

# Leveraging Multiple Analytic Frameworks to Assess the Stability of Students' Knowledge in Physiology

Matthew Lira<sup>†</sup> and Stephanie M. Gardner<sup>\*\*</sup>

<sup>†</sup>Educational Psychology & Learning Sciences, Department of Psychological & Quantitative Foundations and DeLTA Center, University of Iowa, Iowa City, IA 52242; <sup>‡</sup>Department of Biological Sciences, Purdue University, West Lafayette, IN 47907

## ABSTRACT

When a student explains a biological phenomenon, does the answer reflect only the *product of retrieving knowledge* or does it also reflect a *dynamic process of constructing knowledge*? To gain insight into students' dynamic knowledge, we leveraged three analytic frameworks—structures–behaviors–functions (SBF), mental models (MM), and conceptual dynamics (CD). To assess the stability of student knowledge, we asked undergraduate students to explain the same physiological phenomenon three times—once verbally, once after drawing, and once after interpreting a diagram. The SBF analysis illustrated fine-grained dynamic knowledge between tasks. The MM analysis suggested global stability between tasks. The CD analysis demonstrated local instability within tasks. The first two analyses call attention to differences between students' knowledge about the parts of systems and their organization. The CD analysis, however, calls attention to similar learning mechanisms that operate differently vis-à-vis external representations. Students with different mental models deliberated *localization* or where to locate the structures and mechanisms that mediate physiological responses, but students made these deliberations during different tasks and arrived at different conclusions. These results demonstrate the utility of incorporating dynamic approaches to complement other analytic approaches and motivate future research agendas in biology education research.

## INTRODUCTION

When a student responds to a question on an exam, does the answer reflect only a *product of retrieving knowledge* or does it also reflect a *process of constructing knowledge*? We contend that the answer to this question matters for biology education research (BER), because treating knowledge as a product risks missing the dynamic mechanisms of people's thinking and learning (diSessa, 2002). Attending to biology students' dynamic knowledge affords us the opportunity to identify the conditions that engender students' productive thinking and disciplinary engagement.

Students' dynamic knowledge poses a problem for assessment. How do we know what our students know? To make any assessment, we must select a mode. But assessment modes are theory laden—a model of student learning in a domain should inform assessment design (Pellegrino, 2012). For example, if we assume that students either do or do not know a given fact, we might choose to measure students' knowledge once. This, in turn, may reinforce a static and binary conclusion. Other modes that invite iterative or dynamic assessment, however, encourage us to consider the variability in student knowledge and when and how students rely on variable knowledge across assessment contexts. Modes such as clinical interviews open windows into the instability of students' knowledge (Sherin *et al.*, 2012). In interviews and related contexts, students present conflicting ideas, change their minds, and express doubt. In short, assessment modes that capture *moment-to-moment* thinking support efforts to capture students' dynamic knowledge.

Julia Gouvea, *Monitoring Editor*

Submitted Aug 20, 2018; Revised Oct 18, 2019;

Accepted Oct 28, 2019

CBE Life Sci Educ March 1, 2020 19:ar3

DOI:10.1187/cbe.18-08-0160

\*Address correspondence to: Stephanie M. Gardner (sgardne@purdue.edu).

© 2020 M. Lira and S. M. Gardner. CBE—Life Sciences Education © 2020 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

We call attention to the mode of assessment, because part of the solution to the assessment problem involves aligning our modes of assessment with the core assumptions of our analytic frameworks. This point matters, because recent dialogue in BER highlights disagreement regarding how assessments of student knowledge (Maskiewicz and Lineback, 2013; Leonard *et al.*, 2014) and its dynamic properties (Gouvea and Simon, 2018) suggest different theoretical models of student knowledge. We aim to add to this conversation by illustrating the importance of coordinating particular data sources with particular analytic frameworks. Data sources differ in their capacity to capture the various facets of student knowledge. Therefore, we must align our analytic frameworks and data sources with the problems we seek to address in our research. Attention to this alignment supports investigators in gaining traction on research problems and extending, testing, or modifying theoretical claims regarding student knowledge. We intend to illustrate this point by applying multiple analytic frameworks—structures–behaviors–functions (SBF), mental models (MM), and conceptual dynamics (CD)—to a single qualitative data set.

### To Assess Student Knowledge, Frameworks and Data Sources Must Coordinate

When we assess a student's knowledge, we must observe some action taken by the student. Student knowledge consists of at least five observable properties: determinacy, coherence, context dependence, malleability, and variability (Scherr, 2007). "Determinacy" refers to truth value. "Coherence" refers to the internal agreement within student knowledge. Context dependence refers to the reliability or flexibility of knowledge (i.e., how students use the same knowledge across different representations, social interactions, and education settings). "Malleability" refers to the ease whereby students change knowledge in response to instruction. Finally, "variability" refers to stability.

Though distinct, these properties interrelate. For instance, if a student makes an inaccurate statement (determinacy), an instructor may point to different claims (variability) made previously by the student that conflict (coherence). The instructor may then introduce a curricular resource, such as a graph or instructional video, and assess whether and how the student's knowledge changes (malleability). Next, the instructor may assess whether the student continues to use the same knowledge repeatedly and under different conditions (context dependence). In such a case, the instructor may use a variety of assessment modes and frameworks. A savvy instructor might foster class discussion to generate a variety of knowledge at the group level and then leverage that diversity of knowledge to prompt individual students to search for coherence. Thus, the instructor might first frame learning as social activity but then later frame learning as an individual action.

This pedagogical example illustrates a point that extends to educational research, wherein additional concern must be given to the alignment between analytic frameworks and data sources. Although most frameworks possess the capacity to capture all five facets of knowledge, they do not do so with equal ease. If a researcher tacitly assumes that student knowledge is static, then he or she might assess students once and select an analytic framework that reinforces that stance. A single assessment may detect determinacy, but it will fail to detect other facets of knowledge, such as context dependence, malleability, and variability,

because the single measurement occasion prohibits varying contextual factors, lacks a preinstructional assessment, and collects one final response but no process. These limitations hamper researchers and educators from gaining insight into multiple dimensions of student knowledge—especially variability.

In the present investigation, we concern ourselves most with variability. When a student's ideas shift *spontaneously* or else with little prompting during talk, they demonstrate instability. The desire to detect shifts demands a data source that affords insight into how students construct knowledge from *moment to moment* as a student thinks and learns with curricular resources. Video and audio data derived from clinical interviews remain a common and useful source for meeting such goals (Derry *et al.*, 2010). Video and audio data from interviews capture students' dynamic knowledge, because they possess a time course that affords students the opportunity to construct and modify claims (Sherin *et al.*, 2012) and because an interview creates a history of dialogue that allows conflicting ideas (coherence) to emerge across tasks (context dependence) and change in response to instruction (malleability).

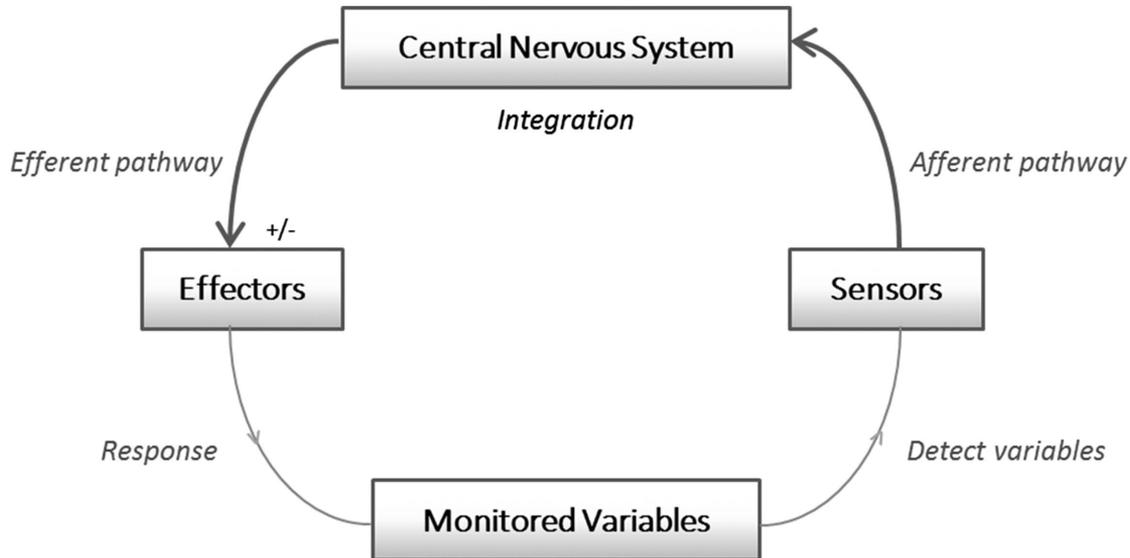
### Disciplinary Learning: Is That Your Final Answer?

Here, we introduce spinal reflexes as a disciplinary context for investigating the stability of students' knowledge. Reflexes are modeled as control loops (see Figure 1). Although they follow predictable sequences, they possess complexity due to their many interacting parts that span multiple levels of organization (e.g., molecules, cells, circuits), include multiple systems (e.g., nervous and muscular systems), and occur over different temporal scales (Michael *et al.*, 2017). The knee-jerk response can be elicited when a person receives a strike to the patellar tendon. After the strike occurs, the person's leg will "jerk" (i.e., kick).

The overall reflex can be modeled as consisting of three modules: detection and afferent relay, central nervous system integration, and efferent relay and response (Silverthorn, 2019). Critical to the canonical model sought by instructors, an explanation includes the organization of neural circuitry located in the spinal cord. The signals should be propagated to both the quadriceps to excite it and the hamstrings to inhibit it—this mechanism is crucial, because it coordinates the musculature to balance and protect the person. In sum, the complete explanation includes how the initial relaxed state and the final kick response are mediated by neurons, their connections, and their interactions.

To illustrate the assessment problem in context, consider a physiology student who says, "Reflexes control movement." What does such a statement demonstrate about this student's knowledge? Because this statement underspecifies the mechanism, a physiology instructor might be left wondering whether the student refers to spinal reflexes and neural circuits, what "control" means, and how reflexes control movement. For example, the student's use of "control" could refer to voluntary or involuntary action, and depending on the case, the control could occur in different locations in the central nervous system.

A follow-up question provides not only an assessment opportunity but also a learning opportunity—the student may demonstrate dynamic knowledge by constructing a novel explanation that cues and reorganizes other knowledge. Imagine that we invite the student to elaborate. The student then says, "Reflexes just happen. We do not have conscious control over them so they are involuntary." By asking the student to elaborate upon



**FIGURE 1.** Schematic of systems-level control loops in physiology. Many systems-level physiological phenomena, like the common knee-jerk response, must be detected by sensory receptor cells, sent to integrating structures where many cells connect to each other, and then relayed to effector cells such as muscle fibers.

the first answer, we learned something about the variability of the students' knowledge that we would have otherwise missed with a single question. Moreover, without the prompting, the student might not have known to characterize volition in such contexts, but perhaps that has now changed. The change may prompt the reorganization of other knowledge stated previously or subsequently. Other concerns might motivate us to inquire further still, but by providing an opportunity to elaborate, we obtained evidence that the student associates reflexes with involuntary not voluntary actions.

Yet adding more assessment items does not solve the assessment problem. It highlights it. Is the final answer the best representation of this student's knowledge? That depends on the goals of the assessment. If a researcher aims to gain insight into what answers mean to students and how students evaluate the meaning and accuracy of their own answers, then our methods demand assessments that afford students opportunities to elaborate and revise. Gaining insight into how students present knowledge from moment to moment will enrich the current dialogue surrounding the future aims of BER in general (Dolan, 2015) and theorizing about the dynamic nature of student knowledge in particular (Gouvea and Simon, 2018). Because our present aim is to assess instability in student knowledge, we contend that there is no single "best" answer. We want to observe how students construct different varieties of knowledge, and that demands attention to the dynamics of thinking from moment to moment and across situations.

#### Multiple Frameworks in BER: Problems and Solutions

We began by discussing the assumption that student knowledge is dynamic, not static. This contention reflects one facet of a theoretical debate regarding the organization of student knowledge and mechanisms of conceptual change. In the history of the learning sciences (e.g., Smith *et al.*, 1994), and more recently in BER (Maskiewicz and Lineback, 2013; Leonard *et al.*, 2014), scholars have continued to address this theoretical prob-

lem by debating its intricacies. We chose to not recapitulate them here. Instead, we aim to illustrate how leveraging more than one framework brings such theoretical problems into relief.

