 3 or Level 2. Using consistency as the response variable, the model included, as fixed effects, item pair (Table 2), the reasoning level the student used on the first item in each pair, the student's GPA at the beginning of the course, and demographic variables.

RQ2: Impact of Which Aspect of the Pressure Gradient Is Kept Constant

To investigate the level/sublevel of bulk flow pressure gradient reasoning students used and the consistency of their reasoning between items that have the same starting pressure but different pressure gradient magnitudes or the same pressure gradient but different starting pressures, we created a final pair of questions that manipulated these variables (Fig. 2).

In our previous work (17) we found that a subset of students explained that a starting pressure with a large magnitude had a strong pushing force that created a higher flow rate. Some of these students indicated that they thought pressure difference was also important but the most important feature was the magnitude of the starting pressure.

Same starting pressures, different pressure gradient

1) The same kind of blood flows through two very long identical blood vessels. Each vessel has the same diameter. A scientist measures the pressure of the blood near the beginning and near the end of each blood vessel.

Blood vessel	Start pressure	End pressure
1	210	180
2	210	150

Which blood vessel has the greatest flow rate (L/min) of blood?

a. Blood vessel 1
b. Blood vessel 2
c. They will have the same rate of flow

Explain the reasoning for your choice about which blood vessel has the greatest flow rate.

Same pressure gradient, different starting pressures

2) Now let's look at another set of tubes. The same kind of blood flows through two very long identical blood vessels. Each vessel has the same diameter. A scientist measures the pressure of the blood near the beginning and near the end of each blood vessel. Both of these tubes have the same pressure difference but different starting and ending pressures.

Blood vessel	Start pressure	End pressure
1	200	140
2	110	50

Which blood vessel has the greatest flow rate (L/min) of blood?

a. Blood vessel 1
b. Blood vessel 2
c. They will have the same rate of flow

Explain the reasoning for your choice about which blood vessel has the greatest flow rate.

Figure 2. Items for RQ2 investigating the impact of which aspect of the pressure gradient is kept constant on students' bulk flow pressure gradient reasoning. Items either had the same starting pressure but different pressure gradients or the same pressure gradient but different starting pressures.

Therefore, to investigate how students would reason if this option was removed, we constructed the first item to have tubes with the same starting pressure. To investigate whether students who used pressure gradient reasoning on an item that kept starting pressure constant would revert to high pressure pushing reasoning when one starting pressure was higher than the other, we created a second item where the tubes had identical pressure gradients but the starting pressure of one tube was nearly double that of the other tube.

We collected responses to these items on an online homework assignment completed between the fifth and sixth weeks of the 10-wk term (n = 537), which was after instruction on bulk flow in both the respiratory and cardiovascular systems. The homework assignment consisted of two bulk flow items in sequential order and one other physiology item, not used in this study. Again, as we were investigating consistency (or lack of consistency) in reasoning across questions, it was important that students were not allowed to go back and change their answers after seeing the subsequent questions in this online homework. We asked all students both questions in the same order. We coded student responses to reasoning level as described for RQ1. Of the 1,074 responses analyzed, only 21 (2%) were Level 1. Again, we omitted these responses from further analysis.

To investigate whether the aspect of the pressure gradient that was kept constant influenced the level or sublevel of reasoning students used (RQ2A), we used logistic regression. We ran two models, one investigating level (Level 2 vs. Level 3) and one investigating sublevel (Sublevel 2.1 vs. Sublevel 2.2). Each model included, as fixed effects, item (Fig. 2), students' GPA at the beginning of the course, and demographic variables. In these models, Student ID was included as a random effect, as each student took both questions. The first model investigated whether the aspect of the pressure gradient that was kept constant (i.e., same starting pressure or same pressure gradient) impacted the probability that students provided a Level 3 or Level 2 explanation. This data set included 1,053 explanations (Level 1 explanations removed). The second model investigated whether the aspect of the pressure gradient that was kept constant (i.e., same starting pressure or same pressure gradient) impacted the probability that students providing a Level 2 explanation also provided a Sublevel 2.1 or Sublevel 2.2 explanation. This data set included 362 explanations (Level 3 explanations removed). There were few Level 2 responses in this data set, as these data were collected in the middle of the quarter, after instruction. Given the infrequent Level 2 answers, the second model did not converge when it included all demographic variables. Therefore, we ran separate model iterations for each EOP status, PEER status, and FGN status, with each iteration including item, gender, and GPA. We did not vary item order for this homework, so we cannot say whether item order impacted the level of reasoning students used.

To investigate to what extent undergraduate students used consistent reasoning across these two items (i.e., same starting pressure or same pressure gradient; RQ2B), we calculated the percentage of students who answered consistently. We coded each student's response as being consistent if they provided an explanation on the "same pressure gradient" item (second item) that was coded to the same level/sublevel of reasoning as on the first item, the "same starting pressure" item. To investigate whether a student's level/sublevel of reasoning influenced the probability that a student was consistent, we used logistic regression. Note that as there is only one item pair for RQ2B we cannot investigate the impact of item pairing. This data set included the 523 students who provided Level 3 or Level 2 reasoning on both items. Using consistency as the response variable, the model included, as fixed effects, the reasoning level the student used on the "same starting pressure" item, the student's GPA, and demographic variables.

RESULTS

RQ1: Impact of Item Scenario Context

RQ1A: Does assessment item scenario context influence the level or sublevel of bulk flow pressure gradient reasoning students use?

We found that the item scenario context, demographic factors, and incoming GPA did not significantly influence the level or sublevel of reasoning in students' explanations. Across items, 49% of students provided explanations at Level 3, whereas 51% provided Level 2 explanations (Fig. 3). Of the Level 2 explanations, Sublevel 2.1 "pressure causes" was used most commonly (59% of Level 2 responses to all items). Although it appears (in Fig. 3) that a higher proportion of students provided Level 3 explanations for the animal item with few scientific terms and that there was variation in sublevel proportion by context, modeling that controlled for GPA and demographic factors showed that this was not a significant difference (Table 3).

RQ1B: To what extent are undergraduate students consistent in the level or sublevel of bulk flow pressure gradient reasoning they use across assessment item scenario contexts?

We found that most students (76% of 578 students) used consistent reasoning at level/sublevel across all item pairs. Modeling that controlled for GPA, demographic factors, and level of explanation of the first item showed that the likelihood of student reasoning being consistent was not significantly impacted by which item pair they answered (Table 3). Gender was associated with consistent reasoning (odds ratio = 0.60; Table 3). An odds ratio of exactly 1 would mean that gender is not associated with the odds of reasoning consistently at level/sublevel across the two items on the homework. As the odds ratio of the gender comparison for RQ1B is <1 and females are the reference category in our analysis, there was a lower odds of males reasoning consistently. Seventy-eight percent of female students reasoned consistently, whereas only 72% of male students did.

