JOURNAL OF CHEMICAL EDUCATION

pubs.acs.org/jchemeduc Article

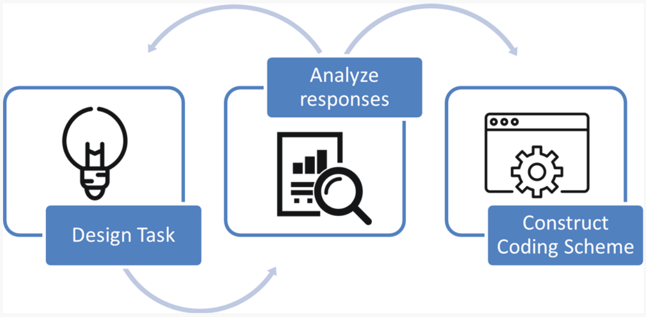
A Deep Look into Designing a Task and Coding Scheme through the Lens of Causal Mechanistic Reasoning

Keenan Noyes,* Clare G. Carlson, Jon R. Stoltzfus, Christina V. Schwarz, Tammy M. Long, and Melanie M. Cooper

Cite This: https://doi.org/10.1021/acs.jchemed.1c00959 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The purpose of this paper is to share the iterative process we used to design a task that elicits causal mechanistic reasoning and how the subsequent student responses can be analyzed. Our goal in this task is to strike a balance between eliciting as much student knowledge as possible without providing so much structure that the answer becomes obvious. The task development was approached using (1) a resources perspective of learning, (2) principles of scaffolding, and (3) evidence-centered design, for which we specified evidence that would be considered a fully causal mechanistic explanation. That is, an explanation which pays explicit attention to the properties, interactions, and behaviors of entities that are involved at a scalar level below the phenomenon under consideration. Since our eventual goal is to characterize how students use knowledge across disciplinary boundaries, the phenomenon of protein−ligand binding was chosen as the context for this task, because it requires students to apply ideas learned in chemistry courses to a biological phenomenon. After three rounds of iterative refinement, a final task was developed. To characterize students' responses to this task, we developed a coding scheme which can be used to code explanations based on the presence or absence of three key ideas relevant to this phenomenon. In this paper, we share the detailed processes and approaches used in task development, which we hope will provide insight into instructors and researchers as they, too, develop such tasks to explore student reasoning.



KEYWORDS: *First-Year Undergraduate/General, Chemical Education Research, Interdisciplinary/Multidisciplinary, Testing/Assessment, Noncovalent Interactions*

FEATURE: Chemical Education Research

■ INTRODUCTION

Our interdisciplinary team of researchers has been studying how students connect and use ideas across the disciplines of chemistry and biology at different scales. In particular, we have chosen the lens of causal mechanistic reasoning (CMR), which connects phenomena to the behaviors and interactions of entities at lower scalar levels and provides a powerful predictive and explanatory strategy.[1] By engaging students in tasks that promote CMR, our goal is to provide approaches that are appropriate across a range of courses and contexts, and through analysis of students' explanations, can inform how we might better help students make interdisciplinary connections. However, the development and evaluation of such tasks is by no means trivial. This paper presents the process by which we designed one such task and the resulting coding scheme that will allow us to characterize student mechanistic explanations both within and across disciplines. While the task and coding scheme will eventually be used in this way, it is the development process and qualitative analysis guiding the revisions which we describe in this paper—such analysis is a finding in and of itself.[2]

Causal Mechanistic Reasoning

Our work is based on the causal mechanistic reasoning framework (hereon referred to as the CMR framework) outlined by Krist et al.[1] This is a simplified framework based on that described by Russ et al.,[3] designed to work across content areas and for written assessments, which involves three steps. The first is to consider the level below the target phenomenon. The second is to identify and unpack the behaviors and interactions of entities at that lower level. And finally, the third is to connect how the lower-level behaviors give rise to the target phenomenon.[1] These steps, both individually and taken together, form a powerful thinking strategy that is central to all science disciplines.

Received: September 8, 2021
Revised: January 8, 2022

It is possible that causal relationships, and the mechanisms that underly them, could provide an important connection between chemistry and biology, as phenomena at the atomic, molecular, and cellular levels are the result of behaviors and interactions of entities at smaller scalar levels. Thus, being able to reason causal mechanistically might better help students navigate between these courses and make predictions about biological phenomena using chemistry ideas. For these reasons, we used this CMR framework to guide the development of a task and subsequent evaluation of responses to that task.