Multiple frameworks exist because of the complexity of knowledge. Selecting one framework over the others means that some potentially valuable "signal" will become lost among the "noise." With the recent surge in BER (Lo *et al.*, 2019), reflecting upon the trade-offs between frameworks contributes to our field's intellectual growth as we cross-pollinate with other areas of education and social science. We may therefore borrow from fields that experienced similar growing pains. For example, one way the learning sciences have addressed the challenge of studying knowledge is by identifying strategies for coordinating multiple frameworks (diSessa *et al.*, 2016). These include strategies such as "competing," "macro- and microcomplementing," "influencing," "synergizing," and "fusing" (diSessa *et al.*, 2016, pp. 3–6). This list offers us a set of strategies for thinking about how multiple frameworks could generate innovative research programs that continue the reflective discourse established by Dolan (2015) and Lo *et al.* (2019) and the conversations began by Gouvea and Simon (2018), Leonard *et al.* (2014), and Maskiewicz and Lineback (2013) that concern student knowledge.

To begin to gain traction on the problem of assessing students' dynamic knowledge, we selected "microcomplement," because we aim to illustrate how different frameworks afford different and sometimes complementary insights into student knowledge.

#### HOW MULTIPLE FRAMEWORKS PROVIDE COMPLEMENTARY INSIGHT WHEN ASSESSING STUDENT KNOWLEDGE

We selected three frameworks—SBF, MM, and CD—that hold value for BER. We aim to provide an empirical case that illustrates how latent theoretical assumptions residing within different analytic frameworks may complement or constrain

**TABLE 1. Analytic frameworks' emphases and key characteristics**

	Structures–behaviors–functions	Mental models	Conceptual dynamics (knowledge in pieces)
Instructional entailments	Encouraging students to relate structures to functions via mechanisms within and across systems	Visibility; feedback; make predictions explicit and permit the world or simulations of it to tune knowledge	Bring prior or conflicting knowledge to explicit attention of learner, deliver experience, foster reflection and knowledge coordination
Primary data sources	Inscriptions, multiple-choice, verbal reports from interviews	Predictions, drawings, verbal reports from protocol analyses	Verbal reports from clinical interviews, dyad problem solving, classroom discourse
Key theoretical assumptions about knowledge	No specifications regarding structures, organization, or properties; structuralism	Mental representations; semantic networks; knowledge for use to predict/explain; functionalism	Heuristic; implicit; fragmented or elemental; genesis of knowledge structures is small (p-prims) and growth large (concepts); empiricism/constructivism
Guiding metaphor(s)	The computer metaphor or mind as symbol processor (sometimes tacit, i.e., <i>analytic</i> not a theoretical framework)	The computer metaphor or mind as symbol processor (sometimes serial or other times parallel and distributed)	The knowledge ecology or mind as a complex, dynamic, and adaptive system(s); the computer metaphor (historical underpinning)

claims regarding the stability of student knowledge and how it is organized and used (see Table 1).

### SBF Frameworks Emphasize Features of Disciplinary Content Knowledge

The SBF framework emerged from computer science (Vattam *et al.*, 2011), and thus it uses symbolic arguments to model student knowledge as data structures. Students' knowledge of a system is "binned" into one of the three a priori categories. "Structures" refer to physical entities or the things in a system. The "behaviors" refer to the mechanisms or the interactions between the entities. Finally, "functions" refer to the purpose of the structures and mechanisms (Hmelo-Silver *et al.*, 2007). Many biological systems lend themselves well to this framework. For example, structures (like the heart) engaging in behaviors (such as pumping blood) serve functions (like delivering nutrients to tissues).

The SBF framework points to certain properties of knowledge structures more readily than others. Propositional knowledge describes the general syntax modeled by this framework. It lends itself to declarative knowledge stated verbally or inscribed explicitly by students. Thus, one way that scholars have leveraged the SBF framework is as an analytic tool that provides a "representation" to decompose and characterize systems and student knowledge of said systems (Hmelo-Silver *et al.*, 2007). The framework supports researchers in cataloguing student knowledge by binning their statements or inscriptions within prespecified terminology determined by a canonical model. Its primary affordance therefore involves assessing the determinacy of student knowledge. This affordance for assessing the stability of student knowledge stems from the fact that the framework offers categories of a relatively small grain size (e.g., sometimes just one word). As a consequence, an iterative assessment offers opportunities to assess what features appear and vanish as students complete disciplinary tasks.

The SBF framework offers a representation for describing students' knowledge of biological systems. By specifying "structures, behaviors, and functions," the representation classifies small-grained declarative knowledge *explicitly*. This point does not hold for the MM framework.

### MM Frameworks Emphasize the Global and Relational Structure of Knowledge

Here a "model" refers to a mental representation of something (Johnson-Laird, 1980)—be it a system, phenomenon, or object. Mental models refer to knowledge structures larger and more systemic than single propositional statements. Moreover, all of the knowledge that comprises a model does not exist in explicit form (Norman, 1983). Therefore, we adopt a classic cognitive science perspective that does not assume the stability of mental models a priori (cf. Norman, 1983). We additionally concern ourselves with attending to characteristics of different varieties of students' models. By adopting these assumptions, we position our investigation to capture any context dependence or dynamic aspects of students' mental models. The classic perspective, similar to the SBF framework, is rooted in the computer metaphor and thus may tacitly assume static knowledge storage and simple retrieval of it. We, however, adopt the assumption that students assemble a mental model with a core structure but allow that they might shift to a new core structure (Sherin *et al.*, 2012).

The primary affordance of the MM framework, then, rests in the strategy to identify the core organization of a system that students represent in their mental models. This grants insight into internal coherence. For example, students often describe the circulatory system as a single path or loop between the heart and the tissues of the body. This contradicts the canonical model by ignoring the gas exchange that occurs in the lungs and the accompanying vasculature that supports it. Such models specify distinct core structural organizations that suggest distinct mental models (Chi, 2000). But if a student cues knowledge about the role of lungs, then he or she may spontaneously shift to a new model. Identifying the essential structural distinctions in students' mental models facilitates broad categorization. Whereas students might present great individual variability regarding knowledge of all of the various structures, behaviors, and functions that comprise a system, an MM analysis brings coherence to aggregate patterns in the sample population.

Yet this strength is also the Achilles' heel of the framework. Identifying broad patterns and labeling individual students as

holding a single mental model comes with the trade-off of losing insight into intra-individual variability from moment to moment as students construct knowledge. Part of the path to addressing this issue involves closer attention to the sequence of how students *use* knowledge. As we will see next, other frameworks attempt to model the temporal dynamics of student knowledge explicitly.

### CD Frameworks Emphasize the Instability of Implicit Knowledge in Use

The CD framework (Sherin *et al.*, 2012) originates from the knowledge-in-pieces (KiP) framework (diSessa, 1988). According to this account, the novice student's knowledge exists initially in a large collection of disconnected notions that only organize into coherent theory-like structures after years of developing disciplinary expertise. In the KiP account, the "pieces" that reorganize are called "knowledge fragments" or "elements." These elements initially represent the vast collection of implicit and intuitive knowledge that exists in a "flat" organization. Here "flat" means that it holds little hierarchical organization, nor is the organization based on theoretical principles. Multiple elements can account for a single phenomenon or a single element can account for many—contextual factors dominate how the knowledge system behaves.

The CD framework's origins in the KiP approach guide researchers to attend to moments of internal contradictions or conceptual conflict. In these moments, students' dynamic knowledge may be observed and leveraged for learning. Say a student claims that a ball is falling fast. What does "fast" mean and does it always mean the same thing to a student? If we present the student with a pendulum, the student might claim that the pendulum is also moving fast. But in physics, we set the goal to have students distinguish between, say, instantaneous velocity and frequency. By encouraging students to coordinate their intuitions about motion (i.e., "fast-ness") with external representations of motions (i.e., graphs), Parnafes (2007) illustrated how conceptual conflict can prompt students to seek coherence and learn. Identifying or even cultivating instability in knowledge provides a window into variability (i.e., students' dynamic knowledge). In such moments, we gain insight into potential mechanisms of learning, because in these moments different varieties of knowledge can compete for status.

For this reason, the CD approach complements the MM approach. Whereas the MM framework calls attention to categorizing static states of knowledge, the CD approach calls attention to the dynamic production and use of knowledge in relation to specific tasks. The KiP framework seeks to understand how different varieties of knowledge interact with one another and with resources such as external representations (e.g., diagrams). Thus, the CD approach provides analytic tools that do not assume prespecified categories of expert-sanctioned disciplinary knowledge. The framework offers utility, because it specifies characteristics of *student* knowledge and models the dynamics of knowledge use with and without resources such as external representations. These features of the framework transform the noise in qualitative data sets into the signal and thereby complement other frameworks that might otherwise miss dynamic knowledge.

### THE PRESENT INVESTIGATION

To accomplish our aims, we will present data from a one-on-one interview protocol—outside the classroom—that provides a context to evaluate the stability of knowledge in physiology students as they explain phenomena and construct and interpret disciplinary representations. Using each of the reviewed framework's core assumptions, affordances, and prior implementations (summarized in Table 1), we aim to illustrate how analyzing a data set with each framework provides insights into student knowledge, its organization, and its dynamic properties. In particular, we pose the broad guiding methodological question:

How do comparative analyses offer convergent or divergent results regarding the stability of student knowledge?

We provided students the opportunity to explain the knee-jerk reflex phenomenon verbally, then again verbally after constructing a diagram, and then once more after viewing an instructional diagram. Note that assessing the stability *within* students' explanations as isolated events across this sequence is a useful analytic strategy, but we point out that, because students experienced the design holistically, we must also attend to stability *between* explanations. Thus, the design operates as a system—one that supports our efforts to answer the following discipline-specific assessment questions:

- How do the SBF, MM, and CD frameworks afford insight into students' dynamic knowledge as used vis-à-vis external representations?
  - What features of the reflex arc do students attend to when explaining with and without different representational resources?
  - What are the characteristics of students' mental models for the knee-jerk response when explaining with and without different representational resources?
  - What conceptual dynamics do students present when explaining with and without different representational resources?

### METHODS

#### Population

Students included 10 undergraduate biology majors (three males; seven females) with a standing rank of junior (1/10) or senior (9/10) at a research-intensive midwestern university. All students were enrolled concurrently and recruited from the same course, Principles of Physiology, an upper-division (i.e., junior/senior) lecture with laboratory course for biology majors.

#### Investigation Context

The course from which students were recruited emphasized the cellular mechanisms responsible for physiological functions and adaptations. Students were recruited inside this course and via email after completing the first exam. The first exam covered neurophysiology and basic systems analysis as applicable to physiology (i.e., system detection, integration, and signaling). The exam was held in the fifth week of class, and thus interviews were conducted between the fifth and 15th weeks of the semester. These data were collected inside an interviewing laboratory. A digital camera was used to collect audio and video

data from students' speech and gesture. A smart pen (LiveScribe pen) was used to collect students' inscriptions and utterances when they explained using a drawing. All work for this study was done in accordance with an approved human studies protocol (Institutional Review Board 1510016662).

### Protocol

The semistructured interview consisted of three tasks: the explanation task, the drawing task, and the interpreting task (see Appendix A in the Supplemental Material). In the explanation task, students answered the primary question of the interview protocol, "How does the knee-jerk response happen?" Note that before the interviewer asked this question, students were provided with context with a few "warm-up" questions designed to elicit domain-specific content knowledge from physiology (e.g., "When I say the word 'mechanism' in the context of physiology, what comes to mind for you?"). See Lira and Gardner (2017) for details. Follow-up questions were asked at the discretion of the interviewer to clarify ambiguous meanings and support students in articulating their knowledge in a comprehensible explanation (e.g., "What do you mean by 'it' goes? What is 'it'?")

In the second task, the drawing task, students were provided the smart pen and dot matrix paper and were prompted to construct a drawing of how the knee-jerk response happens. They were informed that they could construct one or more diagrams at their discretion. Finally, students were informed that upon completing their drawing they would be given the opportunity to explain it. Follow-up questions were asked as in the explanation task.

In the third and final task, the interpretation task, students were provided with an instructional diagram in the physiology chapter from a common undergraduate biology textbook (Campbell *et al.*, 2008). Given the target population for this book (undergraduates) and the simplified nature of this diagram with respect to a variety of details (e.g., the number of cells for each element is far greater than is displayed) this is a reasonable illustration for the target performance for our study. The diagram illustrated how a knee-jerk response occurs (see Appendix A in the Supplemental Material). In specific terms, it illustrated the neural circuitry involved; the sensory receptor cell (a neuron) that detects the pressure from a hammer strike on the patellar tendon, the integration via electrochemical signaling to and from the spinal cord, and the production of the observed coordinated motor response of the leg as a consequence of excitation and inhibition of the leg's musculature. Students were instructed to take their time and review the dia-

gram and were told that whenever they were ready they could describe what they were seeing. The first author (M.L.) conducted all interviews, and interviews lasted no longer than 1 hour. Students were debriefed and compensated US\$20 for their time and travel.