There was also a significant difference in the likelihood of a student answering consistently based on the level of reasoning they used on the first item (Table 3). Eighty-nine percent of students using Level 3 reasoning on the first item used it again on the second item, 77% of students who used Level 2.1 "pressure causes" reasoning on the first item reasoned consistently, and 43% of students using Level 2.2 "pressure indicates" reasoning reasoned consistently (Table 4).

An alluvial diagram represents each student as a ribbon of color that connects the level/sublevel of reasoning a student used on the first item to the level/sublevel of reasoning they used on the second item. The alluvial in Fig. 4 provides a visually informative display of the results shown in Table 4 but also tracks which level/sublevel of reasoning each student used on the first and then the second item. The alluvial shows that of the students who moved from Level 3, Sublevel 2.1, or Sublevel 2.2 on the first item roughly the same proportions moved to each of the other two levels/sublevels. Note that the sublevels of Level 2 are not ordinal but merely represent two different ways students reason with the concept of pressure.

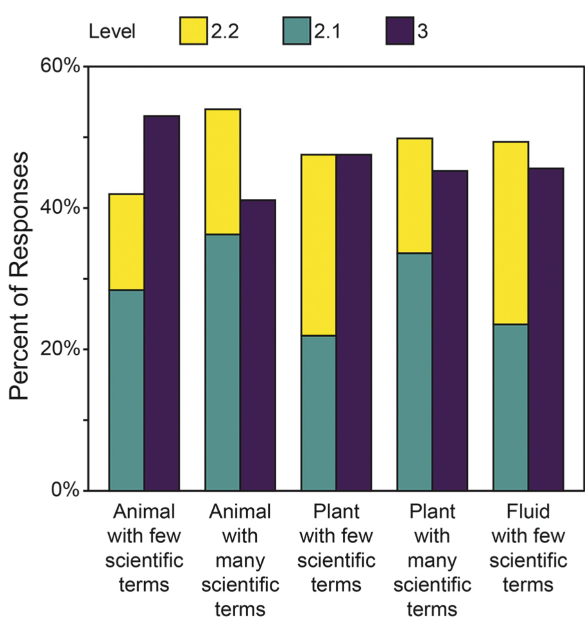

Figure 3. Level/sublevel of bulk flow pressure gradient reasoning by item scenario context on the first question of homework assignment for RQ1 that was given at the beginning of the term prior to instruction. Scenarios vary by taxa (i.e., blood flow in animals vs. phloem sap flow in plants), the amount of scientific terminology (i.e., few or many specific terms), and living vs. nonliving systems (i.e., blood/sap flow in an organism vs. fluid flow in tubes). These data relate to RQ1A: Does assessment item scenario context influence the *level or sublevel* of bulk flow pressure gradient reasoning students use?

RQ2: Impact of Which Aspect of the Pressure Gradient Is Kept Constant

RQ2A: Do different aspects of the pressure gradient being kept constant influence the level or sublevel of bulk flow pressure gradient reasoning students use?

We found that the aspect of the pressure gradient that was kept constant (i.e., same starting pressure or same pressure gradient) and GPA had a significant impact on the level of reasoning students used, although demographic factors did not (Table 4). Students were significantly more likely to reason at Level 3 on the second item where the pressure gradient was kept constant and the starting pressure varied than when they considered tubes where the starting pressure was kept constant but the pressure gradient varied (69% and 62%, respectively). Students with higher GPAs were more likely to provide responses at Level 3.

We also found that which aspect of the pressure gradient was kept constant significantly impacted the likelihood of students using Sublevel 2.1 and Sublevel 2.2 reasoning, although incoming GPA and demographic factors did not. On the first item, "same starting pressure," Sublevel 2.2 "pressures indicate" reasoning was the most common (65% of students using Level 2 reasoning). On Item 2, "same pressure gradient," the most common reasoning was Sublevel 2.1 "pressure causes" (69% of students using Level 2 reasoning).

RQ2B: To what extent are undergraduate students consistent in the level or sublevel of bulk flow pressure gradient reasoning they use between items that differ in which aspect of the pressure gradient is kept constant?

We found that 70% of students were consistent in the level/sublevel of reasoning they provided across the pair of items

Table 3. *Odds ratios, confidence intervals, and P values for analysis models*

	RQ1A Level	RQ1A Sublevel	RQ1B Consistency	RQ2A Level	RQ2A Sublevel (PEER)	RQ2A Sublevel (FGN)	RQ2A Sublevel (EOP)	RQ2B Consistency
Item context (ref: animal with many scientific terms)								
Animal with few scientific terms	0.60 (0.33–1.07, 0.08)	0.98 (0.40–2.44, 0.96)						
Plant with few scientific terms	0.70 (0.40–1.21, 0.20)	1.01 (0.44–2.43, 0.97)						
Plant with many scientific terms	0.76 (0.42–1.38, 0.38)	2.37 (0.97–6.06, 0.06)						
Fluid with few scientific terms	0.70 (0.39–1.24, 0.22)	2.30 (0.97–5.67, 0.06)						
Item pair as described in Table 2 (ref: pair 5)								
Pair 1			1.74 (0.87–3.59, 0.13)					
Pair 2			0.89 (0.44–1.82, 0.74)					
Pair 3			0.99 (0.49–2.04, 0.97)					
Pair 4			0.94 (0.48–1.86, 0.85)					
Pair 6			0.99 (0.51–1.97, 0.99)					
Same pressure gradient item (ref: same starting pressure item)				**1.84 (1.29–2.61, 0.001)**	**0.05 (0.01–0.31, 0.001)**	**0.05 (0.01–0.34, 0.002)**	**0.04 (0.00–0.66, 0.03)**	
Level/sublevel on first item on homework (ref: level 3)								
Sublevel 2.1			**0.39 (0.23–0.66, <0.001)**					**0.28 (0.15–0.51, <0.001)**
Sublevel 2.2			**0.09 (0.05–0.16, <0.001)**					**0.08 (0.05–0.13, <0.001**
GPA	0.91 (0.57–1.44, 0.68)	1.31 (0.67–2.59, 0.44)	0.91 (0.50–1.63, 0.74)	**4.76 (2.22–10.23, <0.001)**	2.62 (0.54–12.57, 0.23)	4.15 (0.75–23.03, 0.10)	1.95 (0.35–10.71, 0.45)	1.70 (0.93–3.12, 0.09)
Gender (ref: female)	1.07 (0.75–1.52, 0.70)	0.78 (0.47–1.30, 0.35)	**0.60 (0.38–0.94, 0.03)**	1.10 (0.63–1.94, 0.73)	3.07 (0.79–11.92, 0.11)	3.13 (0.80–12.34, 0.10)	3.54 (0.65–19.30, 0.14)	**1.78 (1.09–2.93, 0.02)**
PEER (ref: not PEER)	0.97 (0.52–1.83, 0.94)	0.74 (0.31–1.72, 0.49)	2.01 (0.90–4.64, 0.09)	0.67 (0.25–1.82, 0.44)	0.64 (0.14–2.98, 0.570)			0.85 (0.37–1.95, 0.70)
FGN (ref: not FGN)	0.81 (0.54–1.22, 0.32)	1.01 (0.56–1.81, 0.97)	0.66 (0.39–1.12, 0.12)	0.55 (0.28–1.07, 0.08)		3.41 (0.82–14.23, 0.092)		1.23 (0.71–2.18, 0.46)
EOP (ref: not EOP)	0.76 (0.46–1.27, 0.302)	1.29 (0.64–2.59, 0.47)	0.98 (0.52–1.88, 0.96)	1.40 (0.61–3.23, 0.43)			0.24 (0.03–1.68, 0.15)	1.14 (0.56–2.38, 0.73)
R^2	0.012	0.049	0.185	0.070	0.180	0.197	0.202	0.256