Resources Perspective of Student Learning

In the development of tasks, we approach learning through a *resources* perspective, which theorizes that students call upon context-dependent conceptual and epistemological resources to make sense of phenomena.[4,5] This is a departure from previous research identifying misconceptions, which is based on the underlying assumption that students hold coherent and intact conceptions, which should be challenged and replaced with more expert conceptions.[5,6] While research into misconceptions has revealed the range of problematic ideas students hold, this approach tends to emphasize a deficit view of learning and does not adequately account for how these novice ideas are replaced with, or evolve toward, more expert ones.[6] In contrast, the resources perspective takes a constructive approach to student thinking by focusing on the knowledge pieces that students *do* have in contexts and how students use those ideas in productive ways.[4,5] Through this lens, the aim of teaching is not to replace the students' ideas, but to design instruction that activates appropriate and productive resources for the specific context, allowing students to use and advance those ideas.

This perspective on student thinking has important implications for how knowledge is transferred to new contexts, including across disciplinary boundaries. The resources perspective suggests that transfer is not the movement of intact ideas, but rather the activation of similar resources in different contexts.[5] For example, when considering why atoms interact in a chemical context, conceptual resources related to electrostatic interactions and forces may be activated. When then asked to consider a biological phenomenon, such as why a ligand binds to an enzymatic binding site, the goal would be to activate these same cognitive resources in the new situation. If the student can repeatedly use and coordinate such resources in productive ways across multiple contexts, these connections may strengthen. However, developing expertise requires many years of experience working across a range of contexts to develop a connected framework of knowledge. It is therefore important that we give our students ample opportunity to use and develop their resources productively, and one such way we aim to do this is through formative assessment tasks.

Assessment Design

Assessment of student learning can be thought of as a process by which evidence is elicited to make an argument about what students know and can do.[7] The general approach to the design of assessments involves specifying the type of cognition (or theory of learning that the assessment is designed to assess), and, in this case, we are using the resources perspective. Then, we should define what types of observations would produce data that can be ultimately interpreted as evidence of learning. To do this, we used a modified evidence-centered design (ECD) approach,[8] which requires we first identify the resources (or evidence) we wish to elicit from students. That is, what would it look like for a student to reason causal mechanistically through the phenomenon of protein–ligand binding. We decided that explanations which appropriately leverage the electrostatic properties of the atoms and/or amino acids to explain the attraction between the protein and ligand would be evidence of CMR. While other ideas such as the size, shape, or orientation of the ligand also influence its ability to bind with a protein, the impact of all these characteristics on binding are inherently electrostatic in nature—maximizing attractive interactions and minimizing repulsive ones. Further, while we know that protein–ligand binding occurs in an aqueous environment, which brings with it the added complexity of entropic changes, we chose not to probe such ideas in this activity. It is our experience that designing tasks that probe for all these ideas at the same time leads to confusion and problems with coding the student responses.

The next step in ECD is to develop a task capable of eliciting such evidence. The task design is key as it provides the context that determines which of the students' resources are activated. One way to attempt to elicit particular resources is to provide scaffolding that supports students' reasoning and guides them to an appropriate response. Wood et al. coined the term scaffolding while exploring the methods by which an expert may help a beginner accomplish a task that they would be unable to do on their own.[9] This approach may be considered as a way to help students traverse Vygotsky's Zone of Proximal Development (ZPD)—the gap between what a learner can do on their own and what they can do with assistance from an expert.[10,11] While scaffolding was initially studied in one-on-one or in-person environments, this approach has also been applied to assessment design.[12]

Wood et al. identified several productive scaffolding techniques in their initial paper, two of which have been key for the development of our task: reduction in degrees of freedom and marking critical features.[9] Both focus the learners' attention on specific ideas that we, the experts, have deemed important to the phenomenon. Additionally, these techniques align well with Hammer's resources perspective: by attempting to focus the learner's attention to relevant productive ideas, we can activate those resources so they may then use those and any other closely linked resources to reason through the task at hand.[4]