### Transcription

For the overarching goal, we needed to ascertain all the parts of the systems and their relations as described by students. Therefore, we first transcribed all audio to text. Transcripts were constructed to preserve the students' meaning explicitly. As noted earlier, the interviewer asked students to clarify statements in the context of the interview itself. When ambiguities in speech were detected but not resolved during the interview, we used contextual cues to disambiguate. For example, if a student says, "The hammer hits you *here*," our transcription procedure involved examining the video for a complementary gesture to determine whether the student pointed to the patellar tendon when saying "here." Likewise, if a student used pronouns (e.g., "it" happens) or demonstratives (e.g., "that" signals), we attempted to locate the specific nouns in surrounding speech and replace the pronouns and demonstratives. In instances where this procedure was not possible, we left the students' speech as it occurred.

Earlier, we stressed the importance of attending to the sequence of speech production. Because speech production reflects the output of complex cognitive and motor processes, several features of speech reflect these complexities. For our purposes, however, the aim was to reconstruct the *content* of students' explanations for the knee-jerk response. This meant that we ignored many low-level speech production characteristics such as fillers (e.g., "um"), recasts ("So, I ... So, you ..."), and false starts (e.g., "S-s-uh, So, I ..."). If, however, the student made an explicit shift, then we captured this fact in the transcription. We will describe this analysis further in the *CD Analysis* section.

**SBF Analysis.** The SBF analysis was applied to all participant explanations and artifacts. We first leveraged our expertise in neuroscience and physiology to generate a set of anatomical structures necessary for explaining how the knee-jerk response occurs. Then, using the SBF analytic framework (Hmelo-Silver and Pfeffer, 2004), we specified the mechanism supported by the structure and the physiological function supported by both the structure and the mechanism. In sum, this strategy led to seven critical propositions needed to explain the knee-jerk response (see Table 2). For example, "synapses connect neurons

**TABLE 2. Coding scheme as aligned with the SBF analytic framework**

Structures	Behaviors	Functions
Patellar tendons	Connect muscle to bone	Provide a physical anchor for movement
Spindle fibers	Depolarize upon mechanical stimulation	Detect sensory stimuli
Sensory neurons	Propagate action potentials	Send afferent signals
Synapses	Connect neurons with target cells	Support intercellular communication
Interneurons	Inhibit action potentials <sup>a</sup>	Control circuitry and muscle tension
Motor neurons	Propagate action potentials	Excite muscle fibers
Muscle fibers	Contract	Move the skeletal system

<sup>a</sup>Not all interneurons send inhibitory signals.

TABLE 3. Analytical process for identifying students' mental models from explanations

Analytic process	Direct mechanism	Indirect mechanism
Identify exemplary excerpts of transcript	"The <i>stimulus</i> ... just sends a signal to all the <i>muscles</i> ... that <i>stimulus</i> is basically telling the <i>muscles</i> ."	"Something changes ... which is <i>detected</i> by a <i>specialized sensory apparatus</i> ... so then that goes to the <i>spine</i> where it is <i>transformed</i> , a motor output is created ... and then the <i>muscle contracts</i> ."
Generate a schematic based on production sequence	Stimulus → muscles	Change detected → transformed in spinal cord → muscle contracts
Determine critical characteristics	Student indicates that stimulus has its effect <i>directly</i> on a target—muscle in this case.	Student indicates that the stimulus is <i>detected</i> , <i>transformed</i> , and then effects a target—muscle again.

with target cells to support intercellular communication." To ensure that these propositions reflected a general consensus, as opposed to our personal instructional targets, we consulted textbooks in physiology and neuroscience (Campbell *et al.*, 2008; Silverthorn, 2019).

We used the table as a framework to analyze students' explanations, illustrations, and interpretations. Note, however, that we did not use literal language to detect the presence of any of the features. For example, say a student stated, "So, then your *muscles* in your leg *squeeze* to make your leg *kick* up." This student's explanation would have been coded as including the features "muscle fiber," "contract," and "to move skeletal system," respectively.

All relevant excerpts from the transcript were first identified and compiled by M.L. Excerpts were bound first by the initial question and then when a new line of questioning was initiated. Next, using a data corpus from a related pilot study, M.L. trained a second coder who was unaware of the project goals. After the training, the two coders used a random sample of 30% of the data from the corpus to assess interrater reliability. To accomplish this, the two coders (M.L. and a third-party analyst, C.R.) read and reviewed each excerpt line by line (i.e., every time the student started a new sentence to respond to the interviewer's question or comment). Upon reading each portion of the excerpt, each researcher would state either "yes" or "no" to indicate if the statement contained a portion of text that could be coded within the existing framework. If either of the two researchers said "no," that portion of the excerpt was abandoned. If both researchers said "yes," that portion of the excerpt was tagged and retained to be analyzed. Next, the two researchers, in isolation, applied the 21 codes from the list within Table 2. Students' drawings were analyzed after their verbal explanations to detect whether any additional features manifested—the analysis proceeded otherwise the same. This process led to high reliability (Cohen's kappa) for coding SBF in students' responses ( $\kappa = 0.95$ ). For rigor, all disagreements between researchers were not resolved but were instead excluded from further analysis.

**MM Analysis.** The MM analysis was applied to all participants. After we transcribed the audio data, we then read through the transcripts corresponding to the *first* task in order to assess whether or not students constructed the canonical model. By contrasting students' explanations with standard textbook explanations, we determined that no students produced our a priori canonical model. A constant comparative method (Glaser, 1965) revealed that a discriminating characteristic in our data corpus involved whether or not students included in their initial explanations mention of *any* neural circuits responsible for the *mechanisms that mediate* the knee-jerk response. To

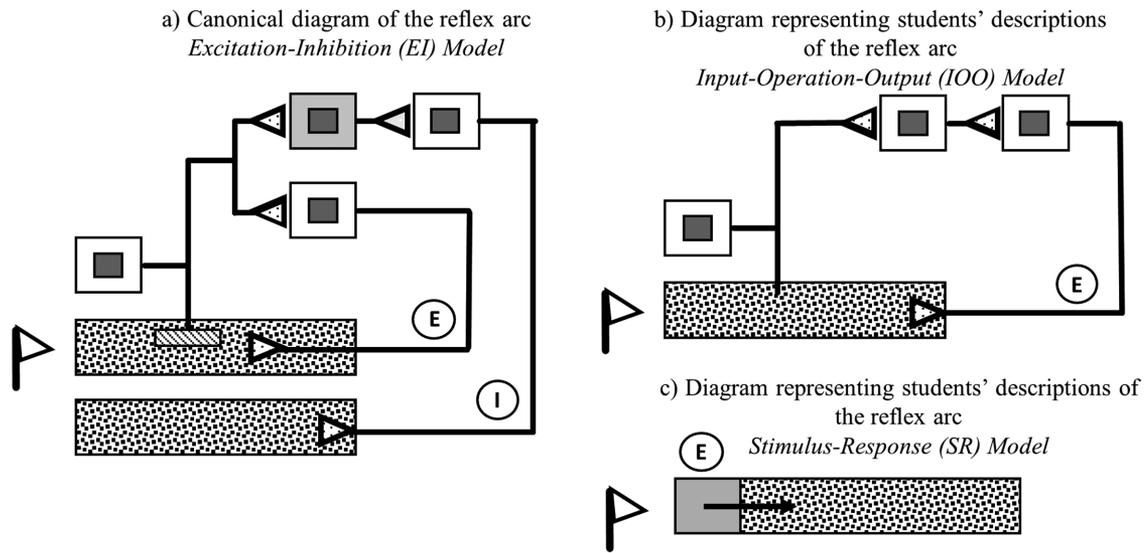
operationalize this discriminating characteristic, we leveraged our complementary SBF analysis to determine the presence or absence of neural circuitry. Two mental models emerged from this analysis. The first is a direct mechanism whereby a stimulus causes the muscles to contract or the leg to kick. The second is the indirect mechanism whereby the stimulus is transduced or transformed before causing muscle contraction or kicking. To arrive at this conclusion, we re-represented students' explanations into linear schematics (see Table 3).

The MM analysis then involved 1) assessing whether this core structure—the mechanism of neural circuitry—was adopted by students; or 2) if it was not, identifying what features were adopted, abandoned, or reorganized across the two subsequent tasks.

**CD Analysis.** Sherin *et al.* (2012) describe the general methodological approach we adopted. In particular, we aimed to observe and analyze students' reasoning *as it unfolds* as opposed to the endpoint of students' reasoning. This means we sought moments of the interview when students deliberated competing ideas. To accomplish this aim, we leveraged our complementary MM analysis. That is, after classifying students with one of the two MM groups, we sought out one exemplar from each class to contrast.

From this point, in essence, we leveraged a constant comparative method (Glaser, 1965) to identify when students verbally indicated a conceptual challenge or when two ideas appeared in conflict or one idea failed to satisfy the student as a plausible explanation or an accurate claim (for related analyses, see Parnafes, 2007). For instance, we examined transcripts for moments when students expressed language such as, "wait, hold on, no that's not right, or maybe it doesn't, I think but I am not sure," and so forth. We then reviewed before and after such moments to ascertain whether or not the student shifted the explanation by adding, removing, or modifying features. In our domain, examples included "No, an action potential does not go to the brain," or "Wait, yes, I think the spinal cord is involved." With respect to the cases we will present, we report such moments of conceptual challenge comprehensively to illustrate how two students presented their knowledge tentatively and dynamically through the sequence of tasks.

We selected our two representative cases by attending first to their relative scientific accuracy as determined by the SBF analysis—these two students stood in stark contrast. We then compared them to their MM grouping to assess whether they presented any aberrant characteristics that deemed them non-representative of their respective groups. They did not. Finally, both students were sufficiently articulate to serve as cases worthy of contrast.



**FIGURE 2.** Canonical and student models of the knee-jerk reflex. Schematics illustrating the critical features of neural circuitry in (a) the accurate excitation–inhibition (EI) model and the two student mental models: (b) IOO ( $n = 5$ ) and (c) SR ( $n = 5$ ). The triangle at left in each schematic represents a mechanical stimulus (i.e., a hammer strike to the tendon). The rectangles represent effectors (i.e., muscle fibers) that are either excited or inhibited by the neurons. In the EI model, notice the excitatory neural pathway (labeled E) and the inhibitory neural pathway (labeled I). Despite mentioning the (inhibitory) interneuron in the middle of the circuit, the IOO model does not include this regulatory mechanism. The SR model represents no neural circuitry at all. A mechanical strike excites and causes the jerk if it is of sufficient intensity.

**RESULTS**

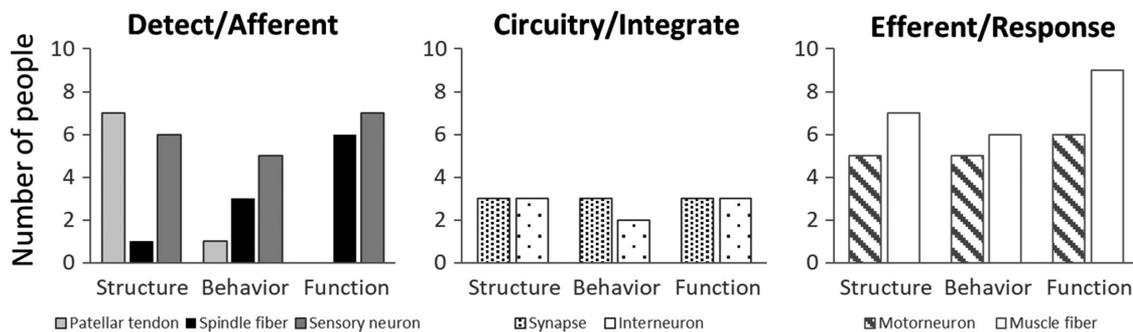
The three analytic frameworks complemented one other by illustrating different properties of student knowledge, its organization, and potential mechanisms of thinking and learning. These insights were gained predominantly by analyzing when and how students incorporated features that mediate stimulus-response relations and where students locate those features during talk and inscription. We next report this process in detail.

**Students’ Initial Explanations for the Knee-Jerk Reflex**

Of the three analytic frames, we begin with the SBF analysis to identify the structures, mechanisms (i.e., behaviors), and functions students presented in their initial explanations for the knee-jerk reflex. This lens enabled us to detect the specific features that distinguished students’ explanations and infer different mental model varieties. To illustrate, notice that a key feature of the canonical model involves the sensory neuron that

transduces the mechanical stimulus (see Figure 2a). Notice further that *this neuron propagates action potentials to send afferent signals*—this syntax reflects the syntax modeled by the SBF framework (see Table 2). By specifying all of the critical structures, mechanisms, and functions necessary to explain how a knee-jerk response occurs, we characterized the domain and generated a tool to define the canonical model necessary for assessing the completeness of students’ explanations, identifying the features students emphasized, and determining how these different emphases might shift dynamically across tasks. Because spinal reflexes consist of three basic modules, we bundled the features coded in Table 2 into each of their three respective modules. This allowed us to assess how students added or selected these features during the explaining, drawing, and interpreting tasks.