Values are odds ratios, with confidence intervals and *P* values in parentheses. Significant predictors are in bold. Demographic categories included students in the educational opportunity program (EOP) for economically or educationally disadvantaged students, persons historically excluded from science because of their ethnicity or race (PEER), and first-generation college students (FGN). GPA, grade point average. RQ1A: Does assessment item scenario context influence the *level or sublevel* of bulk flow pressure gradient reasoning students use? RQ1B: To what extent are undergraduate students *consistent* in the level or sublevel of bulk flow pressure gradient reasoning they use across assessment item scenario contexts? RQ2A: Do different aspects of the pressure gradient being kept constant influence the *level or sublevel* of bulk flow pressure gradient reasoning students use? RQ2B: To what extent are undergraduate students *consistent* in the level or sublevel of bulk flow pressure gradient reasoning they use between items that differ in which aspect of the pressure gradient is kept constant?

Table 4. *Percentage of students who provided short-answer responses consistent in level/sublevel of bulk flow pressure gradient reasoning across scenario context comparisons*

	Taxa Item Pair		Scientific Detail Item Pair		Living vs. Nonliving System Item Pair		
	1	*2*	*3*	*4*	*5*	*6*	
	Animal or plant with few scientific terms	Animal or plant with many scientific terms	Animal with few or many scientific terms	Plant with few or many scientific terms	Plant or fluid with few scientific terms	Animal or fluid with few scientific terms	Across Item Pairs
Level 3 "flow down gradients"	98%	85%	90%	87%	92%	80%	89%
Level 2, 2.1 "pressure causes"	77%	78%	78%	75%	78%	77%	77%
Level 2, 2.2 "pressures indicate"	60%	43%	29%	43%	34%	54%	43%

Percentage of students who provided short-answer responses consistent in level/sublevel of bulk flow pressure gradient reasoning across scenario context comparisons (see Table 2 for item pair descriptions). These data address RQ1B: To what extent are undergraduate students *consistent* in the level or sublevel of bulk flow pressure gradient reasoning they use across assessment item scenario contexts?

addressing the same starting pressure versus the same pressure gradient. Modeling that controlled for GPA, demographic factors, and level of explanation of the first item again showed that gender and level of explanation on the first item were significantly associated with the probability that students were consistent (Table 3). Again, students reasoned most consistently when using Level 3 reasoning and least consistently when using Sublevel 2.2 reasoning. Eighty-six percent of students who used Level 3 reasoning on the first item (same starting pressure) used it again on the second item (same pressure gradient); 61% of students who used Level 2.1 "pressure causes" reasoning on the first item were consistent; and 34% of students who used Level 2.2 "pressure indicates" reasoning were consistent. On this pair of items, however, male students were more consistent than female students (86% and 67%, respectively).

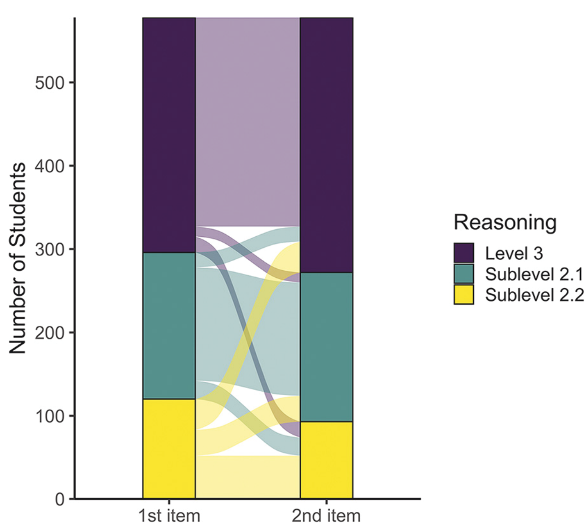

Figure 4. Alluvial diagram of level/sublevel of bulk flow pressure gradient reasoning by item order (1st bulk flow item on the homework or 2nd bulk flow item on the homework) on homework assignment for RQ1 that was given at the beginning of the term, prior to instruction. Data combined for all homework versions (across all item scenario contexts). An alluvial diagram represents each student as a ribbon of color that connects the level/sublevel of reasoning a student used on the first item to the level/sublevel they used on the second item. These data relate to RQ1B: To what extent are undergraduate students *consistent* in the level or sublevel of bulk flow pressure gradient reasoning they use across assessment item scenario contexts?

When we plotted which students changed answers, the alluvial displayed three striking patterns (Fig. 5). First, when a student using Level 3 reasoning on the "same starting pressure" item changed reasoning for the "same pressure gradient" item, they changed almost exclusively to using Sublevel 2.1 "pressure causes" reasoning and not to Sublevel 2.2 "pressures indicate" reasoning. Second, when a student using Sublevel 2.1 "pressure causes" reasoning on the "same starting pressure" item changed reasoning, they changed almost exclusively to using Level 3 reasoning, with few moving to Sublevel 2.2 "pressure indicates." Third, when a student using Sublevel 2.2 "pressures indicate" reasoning on the "same starting pressure" item changed reasoning, approximately two-thirds of those who changed moved to Level 3 reasoning and the remaining third moved to Sublevel 2.1 "pressure causes" reasoning.

DISCUSSION

Accessing cognitive resources when answering questions is a dynamic process that is often context dependent for students (28, 53–55). In physiology, key cognitive resources are understanding the Physiology Core Concepts and their application across a plethora of contexts (e.g., across taxa and across organ systems). These cognitive resources are proposed to be the key to the successful transfer of understanding from one specific situation to a novel situation (5, 9, 13). To accurately predict the rate of blood flow to any organ in the body, or to understand why rising too quickly from a seated position may cause a person to feel light headed or even pass out, how drinking seawater causes diarrhea, how bile moves from the gallbladder to the duodenum, and why plants produce less food during a drought, it is important for students to have a deep understanding of the Physiology Core Concept of "flow down gradients" and be able to apply it across systems with structures that are uniquely named for that system. However, research from other fields on using the "core concepts" of their disciplines to support transfer indicate that superficial features of the problem or scenario often interfere with students' successful transfer (6, 22–27). We found that when students used the Physiology Core

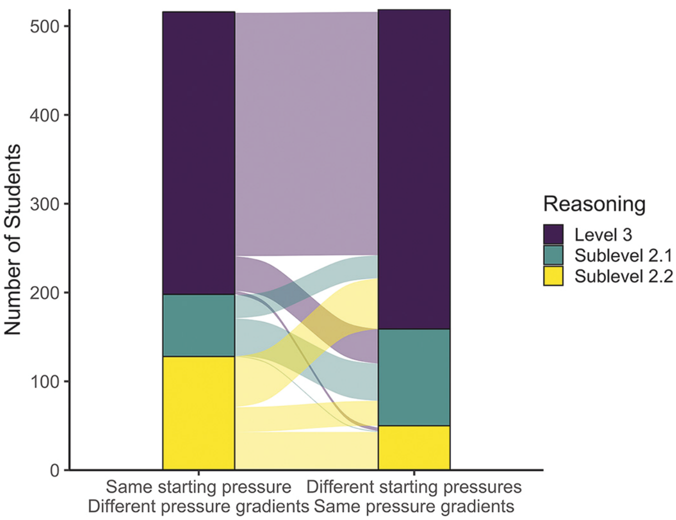

Figure 5. Alluvial diagram of level/sublevel of bulk flow pressure gradient reasoning on homework assignment for RQ2 that was given in the middle of the term. The first item on the assessment was "same starting pressure, different pressure gradients." An alluvial diagram represents each student as a ribbon of color that connects the level/sublevel of reasoning a student used on the first item to the level/sublevel they used on the second item. These data relate to RQ2A: Do different aspects of the pressure gradient being kept constant influence the *level or sublevel* of bulk flow pressure gradient reasoning students use? and RQ2B: To what extent are undergraduate students *consistent* in the level or sublevel of bulk flow pressure gradient reasoning they use between items that differ in which aspect of the pressure gradient is kept constant?

Concept of "flow down gradients" they were able to reason correctly and consistently about fluid flow through tubes with different superficial features. Furthermore, ours is the first research to provide empirical evidence that students who use a Physiology Core Concept successfully transfer mechanistic reasoning across disparate physiological scenarios.

Impact of Item Scenario Context on Student Level/Sublevel of Reasoning

Our previous research on how students reason about bulk flow pressure gradient items found three reasoning levels, with Level 2 consisting of two sublevels (Table 1, Ref. 17). Our present findings indicate that across item scenario context (i.e., items with different taxa, different amounts of scientific detail, and living and nonliving systems) there is no significant difference in the level or sublevel of reasoning students use to explain their answers. Therefore, we conclude that the superficial features of the item scenario context do not impact the difficulty of bulk flow pressure gradient items for students entering a sophomore-level class on introductory physiology. We were surprised that item scenario context did not greatly influence students' level/sublevel of reasoning, as others have found that students' reasoning is impacted by item scenario context (23–26, 31). We had hypothesized that students would be more familiar with problems that dealt with animals and blood flow, rather than with plants and sap flow, and the greater familiarity with animals could positively impact their reasoning. Similarly, we predicted that students would reason at higher levels on an item that was situated in a nonliving system (i.e., fluid in a tube), as this context might be more familiar and less likely to activate content-specific ideas than items set in a living organism. Conversely, we had hypothesized that questions with more scientific terminology could prompt students to inappropriately apply their everyday knowledge or incorrectly apply newly gained knowledge and therefore reason at a lower level.

To explain our findings that item scenario context did not impact students' reasoning level/sublevel, we propose that the format of our items is highly scaffolded, which allowed students to see the items as "fluid flow through a tube" questions and to explain the phenomena regardless of whether they knew what phloem sap or a phloem tube is. It is possible that if we had asked the students a more open-ended, less scaffolded question (e.g., without pressure values or indicating that pressure is important to fluid movement across these contexts) about what causes fluid or sap to flow through an organism, we might have elicited much different student reasoning about their understanding about fluid flow across contexts. Biology students should be able to apply flow down gradients to diverse contexts, and our future work will explore how well students can reason without scaffolding.

Impact of Aspect of the Pressure Gradient Kept Constant on Student Level/Sublevel of Reasoning

Although item scenario context did not appear to influence which cognitive resources students activated and hence the level/sublevel of reasoning they used, changing which aspect of the pressure gradient was kept constant did. More students used Level 3 "pressure gradient" reasoning on the second item, "same pressure gradient," than on the first item, indicating that the second item was easier for students. We propose that the students who used Level 3 reasoning on the second item, but not the first, saw that both the starting and ending pressures were different in the second item and, while exploring the numbers, discovered that the pressure gradients were identical and used that notable pattern to reason.

Within students using Level 2 reasoning, more students used Sublevel 2.2 "pressures indicate" reasoning on the first item, "same starting pressure," whereas more students used Sublevel 2.1 "pressure causes" reasoning on the second item, "same pressure gradient." To try to determine why Sublevel 2.2 "pressures indicate" reasoning was more common on the first item, we looked more closely at the student explanations in relation to the pressure values in the item. These explanations highlighted that different students privileged some values in the tables of pressure values over others. We propose that students using Sublevel 2.2 reasoning noticed that the beginning pressures were the same and used the different ending pressures to explain what happened to generate those different ending pressures, rather than using the difference in pressures to explain differences in flow. These students may have wondered "What could cause the same starting pressures but different ending pressures? What is the difference in ending pressure telling me? Could it be indicating a difference in flow?" In their responses, some of these students reasoned that a high ending pressure was indicative of high fluid volume due to a higher flow rate, whereas others saw the low ending pressure as indicating low volume left in the tube, and hence that a higher flow rate must have occurred.