The SBF analysis conducted on all participant explanations of the knee-jerk reflex revealed a key pattern (Figure 3). Across



**FIGURE 3.** Frequencies of SBF features in the initial explanation task. Students demonstrated less attention to the features that mediated (circuitry/integrate) the knee-jerk response (e.g., synapses and interneurons) than to the structures at the site of the phenomenon (e.g., patellar tendon and muscle fiber). Most neglected among these features was the synaptic connectivity in the circuitry/integrate module.

the three modules of the spinal reflex, there were fewer features within the circuitry/integrate module. This finding suggests that students neglected structures and mechanisms that mediate the response.

With the overall aggregate pattern established, we then sought to identify types of explanations that suggested distinct mental models. Despite the fact that all students had been taught about this reflex in a common, previous class, students demonstrated pronounced differences regarding how they characterized neural circuitry—the critical organization and mechanism underpinning the phenomenon. These differences suggested distinct mental models. In the first class, students (five of 10) presented an explanation that emphasized the features that mediated the stimulus-response relation (indirect mental model; Table 3 and Figure 2b). The first class of explanations we characterized as input–operation–output models, or IOO. In contrast, we characterized the second class of explanations as stimulus–response models, or SR. These students (five of 10) presented an explanation that emphasized the stimulus and the response but not the features (i.e., structures and mechanisms) that mediated the relations between the two (direct mental model; Table 3 and Figure 2c). In addition to including more SBF features generally, the IOO group of students all specified critical neural circuitry, whereas the SR group did not.

To illustrate how SR students neglected mediating features differently from the IOO students, attend to Figure 4. Notice the higher frequencies of the structures sensory neurons, interneurons, motor neurons, and synapses that the IOO students (top) present relative to the SR students (bottom). The same general pattern is found in how students incorporate the behaviors and functions of these structures.

In sum, the aggregated patterns detected by the SBF and MM analyses of students' *initial* statements suggested that they emphasized different features of the reflex arc and constructed

two distinct mental models to explain the knee-jerk response. Both models differed fundamentally from the canonical model, because they neglected the critical role of inhibition. To assess whether students' explanations reflect stable knowledge, we will next analyze how students present knowledge across all tasks. We then conclude our analyses by presenting the results of a comparative case analysis aimed at assessing the moment-to-moment sequence of how representative students use knowledge both between and within the tasks.

### SBF and MM Analyses: Aggregate Data Patterns

In analyzing the aggregate data, we were able to identify and define the attributes of two structurally distinct mental models (i.e., SR/IOO). All subsequent analyses revealed that our coding scheme for students' explanations did not detect any critical changes to the structural organization of the mental models students presented across the remaining tasks.

The different features of students' explanations, however, extended beyond the SR/IOO scheme. We wanted to gain further insight into what specific SBF features might raise conceptual challenges and how such challenges manifested in the aggregate. To make this assessment, we determined each group's initial performance and then assessed the change across tasks (see Figure 5). Because the critical distinction between the SR and IOO models involved specifying either a direct or an indirect mechanism, we attended to when students incorporated features of neural circuitry (e.g., sensory neurons, interneurons, motor neurons, and synapses).

The SBF analysis illustrated that, in the initial explanation task, the IOO students (Figure 5, top) tended to remain stable, or else decreased relative to their performance earlier in task sequence. In contrast, the SR students (Figure 5, bottom) present a different story. Although students with the SR model neglected mediating features more frequently than the students

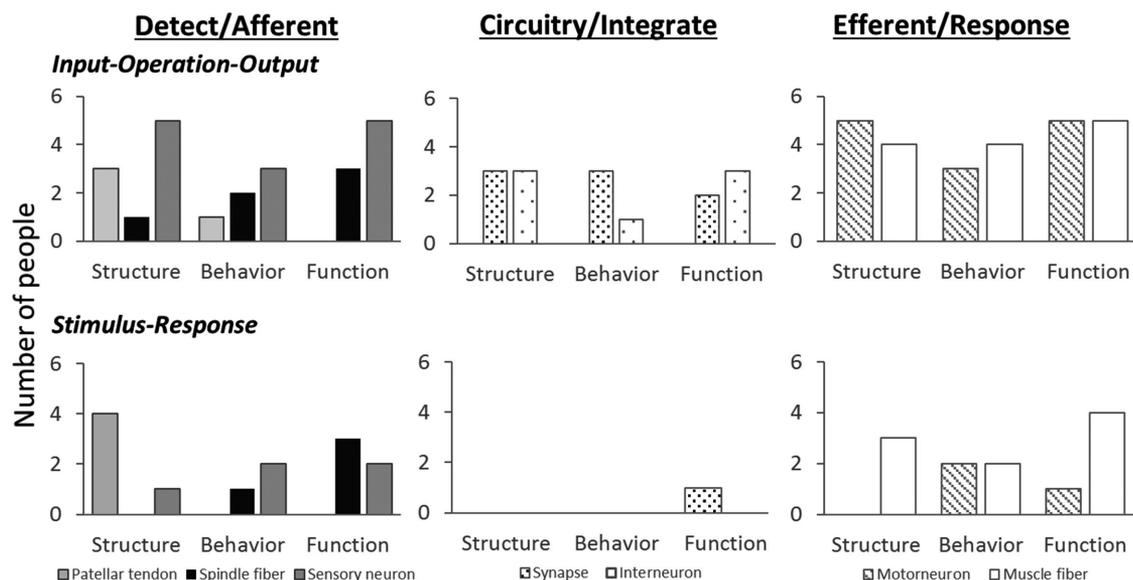
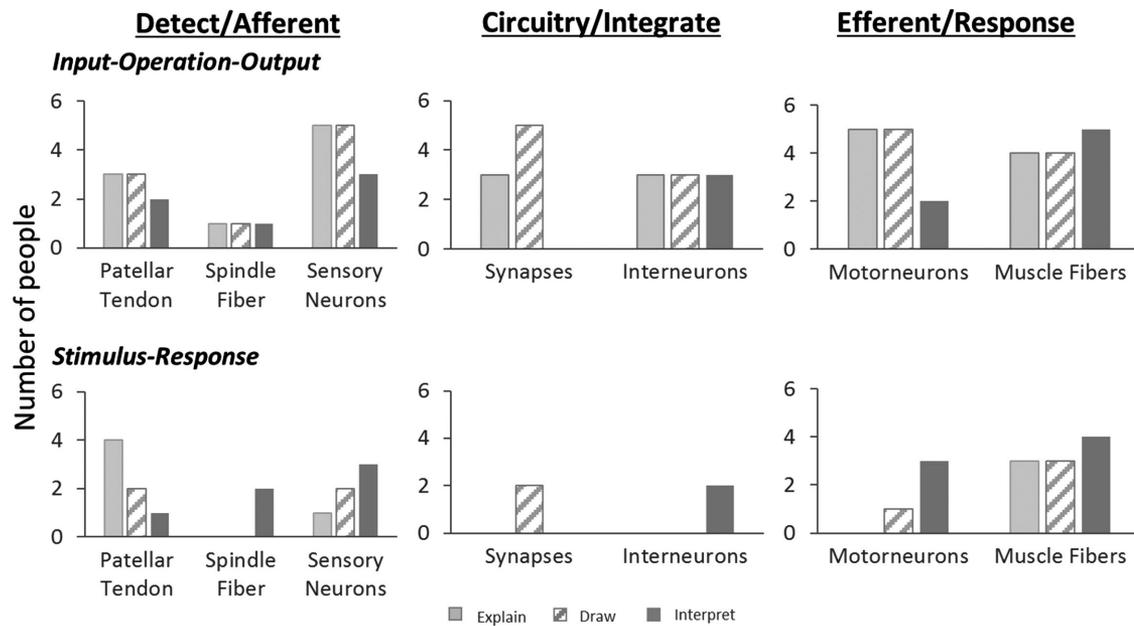


FIGURE 4. SBF features in students' explanations by MM type. The analysis of the SBF features presented in students' initial explanations illustrates the convergent result that IOO students (top) incorporate more SBF features that mediate the knee-jerk response than do SR students (bottom). This finding is particularly salient regarding the circuitry/integrate features (i.e., mediating mechanisms).



**FIGURE 5.** Dynamic knowledge of structures in students’ explanations across interview tasks. Our analysis with the SBF framework illustrates the convergent result that IOO students’ performance changes differently from SR students in that IOO students (top) show fewer novel features in later tasks than do SR students (bottom). This finding is particularly salient regarding the circuitry/integrate structural features (i.e., mediating structures). Notice, for example, how all five IOO students indicate synapses after drawing but none do so when interpreting the diagram. In contrast, no students indicate interneurons until the interpretation task, when two students do so.

with the IOO model, the students added these features as they progressed through the tasks. In contrast, IOO students neglected features they mentioned previously.

Together, this pattern of results hints that students with different mental models present different conceptual dynamics when reasoning vis-à-vis external representations. We next sought to determine whether moment-to-moment patterns at the individual level would further enrich the picture thus far developed at the aggregate level.

**MM and CD Analyses: Two Contrasting Cases**

We will start by using two cases to illustrate, in more detail, the characteristics of student explanations for the knee-jerk reflex. We proceed from students’ mental models as they manifested in the explanation task to student explanations of the same phenomenon during the subsequent tasks (drawing and diagram interpretation). This analysis allows us to evaluate the stability of student knowledge across as well as within tasks. We will then extend our attention from the individual case analyses back to trends in the aggregate data.

**Explaining without Student-Generated Drawings or Instructional Diagrams**

One student, Cindy, provides an exemplary case that illustrates how some students articulate the structures and mechanisms that mediate physiological functions (Table 2 and Figure 2b). She illustrates the IOO mental model. The four other students who produced explanations comparable to that of Cindy’s also 1) specified neural circuitry, 2) located specialized cells or neurons in the integrating center of the spinal cord, and 3) neglected any structure or mechanism responsible for inhibiting the hamstring’s excitatory motor neuron.

In the excerpt that follows, Cindy answers the central explanation task posed to all students in our interview: “How does a knee-jerk response happen?” We call attention to the characteristics of her explanation that will differ from those of the other group of students (SR) who were posed the same question. Like all students in both groups, her explanation starts with the stimulus and unfolds toward the response, but unlike students in the SR group, Cindy describes the specific neural circuitry and signaling mechanisms that mediate the knee jerk response.

*Cindy:* So, the knee-jerk response happens when something hits your patellar tendon and that excites ... the sensory neuron to create an action potential that travels up to your spinal cord where it synapses with an interneuron and then it will synapse with a motor neuron and then an action potential will tell your muscle, at whatever end plate at that end of that muscle fiber, to move according to the reflex making your knee go like that [extends right leg].

Notice that Cindy specified the cellular structures that support the mechanisms responsible for detecting the stimulus (sensory neurons), integrating the information (interneurons), and communicating the information to elicit an appropriate response (motor neurons). Moreover, she located the physical position of the integrating center when she mentioned the spinal cord.

Cindy, like all students in both groups, nevertheless neglected to mention that the interneuron inhibits a second, different motor neuron that excites the antagonistic hamstring muscle—recall that this feature is important, because relaxing the hamstring serves a primary function of coordinating muscle tone to prevent tearing and to stabilize balance to protect the organism

from falling and suffering injury. Despite lacking this feature, Cindy nevertheless adopted a cellular-level explanation that is almost accurate, because it includes neural circuitry (but not excitation *and* inhibition). She delivered the explanation in full and with no further intervention from the interviewer.

In contrast, when posed the same question, notice that a different student, Maggie, fails to articulate the structures and mechanisms that mediate the physiological function of the knee-jerk response. Maggie describes the stimulus and the response, but we observe no evidence of her describing the specific neural circuitry that mediates the response. Moreover, unlike Cindy, notice that Maggie encounters a conceptual challenge when she constructs her explanation—she indicates this by posing a question out loud (i.e., “Does it?”). This challenge will result in further questions posed by the interviewer within the same explanation task.

*Maggie:* OK, so like if you hit the, I do not know what it [the patellar tendon] is called, but if you hit it, your body detects a signal. It goes up into your brain and then an action potential is fired and then [a] kick back [kicks her leg]. Except I do not think that it goes into your brain. Does it?