Similarly, a closer analysis of students' explanations in relation to the pressure values in the second item, "same pressure gradient," also helped to explain why there are more Sublevel 2.1 "pressure causes" responses to the second item. We designed this item specifically to highlight the same pressure gradient across both tubes, with one tube having a higher starting pressure. Students using Sublevel 2.1 "pressure causes" reasoning selected the tube with the highest pressure values, as this tube has the "highest pushing pressure," which causes the highest flow rate. Students could have used this same reasoning on the first item (they would just choose the "flows are the same" option); however, the large difference in magnitudes of the starting pressures on the second item seemed to cue students to use "pressure causes" reasoning. We propose that these students are currently learning to use gradients as driving forces. These students know that gradients can be driving forces, but, to them, high pressure values are too compelling for explaining fluid flow.

Level/Sublevel of Reasoning Impacts Consistency

We that found a high percentage of students reasoned consistently across all item comparisons (item scenario context pairs and the aspect of pressure gradient pair), and we found that the level/sublevel students used to reason on the first item of the homework was predictive of whether or not they used the same level/sublevel of reasoning on the second item. Students who used Level 3 reasoning on the first item were, by far, the most consistent (i.e., a higher percentage of students who used Level 3 on the first item also used Level 3 reasoning on the second item), whereas students who used Sublevel 2.2 reasoning were the least consistent. This pattern held for both the item scenario context pairs (L3 89% > SL2.1 77% > SL2.2 43%) and aspect of the pressure gradient pair (L3 86% > SL2.1 61% > SL2.2 34%). To avoid the possibility that any consistency of reasoning we observed was an artifact of the similar item format, we purposely inserted six other items between the two bulk flow items to act as task-switching items. Task switching has been shown to disrupt student problem-solving skills and often leads to more errors (56, 57). We found that this type of task switching did not disrupt students using Level 3 reasoning but may have been disruptive for students using Level 2 reasoning.

The high degree of consistency among students who used Level 3 reasoning indicated that these students had a firm understanding that the magnitude of the pressure gradient determines flow rate. This suggests that being able to apply the principle of flow down gradients found in the Physiology Core Concepts can support students in transferring their knowledge across contexts (5, 9). We propose that students using Level 2 reasoning have a less robust understanding of bulk flow. These students may have called on different cognitive resources in response to each prompt (28, 58). These students may realize that they should invoke a mechanistic explanation involving pressure, but they do not yet consistently know which of their pressure resources to apply (e.g., Poiseuille's law, ideal gas law, static pressure equation, pressure is directly or inversely proportional to volume).

Even within students using Level 2 reasoning, students who use Sublevel 2.2 "pressure indicates" reasoning are particularly inconsistent (43 or 34% of students using Sublevel 2.2 reasoning on the first item used it on the second item). We that know students using Sublevel 2.2 reasoning can use pressure or pressure gradients as a driving force because when they do not use Sublevel 2.2 reasoning (are inconsistent) they use either Level 3 "pressure gradient" reasoning or Level 2.1 "pressure causes" reasoning. However, these students do not reason consistently across these levels (Level 3 or Sublevel 2.1). We propose that this group of students is not yet committed to reasoning with pressure or pressure gradients as a causal mechanism. When we originally created the bulk flow pressure gradient reasoning framework (17), we did not consider one of the sublevels to be more productive than the other. On the basis of the present results, we now hypothesize that Sublevel 2.1 "pressure causes" reasoning is more productive than Sublevel 2.2 "pressures indicate" reasoning.

When Students Were Inconsistent, Did the Level They Used to Explain the First Item Provide Insight into How They Reasoned on the Second Item?

The alluvial diagram for RQ1 (Fig. 4) did not uncover any obvious patterns as to where students moved from or to in reasoning level between item scenario contexts. However, the alluvial diagram for RQ2 (Fig. 5) showed an interesting pattern in how students using different levels/sublevels of reasoning changed the level/sublevel of reasoning they used on the second item. We had hypothesized that the design of the items in RQ2 would challenge students who vacillated between high pressure pushing and the actual pressure gradient as a cause for rate of flow. Therefore, we were not surprised to see a number of students who provided a Level 3 explanation to the first item, "same starting pressure," change to Sublevel 2.1 "pressure causes" reasoning on the second item, "same pressure gradient," as this item now had one tube with a higher starting pressure. However, the most striking pattern observed was that any student who used driving force reasoning on the first item (i.e., Level 3 "pressure gradient" or Sublevel 2.1 "pressure causes") continued to use some form of driving force reasoning on the second item. This may indicate that these students see pressure as the causal agent in fluid flow. In contrast, students who used Sublevel 2.2 "pressures indicate" moved to either of the other two levels on the second item. We suggest that students using Sublevel 2.2 "pressure indicates" reasoning are less stable in their understanding of what causes fluid flow.

Gender Was Not Associated with Reasoning Level/Sublevel but Was Associated with Reasoning Consistency

There was no difference in the level of reasoning used by students who identified as either female or male across all items. However, we did find that female students were more consistent than male students in the level of reasoning they used when the two items varied by context (78% and 72%, respectively). The reverse was true for the items that varied the aspect of the pressure gradient kept constant; males reasoned more consistently than females (86% and 67%, respectively). On items that varied the text in the prompt but always presented a table with five tubes with the same pattern of starting pressures and pressure differences (Fig. 1), it seems that females may have more easily recognized that these questions were similar and used the same pattern of reasoning for both. However, when the values within the pressure tables were quite different (i.e., items that varied in which aspect of the pressure gradient was kept constant), males were more consistent.

Implications for Teaching

As instructors, one of our goals is for students to transfer knowledge to novel situations. One teaching method that has been shown to enhance students' ability to transfer knowledge is to provide multiple different examples of the same concept (8). Students may not be as familiar with the specific physiological processes of phloem sap through sieve tubes or xylem sap through xylem vessels in plants. However, students using Level 3 reasoning consistently used this same level of reasoning even on items in these potentially unfamiliar contexts. Furthermore, some students using Level 2 reasoning on the first item moved to Level 3 reasoning on a new example, which may indicate that providing a variety of contexts could stimulate students to consider other types of reasoning. Therefore, we suggest that instructors incorporate disparate examples in instruction or homework to help students learn and successfully transfer the Physiology Core Concept of "flow down [pressure] gradients."