Maggie's explanation illustrates a mental model different from Cindy's. Because the characteristics of this explanation differ so far from the accurate model, we contend that this excerpt illustrates a different type of explanation—one based on the SR mental model—that does not specify the organization of cellular structures and mechanisms that mediate the knee-jerk response (Table 2 and Figure 2c). The four other students who produced explanations comparable to that of Maggie's also 1) emphasized gross anatomical structures such as the body, leg, or brain as opposed to neural circuitry; 2) located the integrating center in the body, leg or brain as opposed to the spinal cord<sup>1</sup>; and 3) like the IOO group, neglected any structure or mechanism responsible for inhibiting the hamstring's excitatory motor neuron.

For these reasons, we interpret Maggie's explanation as failing to demonstrate a clear operation or the mediating mechanisms based on neural circuitry. Her explanation possesses emerging features because she specifies some mechanisms (e.g., detection and action potentials), but because she does not coordinate these mechanisms with specific neural circuitry, her explanation does not offer the defining characteristics needed to qualify for the IOO model.

Maggie differed from Cindy in other ways relevant to conceptual dynamics. Notice that, in this very first task, Maggie indicates a conceptual challenge. At the end of her explanation, when she noted that she does not think that “it” goes into the brain,” she asks the interviewer explicitly if “it” does. In contrast to Cindy, Maggie appears to elicit support from the interviewer to complete her first explanation. Because we designed our protocol to afford such opportunities—per the affordances of the cognitive clinical interview (Ginsburg, 1997)—we invited stu-

dents to demonstrate their reasoning as shown in the following excerpt.

*Interviewer:* Well go ahead and talk me through that. So let us consider two options. If it goes into your brain what does that mean in relation to what you said before?

*Maggie:* It would still be subconscious because with that reflex you do not make that decision but I do not think that that reflex goes into your brain. I think that it is just your muscles stretching.

*Interviewer:* If it does not go into your brain then where does it go?

*Maggie:* Your muscles [points to quadriceps]. It is just your muscle reacting but like a fight-or-flight response that would be more subconscious.

In this subsequent exchange, Maggie stabilizes upon an explanation with characteristics that now further embody the SR model. Maggie's explanation suggests that she does not—at this moment—consider the neural circuitry that mediates the knee-jerk response. During this excerpt, she excluded action potentials, neurons, and the central nervous system. She located the entire phenomenon and mechanism within the leg, and she therefore situated any semblance of integration via neural circuitry (i.e., the operation) as occurring at a site proximal to both the stimulus and the response as opposed to the more distal, and correct, spinal cord.

To clarify, we do not defend a literalist interpretation. When Maggie notes that she does not think “that ‘that reflex’ goes into your brain,” she could mean information or action potentials signal throughout the central or the peripheral nervous system's neurons to excite the muscles. The point is that we do not know. By using the same criteria to judge the explicit characteristics of Cindy's and Maggie's explanations, we observe stark differences in their mental models and conceptual challenges. Maggie here suggests that the stimulus produces the knee-jerk response by (somehow) inducing the quadriceps muscle to contract *directly* as opposed to the contraction being mediated by signaling via a neural circuit. Maggie's conceptual challenge (i.e., when she asks if information gets transmitted to the brain or to the leg) further underscores critical differences between Cindy and her. Whereas Maggie wavered when localizing where the “information” traveled, Cindy located the integrating center in the spinal cord, and she did not waver.

These two excerpts provide two different mental models to compare with students' subsequent performance in the drawing task. Cindy illustrated a stable explanation consistent with the IOO model. Maggie illustrated an unstable explanation consistent with the SR model. Maggie resolved her conceptual conflict (incorrectly) when she committed to the idea that information travels to the leg and not the central nervous system.

### Explaining with Student-Generated Drawings but without Instructional Diagrams

People may underspecify spatial relations when speaking (Ehrlich and Johnson-Laird, 1982). Because drawing affords a permanent trace for reflection and holds the computational

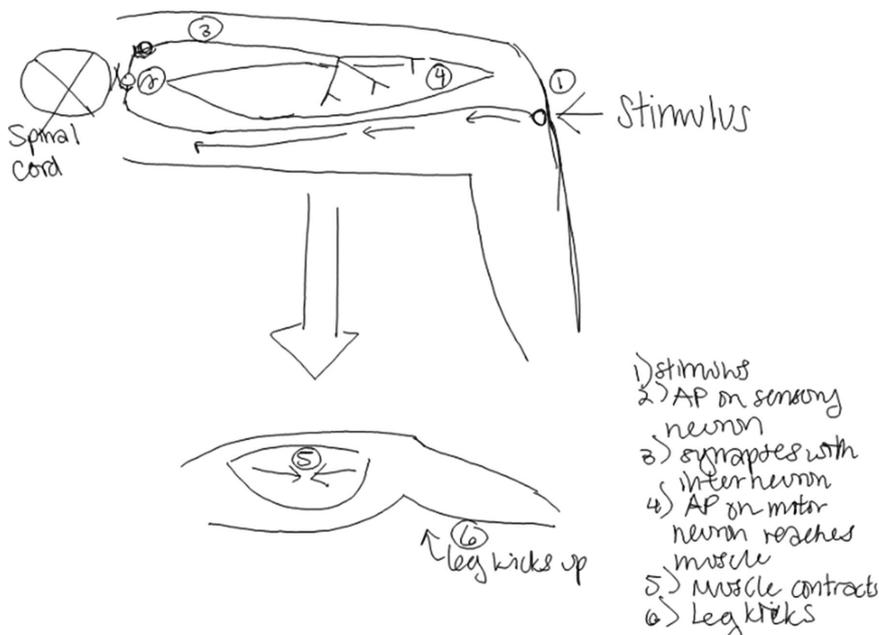
<sup>1</sup>Maggie almost immediately contested her own idea of the brain's involvement in the reflex arc. Conceptual dynamics turns this noise into the signal. In more concrete terms, her first response matters, because this knowledge cued *before* the conflict emerged. Temporal dynamics matter! This is an example of attending to moment-to-moment thinking and its messiness (i.e., dynamics).

property of necessitating specificity (Larkin and Simon, 1987), students should, in part, focus their explanations on localizing physical structures and mechanisms to illustrate the core relational structure of the system (Quillin and Thomas, 2015). In general, then, we aimed to identify further conceptual challenges when students specified neural circuitry in their drawings and assess whether students shifted to a different class of mental model based on constructing their drawing.

When prompted to construct a drawing that explains how the knee-jerk reflex happens and to explain the drawing, Cindy maintained an explanation with the general characteristics of the IOO model both as represented by her drawing (see Figure 6) and her talk after she drew. She did so by depicting different neuron types, their relations, and the direction of information flow from the stimulus to the response through mediating structures and mechanisms (i.e., a neural circuit). She describes this sequence after she completes her drawing:

*Cindy:* So, the reflex is started by a stimulus hitting the patellar tendon, which is right here below the knee cap, and then the signal travels along the sensory neuron to its interneuron which, with my understanding, it is not in the spinal cord, and then it goes to the motor neuron and it travels to and hits the muscle where it synapses with the motor end-plates and then it excites the muscle to contract and the muscle contracts and the leg kicks.

Using the same coding scheme for MM analysis, we did not detect sufficient change in Cindy’s explanation to warrant the claim that the drawing task alone elicited a different mental model.



**FIGURE 6.** Example drawing from an IOO model student. Cindy’s drawing preserves the critical features of neural circuitry (indicated by lines that end with forks) first presented by her in the explanation task. Notice, however, that at the left of her drawing, she crosses out the spinal cord to relocate the circuitry in the leg. Her list on the right recapitulates both her drawing and talk in the explanation and drawing tasks.

The drawing task, however, did operate in other ways that illustrate students’ dynamic knowledge. In this excerpt, for example, Cindy provides evidence of a conceptual challenge related to locating the position of the interneuron (see Figure 6). Cindy calls explicit attention to the issue. Her drawing conflicts with some characteristics of her mental model that we described in the previous section. Cindy noted in her initial explanation that “an action potential ... travels up to ‘your’ spinal cord where ‘it’ synapses with an interneuron.” In contrast, during drawing, Cindy repositions the interneuron in the leg, not the spinal cord. She also states that the interneuron is “not in the spinal cord.”

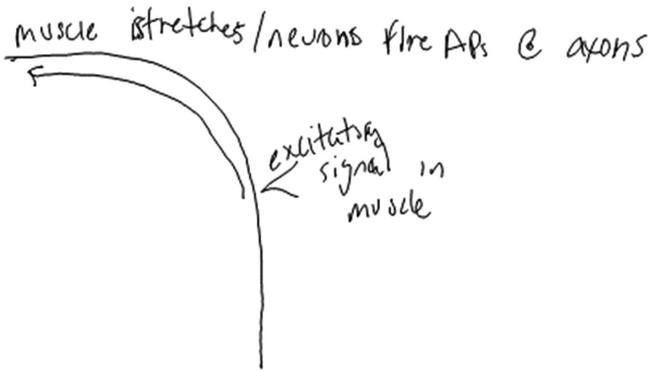
Note that this characteristic assumes a relation to the SR model, because it (incorrectly) locates the integrating center at the site of the stimulus and response (i.e., the leg). But because Cindy’s explanation maintains a commitment to mediating neural circuitry and mechanisms, the explanation remains categorized as IOO and not SR.

Although the critical characteristics of Cindy’s explanation did not cause her to fall into the SR category, they also did not motivate us to generate a new mental model class, because her drawing and talk remained otherwise consistent with her first explanation. Cindy again neglected to indicate that the interneuron inhibited the motor neuron that excites the hamstring—she neglected the hamstring entirely. Like her earlier explanation in the absence of a drawing, Cindy again specifies a sequence of structures and mechanisms that mediate the stimulus and response. Cindy notes the consistency explicitly (i.e., she says, “the same thing I just said.”) when asked.

*Interviewer:* OK, and you have a list over here [next to your diagram]...

*Cindy:* Yes, which is just explaining this [entire diagram], the same thing I just said. So the stimulus happens. The action potential goes down the sensory neuron. Then it synapses with ... the interneuron, then the motor neuron, then the muscle contracts and the leg kicks.

Cindy’s conceptual challenge during the drawing task makes Cindy’s performance appear less accurate, not more. From Cindy’s drawing, we observe evidence of unstable patterns of knowledge use. Cindy first drew a cross-section of a spinal cord as labeled on the left of her drawing. She demonstrated instability when she subsequently crossed out the spinal cord as indicated with the “X” mark. She completed her drawing by repositioning the interneuron in the quadriceps muscle. As noted earlier, her speech confirms this interpretation (i.e., “not in the spinal cord”). Thus, for Cindy, we see the drawing task operating in a way that illustrates instability in her knowledge. The drawing task enabled us to detect different varieties



**FIGURE 7.** Example drawing from an SR model student. Maggie's drawing preserves the critical features of the SR model she first presented in the explanation task. Notice, however, that above her drawing, she now indicates that muscles stretch and neurons fire in the leg. Yet we cannot detect the sequence from her drawing and talk. More importantly, we cannot discern any neural circuitry.

of reasoning either not active for her or else not shared with the interviewer during the first task.

Note further that what we have characterized as Cindy's unstable explanation in the drawing task assumed characteristics similar to Maggie's unstable explanation as described in the previous section. Earlier, in the absence of a drawing, Maggie demonstrated a conceptual challenge when she deliberated whether information traveled to the brain or the leg. From the drawing task, we now observe a new characteristic that relates Cindy's and Maggie's knowledge. Although Cindy maintained the critical characteristics of the IOO model, she too identified a similar conceptual challenge when she attempted to locate the position where the nervous system integrates information—the spinal cord or the leg in her case. Cindy's shift toward characteristics similar to the SR model suggests that students with different mental models can share similar conceptual challenges and dynamics when using their knowledge.

In contrast to Cindy's explanation, Maggie's explanation stabilizes during and after drawing. Thus, again, we observe a case in which the student remained categorized in same class—SR for Maggie. Recall, however, that Cindy and Maggie differed in how they incorporated neural circuitry. Maggie remains committed to locating structures and mechanisms in the leg exclusively, and she never notes any neural *circuitry*—though she does mention neurons. Moreover, she indicates no conceptual challenges regarding this matter. She abandons the brain altogether and never mentions the spinal cord when she explains her drawing (see Figure 7):

*Maggie:* OK, so this is a muscle and the excitatory signal when something hits right there and then the action potential propagates up the muscle stimulating other neurons and action potentials firing at those axons and then the muscle contracts and stretches.

Maggie remains categorized in the same group, because she maintained the critical features of the SR model. First, she did not specify a core structure that qualifies as a circuit. Second, when Maggie emphasized the leg's "muscle," she deemphasized the brain, but in doing so, she again neglected the spinal cord

or any other integrating center in the central nervous system. Finally, recall that for students who presented an explanation consistent with the SR model, inhibiting the hamstring remained an even less salient feature than for the IOO students, because they did not include an (inhibitory) interneuron. It is not surprising that Maggie neglected to indicate that an interneuron inhibits the hamstring's excitatory motor neuron.