In this study, at the beginning of an introductory biology course for sophomores, we found that 51% of students used Level 2 reasoning. By midquarter, after respiratory and cardiovascular instruction designed to teach "flow down [pressure] gradients," we found that 38% of students still used Level 2 reasoning at least once on our items. After another 5 weeks of instruction, one of which was a week that included bulk flow instruction, a similar bulk flow item was given on an end-of-quarter homework assignment, and we found that only 14% of students used Level 2 reasoning (analyses not shown). Although pressure gradients are a key aspect of mechanistic reasoning in physiology, initially many students in our sample did not apply the concept of a pressure gradient and instead used other resources they have about pressure to reason. They may have gained these resources about pressure from their everyday life (water faucets or garden hoses), from their physics courses (Bernoulli's law), their chemistry courses (Dalton's law, Henry's law, ideal gas law), and/or their physiology courses (Poiseuille's law, Fick's law). Therefore, instructors need to help students learn how to differentiate and organize their multiple ideas and concepts about pressure.

We propose that it is important for instructors to provide students with ample time and a variety of practice items to help them develop the understanding that rate of flow in a tube is directly proportional to the magnitude of the gradient (i.e., the difference in pressures between two points). In particular, our findings suggest that students need deliberate practice with questions that change different aspects of the pressure gradient while holding others constant. Too often, textbooks and in-class examples provide only one set of values at each end of the tube, so students have not dealt with variation. In addition to using the items in this article to provide that practice, instructors could ask students to generate a set of tubes with different starting pressures but similar flow rates. This activity could provide the deliberate and focused practice that students need to understand the nature of each of the variables in the bulk flow equations; flow (volume/time) is directly proportional to the magnitude of the gradient (pressure difference between two points) and indirectly proportional to the magnitude of the resistance (which is not the pressure downstream pushing back).

Limitations

We did not create two sets of homework for RQ2 items that presented the two items in a different order, nor did we intersperse six other items not associated with this research. We realize that item order and lack of items between the two items may have influenced student performance on these two items. Given that we found that item order did not impact the reasoning level students used across the items pairs used in RQ1, we posit that item order did not impact student reasoning level on items in RQ2. However, we did not test item order for RQ2.

This research was done in one class of 614 STEM undergraduates taking the third and final course in the Introductory Biology series at a competitive R1 university. To gain access to this course students had to earn a 2.0 on a 4.0 scale in the previous course, whose course mean is usually a 2.8. Therefore, our conclusions must be limited to that student population. Moving forward, it will be important to greatly expand institutional diversity in our study population. In particular, investigating student populations with less biology experience may yield more Level 1 responses, allowing us to explore how item characteristics influence students using that type of reasoning.

Although having two bulk flow items on the same homework allowed us to investigate consistency in reasoning level/sublevel, it is possible that the format of the two questions was similar enough that students may have ignored the surface feature presented in the scenario (e.g., blood vessels or phloem tubes) and only focused on the numbers in the table. However, given that we found the level of reasoning used to answer the first item had a large impact on the probability that a student reasoned consistently, we feel that the similar format of items may have had only a minimal contribution.

Conclusions

We investigated the impact of item scenario context (i.e., taxa, amount of scientific details, living or nonliving system) and which aspect of the pressure gradient is kept constant on the level/sublevel and consistency of student reasoning when answering bulk flow pressure gradient questions. Item scenario context did not impact reasoning level/sublevel or consistency of reasoning across multiple items on the same assessment by students entering a sophomore-level introductory physiology course. We did find that the aspect of the pressure gradient that is kept constant impacts the level/

sublevel that students use at the midpoint of this same course (after completion of both respiratory and cardiovascular instruction). More students used Level 3 reasoning when the pressure gradient was kept constant and the starting pressures were different, whereas more students used Sublevel 2.2 "pressure indicates" reasoning when the starting pressures were the same and the pressure gradients were different. Varying the patterns in the pressure values given to students seems to cue students to use different cognitive resources they have about the concept of pressure.

We found that students using Level 3 "pressure gradient" reasoning were the most consistent, whereas students using Level 2 reasoning were less consistent. We argue that these results support the claim that learning to reason with the Physiology Core Concept of "flow down gradients" supports successful transfer of knowledge. More broadly, these findings are the first empirical evidence to support the claim that using Physiology Core Concept reasoning supports transfer of knowledge across different physiological systems. We suggest that instructors who wish to develop their students' ability to transfer this key physiological principle should consider integrating it into their physiology curriculum.

DATA AVAILABILITY

Data will be made available upon reasonable request.

ACKNOWLEDGMENTS

We thank Jennifer White and Kyle Loucks for assistance in implementing this project and the UW Biology Education Research Group for feedback on this project. We also thank three anonymous reviewers and our managing editor for constructive feedback on earlier versions of the manuscript.

GRANTS

This work was supported by National Science Foundation awards (NSF DUE 1661263/1660643).

DISCLAIMERS

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

DISCLOSURES

J. H. Doherty and M. P. Wenderoth are editors of *Advances in Physiology Education* and were not involved and did not have access to information regarding the peer-review process or final disposition of this article. An alternate editor oversaw the peer-review and decision-making process for this article. None of the other authors has any conflicts of interest, financial or otherwise, to disclose.

AUTHOR CONTRIBUTIONS

J.H.D., J.A.C., E.E.S., L.N.J., J.M., K.C.H., and M.P.W. conceived and designed research; J.H.D., J.A.C., and E.E.S. performed experiments; J.H.D., J.A.C., and E.E.S. analyzed data; J.H.D., J.A.C., E.E.S., and M.P.W. interpreted results of experiments; J.H.D., J.A.C., and M.P.W. prepared figures; J.H.D., J.A.C., and M.P.W. drafted manuscript; J.H.D., J.A.C., E.E.S., L.N.J., J.M., K.C.H., and M.P.W. edited and revised manuscript; J.H.D., J.A.C., E.E.S., L.N.J., J.M., K.C.H., and M.P.W. approved final version of manuscript.