We illustrated with Cindy's case that a student with an IOO model can show signs of conceptual challenges during drawing. Cindy did so by deliberating the location of neural circuitry. Although this deliberation reflects a productive feature of mechanistic reasoning (van Mil *et al.*, 2011), Cindy did not arrive at an accurate conclusion. In Maggie's case, however, we observed no evidence of conceptual dynamics. Maggie did mention neurons, whereas she did not do so in the absence of a drawing. This finding suggests that she continues to cue other knowledge, but because she did not specify a core structure that could be deemed a neural circuit, her explanation also does not provide evidence of a mental model shift.

Cindy's and Maggie's explanations suggested different mental models that possessed different features and organization. These different characteristics corresponded with them presenting conceptual dynamics differently across the first two tasks. Yet, in both cases, we observed students, who after drawing, stabilized upon explanations that moved them further from the scientifically accurate model they were taught. This occurred as they attempted to localize structures and mechanisms. After drawing, Cindy called explicit attention to a conceptual challenge regarding the physical position of an interneuron responsible for integration. She placed it at an incorrect site—the leg. Maggie committed to an explanation further still from the canonical model when she located the entire phenomenon in the leg but with no neural circuitry. Moreover, she indicated no conceptual challenge during or after she drew.

### Revising Explanations with Student-Generated Drawings and Instructional Diagrams

This brings us to the final task, in which students interpreted an instructional diagram. Because diagrams afford explicit representation of structural relations (Ehrlich and Johnson-Laird, 1982; Larkin and Simon, 1987) and because Cindy and Maggie verbalized a conceptual challenge regarding localization, both students should notice and revise their errors regarding the location of the integrating center (i.e., the spinal cord). Maggie, in particular, should repair features of her explanation by relocating integration to the spinal cord (and brain).

We began the instructional diagram task by giving students a few moments to review the diagram (see Appendix A in the Supplemental Material), and then the interviewer asked them to "describe" what they were seeing when they were ready. Students were not asked explicitly to compare the diagram to their drawings. They were, however, reminded explicitly that they could review their drawings at their own discretion with the statement, "Here is your drawing if you wish to look back to it at any time." The instructional diagram was placed next to the student's drawing.

All students did review their drawings but in different ways and with varying degrees of explicit reference. It was in these moments that students revised aspects of their prior explanations, but students did not offer explanations in full during

these episodes. Therefore, we characterize the cognitive behavior of these episodes as *revising* explanations rather than explanations in their own right. These data therefore suggest varying degrees of such awareness based on variability in students' explicit verbal attention to differences between the instructional diagram and their earlier drawings and explanations.

Cindy and Maggie demonstrated two different forms of revision representative of their respective groups. Cindy revised her explanation explicitly by comparing the features of the instructional diagram to her own drawing. In more specific terms, Cindy indicates explicitly that she might be "wrong," and she resolves her earlier conceptual challenge regarding the involvement of the spinal cord. By providing explicit verbal attention to the inaccuracy of her drawing and explanation (e.g., "I ... left out some steps") and selecting the specific features of the diagram needed to revise her explanation, Cindy illustrates how the instructional diagram affords a window into students' conceptual dynamics.

*Interviewer:* Can you describe to me what you are seeing?

*Cindy:* I am seeing that I either left out some of the steps or I was a little bit wrong about some of the things that I remember. But mostly it was the same as what I drew except for that the spinal cord is actually involved and there is only an interneuron and the pathway involves the hamstrings ... actually both muscles [including the quadriceps] seem to be involved in the reflex.

Cindy added new features not mentioned in her earlier explanations, and she revised her earlier explanation specific to her conceptual challenge. In addition to relocating the interneuron to the spinal cord, she now also notes that "the pathway involves the hamstrings" and that it too is "involved" in the reflex. Recall that the role of the hamstrings was a chief concern to us, because we aimed for students to recognize the antagonistic relationship between it and the quadriceps muscles that supports movement around the knee joint. Further, we aimed for students to reason that an excitatory signal to the quadriceps *and* an inhibitory signal to the hamstring ensure proper mechanical functioning of the limb and protection of it and the organism. We did not observe any students present this knowledge.

Cindy's case illustrates how the diagram and the interviewer provide a context for her to shift her explanation closer to elements of the canonical model as represented in the diagram. Knowledge that was before inactive for Cindy now emerges when prompted by the interviewer. Whereas before Cindy did not mention the critical feature of the antagonistic relationship between the quadriceps and hamstrings, she does so now:

*Interviewer:* Could you elaborate upon how both muscles are involved?

*Cindy:* So the quadriceps would contract and hamstrings would extend.

*Interviewer:* By your own assessment, does that make sense?

*Cindy:* Yes.

*Interviewer:* OK, so how does that make sense?

*Cindy:* It makes sense because if the knee is bent then the hamstrings would originally be not contracted but just in a shorter kind of relaxed state so to extend them would make the leg long because it is in this [forward extending] direction and pulling this [quadriceps] backwards on this tendon would make the leg shift up like that [extending the leg].

In this exchange, Cindy and the interviewer did not reach the mechanisms of neural inhibition and relaxation of the muscle, but the conversation demonstrates fodder for such learning. Cindy selected and revised features of her prior explanations that underspecified critical mechanical relations. Cindy did not learn this material *de novo* via the instructional diagram. Instead, Cindy used the context of the interview and the affordances of the diagram to revise her prior underspecified explanation, and she mobilized her intuitions and prior learning in her physiology courses to dynamically assembled a new explanation. She accomplished this by attending to both her own drawing and explanation and the structural relations encoded in the diagram. The interviewer supported her by providing the context for her to reflect and elaborate.

In contrast to Cindy, Maggie revised her explanation implicitly by incorporating additional features of the diagram into her description. Maggie therefore uses the diagram differently from Cindy. Unlike Cindy, Maggie does not revise her earlier explanations or drawing—she does not indicate that she "might have been wrong" earlier. As a consequence, her description of the diagram does not lead to further dialogue with the interviewer. Instead, after the following exchange, the conversation evolves into the debrief phase of the interview.

*Interviewer:* Take a look at that [diagram] and whenever you are ready tell me what you are seeing.

*Maggie:* The stimulus hitting the patellar tendon and through the afferent neuron pathways it goes to the brain and then back in the efferent pathways the hamstring is, and quadriceps is stimulated and the kick occurs.

Although Maggie and Cindy differed in how they interpreted the diagram, we identified particular features of overlap between the two students. First, regarding the differences, Maggie's description of the diagram bears strong relation to her own earlier statements made during the explanation and drawing tasks. Maggie neglected the spinal cord as the integrating site, and she failed to reason about the antagonistic relationship between the quadriceps and hamstring muscles. Maggie instead suggested that both muscles contract. Thus, she differed from Cindy, because she did not relocate the integrating center to the spinal cord, she did not attend to antagonistic relations, and she did not explicitly correct her incorrect explanations and drawing from earlier tasks.

To assess the similarities between these students' conceptual dynamics, first recall that, in the explanation task, Maggie experienced a conceptual challenge regarding whether information traveled to the leg or brain. She concluded that it was the brain. Then, when she explained her drawing, she neglected to mention the brain. When she interpreted the diagram she reincorporated

the brain into her explanation. These conceptual dynamics parallel Cindy's relocation of the interneuron from the spinal cord to the leg and then back to the spinal cord. In addition, Maggie, like Cindy, noticed and mentioned the hamstring. Finally, Maggie, like Cindy, added new features to her description when she mentioned specific structures (e.g., afferent neuron and hamstring).

These results illustrate how students with ostensibly different mental models experience similar conceptual challenges and dynamics. But they did so at different times when reasoning with external representations. The diagram interpretation task illustrated how students' prior knowledge and interview history can draw their attention to different salient features represented. In particular, Maggie's case illustrated that a student with an SR model can reinforce rather than revise inaccurate understandings. Moreover, what revisions students do make might occur during different segments of the assessment sequence. Finally, such students may revise their explanations without explicit attention to their earlier inaccuracies. Thus, Cindy's case illustrates a more manifest (i.e., explicit) example of conceptual dynamics unfolding than does Maggie's case.

## DISCUSSION

The results we presented call attention to biology students' efforts to construct coherent knowledge when locating the structures and mechanisms that mediate physiological phenomena. The noise (i.e., conceptual challenges) students present during their efforts also call attention to dynamic aspects of knowledge construction during assessment tasks. We will therefore next discuss how these findings motivate assessments that extend further than determinacy (i.e., accuracy) to capture other dimensions of knowledge and how curricular resources (e.g., external representations) and multiple frameworks might assist in this effort.

### Students Presented Conceptual Dynamics When Localizing Structures

We found that students deviated from the canonical model for the knee-jerk response by neglecting mediating structures (e.g., interneurons or neural circuitry) more so than sensory or motor structures (e.g., sensory neurons or quadriceps). That students focused on the location of structures echoes prior work adopting the SBF approach. Students tend to learn structures more readily than functions or mechanisms (Hmelo-Silver and Pfeffer, 2004). We note, however, that this may be the consequence of instruction as opposed to how students organize knowledge—indeed, function-centered instruction improves students' learning in physiology better than structure-centered instruction (Liu and Hmelo-Silver, 2009).

But why? One explanation might be that people tend to construct function-centered mental models (Norman, 1983). We construct them by acting upon the world in goal-oriented ways. When tasked with learning volumes of decontextualized information, however, students sometimes cope by memorizing parts of systems divorced from their functions (Liu and Hmelo-Silver, 2009). This explanation resonates with classic cognitive science. For example, people remember lists of objects better when they couple them in meaningful relations (e.g., scissors cut paper; Bower and Clark, 1969). Students, however, are still developing such learning strategies.

Moreover, this explanation also aligns with KiP accounts, in that students memorize “slogans” or build up collections of

knowledge elements—knowledge of structures in this case—before these elements self-assemble into theory-like knowledge (diSessa, 1993). Cindy recalled many of the correct structures (e.g., the interneuron) but did not specify their mechanisms or functions (e.g., integration), because her mental model was not organized around the function of *controlling* the two muscles that exist in an antagonistic relationship (see Figure 2, a and b). That said, we caution against the inference that knowledge of structure is inherently either fragmented or meaningless. Maggie, for example, did not mention the correct mediating circuitry, but she nevertheless contemplated where to locate the destination of the “signal,” because she understood the reflex as a “subconscious” process. Thus, for any CD analysis, the initial state of a knowledge system constrains how students reason and experience conceptual challenges.

### Multiple Dynamic Assessments Locate Students' Dynamic Knowledge

By assessing students multiple times, we gained insight into the properties of students' knowledge—especially variability. Chief among these insights was that students demonstrated variable knowledge of localization, or where to position the structures and mechanisms that mediate the knee-jerk. Our three assessments permitted us to see that features of students' first answers do not remain stable. Cindy demonstrated variability during drawing when she relocated mediating features from the spinal cord to the leg. Although she retained the neural circuitry, her drawing and talk illustrated unstable knowledge regarding the circuitry's relation to the spinal cord. Because much of biology (van Mil *et al.*, 2011) and science learning in general (Russ *et al.*, 2008) involves learning to locate the position of structures and mechanism, this insight illustrates the value of assessing students multiple times when feasible.

We also gained insight into variability by attending to students' processes of constructing knowledge within the temporal bounds of each assessment task. Once again, when students deliberated localization, we observed dynamic knowledge. *While* drawing, Cindy located the mediating features in the spinal cord, crossed out that portion of her drawing, and then relocated the features to the leg—she noted her deliberation explicitly in her subsequent talk. Likewise, Maggie located the mediating features in the brain and then relocated them to the leg *within* her first explanation. This suggests that these students, upon deciding to abandon the central nervous system, determined that the leg served as the most reasonable locus for signal integration, because the reflex happens fast and without “conscious” control. If we had attended to one final answer and ignored the moment-to-moment knowledge construction, we would have failed to detect students' dynamic knowledge, because we would have not observed it, not noticed, or else cast aside these deliberations as noise and not signal. This point highlights the dynamic nature of student knowledge and underscores the fact that assessments provide a window into students' knowledge, but they do so through the vantage point of a framework.

### Multiple Analytic Frameworks Complement One Another

Regarding our specific analytic frameworks, we found that one broad answer to our guiding methodological question was that different frameworks complement one another by directing researchers' attention to different levels of data analysis. These

include the granularity of a data source and its temporal scale. The SBF analysis directed our attention to the presence or absence of features of the system, and these were sometimes marked by the inclusion or exclusion of a single word. The MM analysis directed our attention to the global structure of the system's organization as described by students. Finally, the CD analysis directed our attention to the moment-to-moment construction and shifts marked by changes in students' sentences and the ideas within and between them. Although using multiple frameworks will not be necessary or even desirable in most investigations, our analyses suggest that turning to one framework relieves us of the limits of the others. For example, when similarities between students' overall shifts or fine-grained terminology made CD or SBF comparative analyses arduous, our MM analysis enabled us to draw quick and easy distinctions.

This idea is a two-way street. The easy distinctions made via the MM analysis challenged us to consider continuities between students, as illustrated by the contrasting MM cases. We see this contribution as perhaps the most valuable to our community, because continuities in the conceptual dynamics between students with distinct mental models motivate researchers to revise simplistic models of static knowledge and attend to dynamic learning mechanisms (Smith *et al.*, 1994; Gouvea and Simon, 2018). Both IOO and SR students presented stable *and* unstable knowledge. Thus, we found evidence of variability in student knowledge. This argument must be tempered by noting a critical analytic distinction between *global* as opposed to *local* instability. The MM analysis suggested global stability, in that all students retained their mental models across all tasks—no students shifted dramatically by either including or excluding all neural circuitry. Yet the CD analysis suggested local instability when students deliberated localization. These findings support the idea that knowledge (i.e., long-term memory) itself may be stable, but knowledge *use* is dynamic, because it is poorly indexed (Redish, 2014).

This idea was described previously by Sherin and colleagues (2012) who introduced the phrase “dynamic mental constructs” (DMCs). In the context of the present results, students seemed to either know that neural circuits were responsible for spinal reflexes or else reasoned that tissue in the brain or leg was responsible. This knowledge corresponds to Sherin and colleagues' “modes” and “nodes”—relatively stable knowledge structures that are stored in people's long-term memory. Modes refer to a portion of interconnected nodes that exist in a knowledge system. Once reasoning with a given mode, a student might activate or inhibit any particular node or even shift to a new mode. When tasked with explaining phenomena, students assemble these modes and nodes into a DMC. This mental construct is time bound—it exists within a moment and it changes from moment to moment. This perspective entails the commitment that a student's first explanation is not “the” (only) explanation (Sherin *et al.*, 2012, p. 174). DMCs offer an alternative to stable mental models. For researchers interested in how students' mental models change, DMCs hold value.

### Multiple Frameworks and Assessment Modes Provide Different Insights into the Properties of Knowledge

We introduced three frameworks by noting that they encouraged researchers to attend to five different properties of student knowledge (Scherr, 2007). Here we focus on determinacy to

challenge it as the only goal of assessment. Applying the SBF and MM analyses to both talk and drawing alerted us to problems regarding determinacy, because these two modes highlight inaccuracy and incompleteness differently. For example, students' verbal statements were incorrect with respect to mediating features. When a student states that a signal goes to the brain or the leg, as opposed to the spinal cord, they are wrong for cases that involve spinal reflexes.

Students' drawings, however, were incomplete with respect to the system's organization (e.g., the mediating neural circuitry). Drawing tasks do demand specificity not demanded by verbal reporting (Ehrlich and Johnson-Laird, 1982), but students may fail to include the necessary features. Features specified explicitly in talk were easily assessed as accurate or not by the SBF analysis, but these features were not demanded in talk—a person may say a fork is beside a knife and fail to specify left–right relations. Likewise, a student may say neurons are “involved” without verbalizing where and how. In contrast, features specified in drawings were not easily assessed as accurate or not, but these features were demanded in drawing (Larkin and Simon, 1987). Thus, drawing tasks might encourage researchers to attend to students' internal coherence as opposed to determinacy (Quillin and Thomas, 2015). Comparative assessments of talk with and without drawing provide insight into the variability of student knowledge.

These points matter in relation to assessments of determinacy. People learn to behave according to context. For example, jokes, slang, relaxed body posture, and informal attire are not in poor taste until we understand their use in a given context. Likewise, the variety of heuristics that students use to explain biological phenomena might serve as useful resources depending on the context (Zohar and Ginossar, 1998; Hammer and Elby, 2003). We therefore echo Gouvea and Simon (2018) by suggesting that assessments and analyses might benefit from expanding the unit of analysis beyond the accuracy of the statements students make to include the statements *and* the contextual factors (e.g., features of the task) that elicit them. Our CD analysis around localization revealed the presence of different properties of student knowledge depending on the availability of representational resources. When students noticed problems regarding localization, they revealed awareness of the importance of spatial organization in biological systems and that they sought coherent knowledge. This finding resonates with the claim that people reveal knowledge when they identify “snags” in the context of problem solving (Lave *et al.*, 1984). Moreover, such assessments support efforts to identify learning mechanisms by documenting the conditions that promote students to seek coherence (Parnafes, 2007). Thus, when students fail to state scientifically accurate propositions, we may still gain insight into other properties of their knowledge.

### OPPORTUNITIES FOR EXTENDING THE STABILITY CONVERSATION: NOTICING, FRAMING, REPRESENTING, SEQUENCING, AND EVALUATING

Several limitations set boundaries for the scope of our claims. We will therefore conclude by discussing opportunities for continuing the conversation on students' dynamic knowledge based on our work, but situated in the broader literature on student learning, instruction, and cognition. We invite our community to join us in 1) extending assessments to include

nonverbal metrics such as *noticing*, 2) experimenting on how *framing* the same task differently elicits different knowledge from students, 3) identifying how temporal *sequencing* constrains learning mechanisms, 4) determining how *representing* ideas differently offers different insight into student knowledge, and 5) *evaluating* the ways we wish to use theory in BER.

### Noticing Students' Emerging and Dynamic Knowledge

When a student responds to a question on an exam or during a lecture, does the answer reflect only a *product of retrieving knowledge* or also a *dynamic process of constructing knowledge*? If we accept the primary tenet of constructivism—that new knowledge is constructed from old knowledge—we must seek mechanisms for how this construction occurs. Attending to students' use of knowledge from moment to moment when speaking or drawing is one way.

Noticing is another (Sherin *et al.*, 2011). For example, investigations with in-service math teachers illustrate how *before* teachers change their pedagogical beliefs or practices, they might notice moments during instruction worthy of attention and reflection. These moments may be problematic or productive aspects of teaching or learning. Our point is that knowledge can be demonstrated in ways that do not require explicit verbal production or other explicit tasks (e.g., inscription or forced choice) that entice us to focus on determinacy. Instead, we can observe students' knowledge by attending to what students notice as salient features in a given context. We can then design for what we want them to notice.

In our study, noticing was observed during the diagram interpretation task. When students interpreted the diagrams, those with the SR models noticed features they neglected to include previously. In contrast, the students with the IOO models noticed fewer features, but they noticed the ones that they determined were incorrect. Just as noticing classroom events can reflect a teacher's pedagogical knowledge, noticing features in external representations can reflect a student's emerging but unstable biological knowledge. Thus, by leveraging the construct of noticing, we can detect and assess moments critical to future learning. Other implicit measures such as gesture analysis (Lira and Stieff, 2018) and eye tracking (Ubben *et al.*, 2018) provide windows into what students notice in cases when they do not state their noticing verbally. In sum, noticing offers one way to observe dynamic knowledge construction.

### The Impact of Task Framing on Eliciting Student Knowledge

Framing—not to be confused with our use of analytic frameworks—refers to the process whereby people come to understand particular instances of a social activity as belonging to more general classes (cf. Russ *et al.*, 2012). For example, a particular conversation between interlocutors may signal clues (e.g., jokes, smiling, winking) that the conversation is informal and jovial. Likewise, a particular conversation between a teacher and student may signal clues (e.g., questions, evaluations, explanations) that the conversation is didactic. Although we observed framing at play elsewhere, the diagram interpretation task highlights this phenomenon best. Maggie and her peers who presented the SR model did not correct their explanations or drawings explicitly. To our surprise, the students who presented the IOO model did make explicit corrections,

despite their explanations more closely reflecting the instructional diagram.

Several reasons might explain this pattern, but framing seems a likely candidate, because we did not frame the diagram interpretation task as an explicit opportunity to correct students' earlier explanations. Students do not always frame disciplinary activities in the ways instructors and designers intend (Redish, 2014). An instructor might engage in a thought experiment or construct a hypothetical world to illustrate a point to students. If the rules and goals of such an activity are not stated, students may perceive the instructor's actions as purely pedantic exercises. Put differently, some students might not recognize the epistemic game and its value (Hammer and Elby, 2003). Cindy and the IOO students presented evidence that suggested that they framed the diagram interpretation task as a learning opportunity. Maggie and the SR students, however, failed to present such direct evidence. This result may be due not to variable knowledge per se but differences in student framing. We therefore suggest that future research could experiment on how different framings elicit different knowledge within or between groups of students. At a minimum, designers of learning and assessment tasks should frame tasks explicitly when they aim to detect the acquisition of content knowledge or skill.

### Representing to Learn and to Assess

External representations may be used for learning (Ainsworth, 2006) or assessment (Quillin and Thomas, 2015). But they also support students in constructing different forms of knowledge (Sherin, 2001; Parnafes, 2007). This point matters when assessing dynamic knowledge. We assessed students by asking them to explain the knee-jerk response verbally, then again after drawing, and then once more after interpreting an instructional diagram. Such a design invites instability. Perhaps a more rigorous design would have eliminated representations and instead asked students to explain three times.

Although we acknowledge this critique, we defend our design on empirical and theoretical grounds. We demonstrated empirically that students present dynamic knowledge *within* tasks not just between—Maggie demonstrated this during her first explanation and Cindy did so while drawing. Moreover, in theory, a design that includes representations approximates ecological validity. Engagement of student knowledge, eliciting representations, and encouraging learning in interviews all reflect authentic activities that manifest in classrooms (diSessa, 2007). Instructors task biology students with coordinating their knowledge with representations (Tsui and Treagust, 2013; Lira and Stieff, 2018), and students must learn to stabilize their knowledge when doing so (Parnafes, 2007).

### Sequencing Tasks and Sequencing Effects

The representations educators select matter in relation to their sequencing (Ainsworth, 2006; Fyfe *et al.*, 2014). Assessing students with external representations provides insight into their dynamic knowledge. But what do these particular representations in this particular sequence mean for assessment or the dynamic knowledge that manifests during it? To gain insight into students' dynamic knowledge from a different vantage point, we might have included a quantitative representation such a graph or another display of data rather than a diagram. This decision would have tasked students with coordinating

knowledge across multiple representations that highlight distinct aspects of the phenomenon and perhaps engendered the construction of new or different knowledge (Ainsworth, 2006; Parnafes, 2007). Such a design holds promise to elicit variability, but we note that neither design is inherently superior. Using the same representational forms (drawings and diagrams) facilitates comparisons by students.

Had we opted to introduce the diagram first and then asked students to explain, students might have presented accurate and stable knowledge. They might have simply reported what they had learned from the diagram as well as what the diagram cued them to remember from their prior instruction. Perhaps no conceptual challenges would have appeared. Such a design would have placed a rigorous test upon claims regarding students' dynamic knowledge. But, once again, it would have afforded less insight into the dynamics of moment-to-moment thinking. Moreover, such a design reflects undesirable learning conditions. More favorable designs tend to support students in cueing prior knowledge, identifying learning challenges, drawing distinctions between ideas, and *then* experiencing explicit instruction (cf. Schwartz and Bransford, 1998).

### Selecting and Evaluating Theory

We selected three frameworks among many that exist. Other theoretical or assessment frameworks and their associated research paradigms could have granted different insight into determinacy or the other properties of knowledge. We selected these three because BER investigators use SBF (Dauer *et al.*, 2013) and because MM and CD approaches have commonly been positioned in opposition in science education research (Southerland *et al.*, 2001). Had we, however, adopted a knowledge integration framework (Linn, 2000), we might have better designed for and observed the malleability of students' knowledge. If we had set out to understand the coconstruction of knowledge, we might have opted for an interaction analysis (Jordan and Henderson, 1995) that granted insight into how two or more students fostered inter-subjective coherence (i.e., shared knowledge). We stress the point that each framework is a framework but not *the* framework.

This points lead us back to our decision on how to use and evaluate multiple frameworks. We selected a microcomplementary approach, because we aimed to illustrate how different frameworks afford different and sometimes complementary insights into student knowledge (diSessa *et al.*, 2016). We see value in our community reading and understanding the theoretical constructs and methodological approaches of more than one framework, because our field is still emerging and reflecting (Lo *et al.*, 2019). A competition strategy might reject a useful framework prematurely. Competition will pave the way for rejecting poor accounts, but we think it wise to do so after we explore what frameworks have to offer across multiple contexts.