REFERENCES

1. **NRC.** *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Washington, DC: National Academies Press, 2012.
2. **Bao L**, **Koenig K.** Physics education research for 21st century learning. *Discip Interdiscip Sci Educ Res* 1: 2, 2019. doi:10.1186/s43031-019-0007-8.
3. **AAAS.** *Vision and Change in Undergraduate Biology Education: a Call to Action*. Washington, DC: American Association for the Advancement of Science, 2011.
4. **NRC.** *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (Online). National Academies Press. 2012. http://www.nap.edu/catalog.php?record_id=13165 [2014 Sep 30].
5. **Modell HI.** How to help students understand physiology? Emphasize general models. *Adv Physiol Educ* 23: 101–107, 2000. doi:10.1152/advances.2000.23.1.S101.
6. **Goldstone RL**, **Son JY.** The transfer of scientific principles using concrete and idealized simulations. *J Learn Sci* 14: 69–110, 2005. doi:10.1207/s15327809jls1401_4.
7. **Hartley LM**, **Wilke BJ**, **Schramm JW**, **D'Avanzo C**, **Anderson CW.** College students' understanding of the carbon cycle: contrasting principle-based and informal reasoning. *BioScience* 61: 65–75, 2011. doi:10.1525/bio.2011.61.1.12.
8. **Kaminske AN**, **Kuepper-Tetzel CE**, **Nebel CL**, **Sumeracki MA**, **Ryan SP.** Transfer: a review for biology and the life sciences. *CBE Life Sci Educ* 19: es9, 2020. doi:10.1187/cbe.19-11-0227.
9. **Michael JA.** Use of core concepts of physiology can facilitate student transfer of learning. *Adv Physiol Educ* 46: 438–442, 2022. doi:10.1152/advan.00005.2022.
10. **National Research Council (NRC).** *How People Learn: Brain, Mind, Experience, and School: Expanded Edition* (Online). National Academies Press, 2000. http://www.nap.edu/openbook.php?record_id=9853&page=31 [2014 Sep 30].
11. **Michael JA.** What makes physiology hard for students to learn? Results of a faculty survey. *Adv Physiol Educ* 31: 34–40, 2007. doi:10.1152/advan.00057.2006.
12. **Slominski T**, **Grindberg S**, **Momsen J.** Physiology is hard: a replication study of students' perceived learning difficulties. *Adv Physiol Educ* 43: 121–127, 2019. doi:10.1152/advan.00040.2018.
13. **Michael JA**, **Cliff W**, **McFarland JL**, **Modell H**, **Wright A.** *The Core Concepts of Physiology: A New Paradigm for Teaching Physiology* (Online). Springer-Verlag, 2017. https://www.springer.com/gp/book/9781493969074 [2019 Aug 7].
14. **Michael JA**, **McFarland JL.** The core principles ("big ideas") of physiology: results of faculty surveys. *Adv Physiol Educ* 35: 336–341, 2011. doi:10.1152/advan.00004.2011.
15. **Besson U.** Students' conceptions of fluids. *Int J Sci Educ* 26: 1683–1714, 2004. doi:10.1080/0950069042000243745.
16. **Brown S**, **Beddoes K**, **Montfort D**, **Baghdanov A.** Engineering students' fluid mechanics misconceptions: a description and theoretical explanation. *Int J Eng Educ* 33: 1149–1162, 2017.
17. **Doherty JH**, **Scott EE**, **Cerchiara JA**, **Jescovitch LN**, **McFarland JL**, **Haudek KC**, **Wenderoth MP.** What a difference in pressure makes! Undergraduate students' reasoning about bulk flow down pressure gradients. *CBE Life Sci Educ*. In press. doi:10.1187/cbe.20-01-0003.
18. **Michael JA**, **Wenderoth MP**, **Modell HI**, **Cliff W**, **Horwitz B**, **McHale P**, **Richardson D**, **Silverthorn D**, **Williams S**, **Whitescarver S.** Undergraduates' understanding of cardiovascular phenomena. *Adv Physiol Educ* 26: 72–84, 2002. doi:10.1152/advan.00002.2002.
19. **Vitharana PR.** Student misconceptions about plant transport—a Sri Lankan example. *Eur J Sci Math Educ* 3: 275–288, 2015. doi:10.30935/scimath/9437.
20. **Wang JR.** Development and validation of a two-tier instrument to examine understanding of internal transport in plants and the human circulatory system. *Int J Sci Math Educ* 2: 131–157, 2004. doi:10.1007/s10763-004-9323-2.