Likewise, at present, it seems premature for us to fuse frameworks. Although we see great value in setting this goal, the field's recent surge and relatively recent emergence suggest that this goal lies ahead of us, but not today. Our microcomplementary approach highlighted the point that different frameworks apply distinct *ontologies*—they categorize, characterize, and call attention to different kinds of properties of knowledge (de Jong *et al.*, 1998) to form philosophical foundations for educational research (cf. Bredo, 2006). In recent years, scholars have

addressed the issue of dynamic knowledge by arguing matters of usage (e.g., *misconceptions* vs. *alternative* conceptions) and directing BER away from the theoretical debate and toward a focus on the research shift from “cold” to “hot” cognition (Leonard *et al.*, 2014). For scholars who see value in continuing the dynamic knowledge conversation, leveraging multiple frameworks to analyze a data set offers one strategy for doing so.

In the future, more methodological work of the nature we presented will add rigor to the stability conversation and enrich theory in BER (Dolan, 2015). It will take time for our community to stabilize how we notice stability.

### ACKNOWLEDGMENTS

We thank Carly Richards (C.R.) for her support on portions of the analysis presented in this paper. We also thank two anonymous reviewers for their insightful and constructive comments.

### REFERENCES

- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, 16(3), 183–198. <https://doi.org/10.1016/j.learninstruc.2006.03.001>
- Bower, G. H., & Clark, M. C. (1969). Narrative stories as mediators for serial learning. *Psychonomic Science*, 14(4), 181–182. <https://doi.org/10.3758/BF03332778>
- Bredo, E. (2006). Philosophies of educational research. In Green, J. L., Camilli, G., & Elmore, P. B. (Eds.), *Complementary methods in educational research* (pp. 3–31). Washington, DC: Erlbaum.
- Campbell, N. A., Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Jackson, R. B. (2008). *Biology* (8th ed.). Berkeley, CA: Pearson.
- Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In Glaser, R. (Ed.), *Advances in instructional psychology* (pp. 161–238). Mahwah, NJ: Erlbaum.
- Dauer, J. T., Momsen, J. L., Speth, E. B., Makohon-Moore, S. C., & Long, T. M. (2013). Analyzing change in students' gene-to-evolution models in college-level introductory biology. *Journal of Research in Science Teaching*, 50(6), 639–659. <https://doi.org/10.1002/tea.21094>
- de Jong, T., Ainsworth, S., Dobson, M., van der Hulst, A., Levonen, J., Reiman, P., ... & Swaak, J. (1998). Acquiring knowledge in science and mathematics: The use of multiple representations in technology-based learning environments. In van Someren, M. W., Reiman, P., Boshuizen, H. P. A., & de Jong, T. (Eds.), *Learning with multiple representations* (pp. 9–40). Oxford, UK: Pergamon.
- Derry, S. J., Pea, R. D., Barron, B., Engle, R. a., Erickson, F., Goldman, R., ... & Sherin, B. L. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *Journal of the Learning Sciences*, 19(1), 3–53. <https://doi.org/10.1080/10508400903452884>
- diSessa, A. A. (1988). Knowledge in pieces. In Forman, G., & Pufall, P. B. (Eds.), *Constructivism in the computer age* (pp. 49–70). Hillsdale, NJ: Lawrence Erlbaum Associates. <https://doi.org/10.1159/000342945>
- diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, 10(2/3), 105–225. <https://doi.org/10.1207/s1532690xci1002>
- diSessa, A. A. (2002). Why “conceptual ecology” is a good idea. In Limon, M., & Mason, L. (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 28–60). Dordrecht, The Netherlands: Kluwer Academic Publishers. [https://doi.org/10.1007/0-306-47637-1\\_2](https://doi.org/10.1007/0-306-47637-1_2)
- diSessa, A. A. (2007). An interactional analysis of clinical interviewing. *Cognition and Instruction*, 25(4), 523–565. <https://doi.org/10.1080/07370000701632413>
- diSessa, A. A., Levin, M., & Brown, N. J. (2016). *Knowledge and interaction: A synthetic agenda for the learning sciences*. In diSessa, A. A. (Ed.), New York: Routledge.
- Dolan, E. L. (2015). Biology education research 2.0. *CBE—Life Sciences Education*, 14(4), ed1. <https://doi.org/10.1187/cbe.15-11-0229>
- Ehrlich, K., & Johnson-Laird, P. N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behavior*, 21(3), 296–306. [https://doi.org/10.1016/S0022-5371\(82\)90626-0](https://doi.org/10.1016/S0022-5371(82)90626-0)

- Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concrete fading in mathematics and science instruction: A systematic review. *Educational Psychology Review*, 26(1), 9–25. <https://doi.org/10.1007/s10648-014-9249-3>
- Ginsburg, H. P. (1997). *Entering the child's mind: The clinical interview in psychological research and practice*. New York: Cambridge University Press.
- Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Problems*, 12(4), 436–445.
- Gouvea, J. S., & Simon, M. R. (2018). Challenging cognitive construals: A dynamic alternative to stable misconceptions. *CBE—Life Sciences Education*, 17(2), ar34. <https://doi.org/10.1187/cbe.17-10-0214>
- Hammer, D., & Elby, A. (2003). Tapping epistemological resources for learning physics. *Journal of the Learning Sciences*, 12(1), 53–90. <https://doi.org/10.1207/S15327809JLS1201>
- Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-novice understanding of complex systems. *Journal of the Learning Sciences*, 16(3), 307–331. <https://doi.org/10.1080/10508400701413401>
- Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science*, 28, 127–138. [https://doi.org/10.1016/S0364-0213\(03\)00065-X](https://doi.org/10.1016/S0364-0213(03)00065-X)
- Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, 4, 71–115. [https://doi.org/10.1207/s15516709cog0401\\_4](https://doi.org/10.1207/s15516709cog0401_4)
- Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences*, 4(1), 39–103. [https://doi.org/10.1207/s15327809jls0401\\_2](https://doi.org/10.1207/s15327809jls0401_2)
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11(1), 65–100. <https://doi.org/10.1111/j.1551-6708.1987.tb00863.x>
- Lave, J., Murtaugh, M., & De La Rocha, O. (1984). The dialectical constitution of arithmetic practice. In Rogoff, B., & Lave, J. (Eds.), *Everyday cognition: Its development in social context* (pp. 67–94). Cambridge, MA: Harvard University Press.
- Leonard, M. J., Kalinowski, S. T., & Andrews, T. C. (2014). Misconceptions yesterday, today, and tomorrow. *CBE—Life Sciences Education*, 13(2), 179–186. <https://doi.org/10.1187/cbe.13-12-0244>
- Linn, M. C. (2000). Designing the knowledge integration environment. *International Journal of Science Education*, 22(8), 781–796. <https://doi.org/10.1080/095006900412275>
- Lira, M. E., & Gardner, S. M. (2017). Structure-function relations in physiology education: Where's the mechanism? *Advances in Physiology Education*, 41, 270–278.
- Lira, M. E., & Stieff, M. (2018). Using gesture analysis to assess students developing representational competence. In Daniel, K. L. (Ed.), *Towards a framework for representational competence in science education: Models and modeling in science education* (Vol. 11, pp. 205–228). Cham, Switzerland: Springer. <https://doi.org/10.1007/978-3-319-89945-9>
- Liu, L., & Hmelo-Silver, C. E. (2009). Promoting complex systems learning through the use of conceptual representations in hypermedia. *Journal of Research in Science Teaching*, 46(9), 1023–1040. <https://doi.org/10.1002/tea.20297>
- Lo, S. M., Gardner, G. E., Reid, J., Napoleon-Fanis, V., Carroll, P., Smith, E., & Sato, B. K. (2019). Prevailing questions and methodologies in biology education research: A longitudinal analysis of research in *CBE—Life Sciences Education* and at the Society for the Advancement of Biology Education Research. *CBE—Life Sciences Education*, 18(1), ar9. <https://doi.org/10.1187/cbe.18-08-0164>
- Maskiewicz, A. C., & Lineback, J. E. (2013). Misconceptions are “so yesterday!” *CBE—Life Sciences Education*, 12(3), 352–356. <https://doi.org/10.1187/cbe.13-01-0014>
- Michael, J., Martinkova, P., McFarland, J., Wright, A., Cliff, W., Modell, H., & Wenderoth, M. P. (2017). Validating a conceptual framework for the core concept of “cell-cell communication.” *Advances in Physiology Education*, 41(2), 260–265. <https://doi.org/10.1152/advan.00100.2016>
- Norman, D. (1983). Some observations on mental models. In Gentner, D., & Stevens, A. L. (Eds.), *Mental models* (pp. 7–14). Hillsdale, NJ: Erlbaum.
- Parnafes, O. (2007). What does “fast” mean? Understanding the physical world through computational representations. *Journal of the Learning Sciences*, 16(3), 415–450. <https://doi.org/10.1080/10508400701413443>
- Pellegrino, J. W. (2012). Assessment of science learning: Living in interesting times. *Journal of Research in Science Teaching*, 49(6), 831–841. <https://doi.org/10.1002/tea.21032>
- Quillin, K., & Thomas, S. (2015). Drawing-To-Learn: A framework for using drawings to promote model-based reasoning in biology. *CBE—Life Sciences Education*, 14(1), es2. <https://doi.org/10.1187/cbe.14-08-0128>
- Redish, E. F. (2014). Oersted Lecture 2013: How should we think about how our students think? *American Journal of Physics*, 82(6), 537–551. <https://doi.org/10.1119/1.4874260>
- Russ, R. S., Lee, V. R., & Sherin, B. L. (2012). Framing in cognitive clinical interviews about intuitive science knowledge: Dynamic student understandings of the discourse interaction. *Science Education*, 96(4), 573–599. <https://doi.org/10.1002/sce.21014>
- Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, 92(3), 499–525. <https://doi.org/10.1002/sce.20264>
- Scherr, R. E. (2007). Modeling student thinking: An example from special relativity. *American Journal of Physics*, 75(3), 272–280. <https://doi.org/10.1119/1.2410013>
- Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, 16(4), 475–522.
- Sherin, B. L. (2001). A comparison of programming languages and algebraic notation as expressive languages for physics. *International Journal of Computers for Mathematical Learning*, 6(1), 1–61. <https://doi.org/10.1023/A:1011434026437>
- Sherin, B. L., Krakowski, M., & Lee, V. R. (2012). Some assembly required: How scientific explanations are constructed during clinical interviews. *Journal of Research in Science Teaching*, 49(2), 166–198. <https://doi.org/10.1002/tea.20455>
- Sherin, M. G., Russ, R. S., & Colestock, A. A. (2011). Accessing mathematics teachers' in-the-moment noticing. In Sherin, M. G., Jacobs, V. A., & Philipp, R. A. (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 79–94). New York: Routledge. <https://doi.org/10.4324/9780203832714>
- Silverthorn, D. U. (2019). *Human physiology: An integrated approach* (8th ed.). Upper Saddle River, NJ: Pearson.
- Smith, J. P., III, diSessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences*, 3(2), 115–163. <https://doi.org/10.1207/s15327809jls0302>
- Southerland, S. A., Abrams, E., Cummins, C. L., & Anzelmo, J. (2001). Understanding students' explanations of biological phenomena: Conceptual frameworks or p-prims? *Science Education*, 85(4), 328–348. <https://doi.org/10.1002/sce.1013>
- Tsui, C.-Y., & Treagust, D. F. (2013). Introduction to multiple representations: Their importance in biology and biological education. In Tsui, C.-Y. & Treagust, D. F., (Eds.), *Multiple representations in biological education* (pp. 3–18). Dordrecht, The Netherlands: Springer. [https://doi.org/10.1007/978-94-007-4192-8\\_1](https://doi.org/10.1007/978-94-007-4192-8_1)
- Ubben, I., Nitz, S., Daniel, K. L., & zu Belzen, A. U. (2018). Assessing representation competence with eye tracking technology. In Daniel, K. L. (Ed.), *Towards a framework for representational competence in science education. Models and modeling in science education* (pp. 229–246). Cham, Switzerland: Springer.
- van Mil, M. H. W., Boerwinkel, D. J., & Waarlo, A. J. (2011). Modelling molecular mechanisms: A framework of scientific reasoning to construct molecular-level explanations for cellular behaviour. *Science and Education*, 22(1), 93–118. <https://doi.org/10.1007/s11191-011-9379-7>
- Vattam, S. S., Goel, A. K., Rugaber, S., Hmelo-Silver, C. E., Gray, S., & Sinha, S. (2011). Understanding complex natural systems by articulating Structure-Behavior-Function models. *Educational Technology and Society*, 14(1), 66–81.
- Zohar, A., & Ginossar, S. (1998). Lifting the taboo regarding teleology and anthropomorphism in biology education—Heretical suggestions. *Science Education*, 82(6), 679–697. [https://doi.org/10.1002/\(SICI\)1098-237X\(199811\)82:6<679::AID-SCE3>3.3.CO;2-9](https://doi.org/10.1002/(SICI)1098-237X(199811)82:6<679::AID-SCE3>3.3.CO;2-9)