21. **Goldstone RL**, **Day SB.** Introduction to "New Conceptualizations of Transfer of Learning". *Educ Psychol* 47: 149–152, 2012. doi:10.1080/00461520.2012.695710.
22. **Docktor JL**, **Mestre JP**, **Ross BH.** Impact of a short intervention on novices' categorization criteria. *Phys Rev Spec Top Phys Educ Res* 8: 020102, 2012. doi:10.1103/PhysRevSTPER.8.020102.
23. **Mamede S**, **Splinter TA**, **van Gog T**, **Rikers RM**, **Schmidt HG.** Exploring the role of salient distracting clinical features in the emergence of diagnostic errors and the mechanisms through which reflection counteracts mistakes. *BMJ Qual Saf* 21: 295–300, 2012. doi:10.1136/bmjqs-2011-000518.
24. **Mohan L**, **Chen J**, **Anderson CW.** Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *J Res Sci Teach* 46: 675–698, 2009. doi:10.1002/tea.20314.
25. **Nehm RH**, **Ha M.** Item feature effects in evolution assessment. *J Res Sci Teach* 48: 237–256, 2011. doi:10.1002/tea.20400.
26. **Teichert MA**, **Tien LT**, **Anthony S**, **Rickey D.** Effects of context on students' molecular-level ideas. *Int J Sci Educ* 30: 1095–1114, 2008. doi:10.1080/09500690701355301.
27. **Chi MT**, **Roscoe RD**, **Slotta JD**, **Roy M**, **Chase CC.** Misconceived causal explanations for emergent processes. *Cogn Sci* 36: 1–61, 2012. doi:10.1111/j.1551-6709.2011.01207.x.
28. **Hammer D**, **Elby A**, **Scherr RE**, **Redish EF.** Resources, framing, and transfer. In: *Transfer of Learning from a Modern Multidisciplinary Perspective*. Greenwich, CT: Information Age Publishing, 2005, p. 89–119.
29. **diSessa AA.** Toward an epistemology of physics. *Cogn Instr* 10: 105–225, 1993. doi:10.1080/07370008.1985.9649008.
30. **Minstrell J.** Facets of students' knowledge and relevant instruction. In: *Physics Learning: Theoretical Issues and Empirical Studies*, edited by Duit R, Goldberg F, Niedderer H. Kiel, Germany: University of Kiel, 1992, p. 110.
31. **Federer MR**, **Nehm RH**, **Opfer JE**, **Pearl D.** Using a constructed-response instrument to explore the effects of item position and item features on the assessment of students' written scientific explanations. *Res Sci Educ* 45: 527–553, 2015. doi:10.1007/s11165-014-9435-9.
32. **Weston M**, **Haudek KC**, **Prevost L**, **Urban-Lurain M**, **Merrill J.** Examining the impact of question surface features on students' answers to constructed-response questions on photosynthesis. *CBE Life Sci Educ* 14: ar19, 2015. doi:10.1187/cbe.14-07-0110.
33. **McFarland JL**, **Price RM**, **Wenderoth MP**, **Martinková P**, **Cliff W**, **Michael JA**, **Modell H**, **Wright A.** Development and validation of the homeostasis concept inventory. *CBE-Life Sci Educ* 16: ar35, 2017. doi:10.1187/cbe.16-10-0305.
34. **Kaplan JJ**, **Lyford A**, **Jennings JK.** Effects of question stem on student descriptions of histograms. *Stat Educ Res J* 17: 85–102, 2018. doi:10.52041/serj.v17i1.177.
35. **Uhl JD**, **Shiroda M**, **Haudek KC.** Developing assessments to elicit and characterize undergraduate mechanistic explanations about information flow in biology. *J Biol Educ* 2022: 1–20, 2022. doi:10.1080/00219266.2022.2041460.
36. **Kampourakis K**, **Zogza V.** Preliminary evolutionary explanations: a basic framework for conceptual change and explanatory coherence in evolution. *Sci Educ* 18: 1313–1340, 2009. doi:10.1007/s11191-008-9171-5.
37. **Heckler AF.** Concrete vs. abstract problem formats: a disadvantage of prior knowledge. In: *ICLS '10 Proceedings of the 9th International Conference of the Learning Sciences*. Bloomington, IN: International Society of the Learning Sciences, 2010, vol. 1, p. 365–371.
38. **Kaminski JA**, **Sloutsky VM**, **Heckler AF.** The cost of concreteness: the effect of nonessential information on analogical transfer. *J Exp Psychol Appl* 19: 14–29, 2013. doi:10.1037/a0031931.
39. **Carraher TN**, **Carraher DW**, **Schliemann AD.** Written and oral mathematics. *J Res Math Educ* 18: 83–97, 1987. doi:10.2307/749244.
40. **Kintsch W**, **Greeno JG.** Understanding and solving word arithmetic problems. *Psychol Rev* 92: 109–129, 1985. doi:10.1037/0033-295X.92.1.109.
41. **Pollard P**, **Evans JS.** Content and context effects in reasoning. *Am J Psychol* 100: 41–60, 1987. doi:10.2307/1422641.
42. **Kolpikova EP**, **Chen DC**, **Doherty JH.** Does the format of preclass reading quizzes matter? An evaluation of traditional and gamified, adaptive preclass reading quizzes. *CBE Life Sci Educ* 18: ar52, 2019. doi:10.1187/cbe.19-05-0098.
43. **Freeman S**, **O'Connor E**, **Parks JW**, **Cunningham M**, **Hurley D**, **Haak D**, **Dirks C**, **Wenderoth MP.** Prescribed active learning increases performance in introductory biology. *CBE Life Sci Educ* 6: 132–139, 2007. doi:10.1187/cbe.06-09-0194.
44. **Jackson MA**, **Tran A**, **Wenderoth MP**, **Doherty JH.** Peer vs. self-grading of practice exams: which is better? *CBE Life Sci Educ* 17: es44, 2018. doi:10.1187/cbe.18-04-0052.
45. **Moorleghen DM**, **Oli N**, **Crowe AJ**, **Liepkalns JS**, **Self CJ**, **Doherty JH.** Impact of automated response systems on in-class cell phone use. *Biochem Mol Biol Educ* 47: 538–546, 2019. doi:10.1002/bmb.21257.
46. **Michael JA**, **McFarland JL.** Another look at the core concepts of physiology: revisions and resources. *Adv Physiol Educ* 44: 752–762, 2020. doi:10.1152/advan.00114.2020.
47. **Scott EE**, **Wenderoth MP**, **Doherty JH.** Design-based research: a methodology to extend and enrich biology education research. *CBE Life Sci Educ* 19: es11, 2020. doi:10.1187/cbe.19-11-0245.
48. **Scott EE**, **Cerchiara J**, **McFarland JL**, **Wenderoth MP**, **Doherty JH.** How students reason about matter flows and accumulations in complex biological phenomena: an emerging learning progression for mass balance. *J Res Sci Teach* 60: 63–99, 2023. doi:10.1002/tea.21791.
49. **R Core Team.** R: a language and environment for statistical computing (Online). R Foundation for Statistical Computing, 2021. https://www.R-project.org/.
50. **Lüdecke D.** 2021. sjPlot: Data Visualization for Statistics in Social Science. doi:10.5281/zenodo.1308157. R package ver. 2.8.7, https://CRAN.R-project.org/package=sjPlot.
51. **Brunson JC**, **Read QD.** ggalluvial: Alluvial Plots in "ggplot2" (Online). 2020. http://corybrunson.github.io/ggalluvial/.
52. **Wickham H.** *Create Elegant Data Visualisations Using the Grammar of Graphics* (Online). Springer-Verlag, 2021. https://ggplot2.tidyverse.org.
53. **Lira M**, **Gardner SM.** Leveraging multiple analytic frameworks to assess the stability of students' knowledge in physiology. *CBE Life Sci Educ* 19: ar3, 2020. doi:10.1187/cbe.18-08-0160.
54. **Redish EF.** Oersted Lecture 2013: How should we think about how our students think? *Am J Phys* 82: 537–551, 2014. doi:10.1119/1.4874260.
55. **Sherin BL**, **Krakowski M**, **Lee VR.** Some assembly required: how scientific explanations are constructed during clinical interviews. *J Res Sci Teach* 49: 166–198, 2012. doi:10.1002/tea.20455.
56. **Monsell S.** Task switching. *Trends Cogn Sci* 7: 134–140, 2003. doi:10.1016/S1364-6613(03)00028-7.
57. **Kiesel A**, **Steinhauser M**, **Wendt M**, **Falkenstein M**, **Jost K**, **Philipp AM**, **Koch I.** Control and interference in task switching—a review. *Psychol Bull* 136: 849–874, 2010. doi:10.1037/a0019842.
58. **Scott EE**, **Anderson CW**, **Mashood KK**, **Matz RL**, **Underwood SM**, **Sawtelle V.** Developing an analytical framework to characterize student reasoning about complex processes. *CBE-Life Sci Educ* 17: ar49, 2018. doi:10.1187/cbe.17-10-0225.
59. **Asai DJ.** Race matters. *Cell* 181: 754–757, 2020. doi:10.1016/j.cell.2020.03.044.