A good task though, must do more than support the activation of relevant resources in students' minds; it must also communicate what information we expect them to include in their response. However, like Goldilocks and her porridge, a task must be carefully balanced, providing enough information for the student to understand what is being asked, but not so much information that it can be answered without thoughtful effort. To design such a task, we drew on the work done by Graulich et al. to incorporate scaffolding through contrasting cases.[13−15] In their work, they asked students to compare two phenomena using a series of questions helping students to consider (1) what explicit structural features are different between the two cases, (2) how the phenomenon at hand differs for each case, and finally (3) how the resulting implicit properties of the structures lead to this change. This scaffolding design was illuminating, both because of its ability to foreground aspects of the phenomenon that might otherwise be missed, and the subtlety of the scaffold. To answer the question, the student must go down to the scalar level where the two cases differ, thereby providing a more "natural" way of framing the scalar level at which we wanted them to explain the phenomenon.

Of course, designing an assessment that provides the right amount of scaffolding often takes multiple rounds of task design and evaluation of the students' responses (Figure 1). The iterative nature of this process is important to designing high-quality assessments.[7] While some researchers have shared portions of this process when discussing their tasks,[16−19] in this paper we provide an in-depth look at how we designed a task which reveals how students reason through the mechanism of protein−ligand binding, detailing the task revisions and underlying decisions which drew from principles of ECD, scaffolding, and the resources perspective. Although this project is part of a larger interdisciplinary research endeavor, in this paper, we only discuss the development of this assessment (hereon referred to as the "PL task", for Protein−Ligand) and the subsequent coding scheme used to characterize students' explanations in relation to CMR.

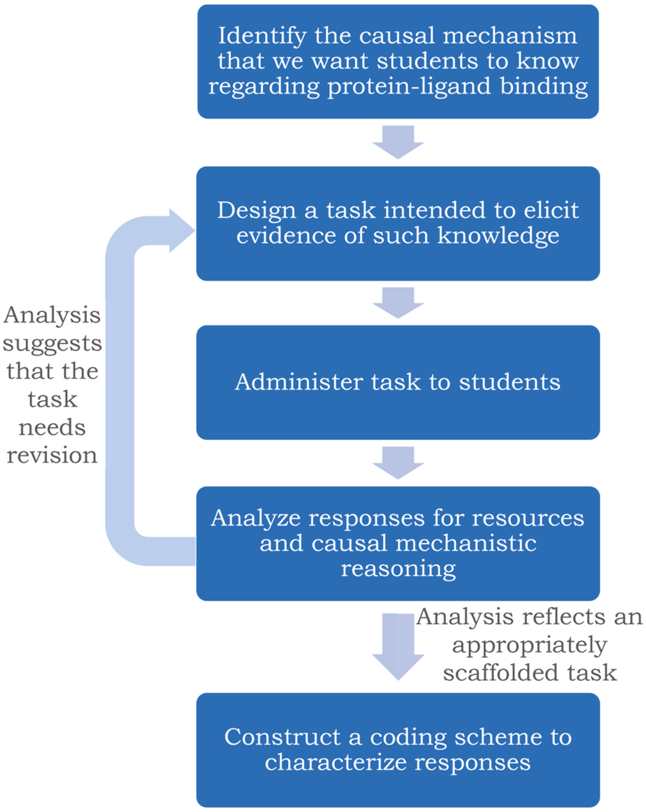


Figure 1. Our iterative process of designing a well-scaffolded assessment task.

RESEARCH QUESTIONS

The following questions will be addressed in this paper.

1. What is the impact of different types of scaffolding on the resources students use to respond to the task?
2. In what ways can we characterize the degree to which students are engaging in CMR to explain this phenomenon?

METHODS

Overview of the Rationale for the Design of the PL Task and Associated Coding Scheme

We chose to situate our task in the phenomenon of protein−ligand (PL) binding, because, while it is positioned in a molecular biology context, it requires that students use core ideas from chemistry to construct a mechanistic explanation. To construct such an explanation, the student must think about the scalar level below the phenomenon, unpack the properties and behaviors of the entities at that scalar level, and link those properties and behaviors to the phenomenon. In this context, an ideal response identifies the charges or partial charges of the atoms or functional groups in the protein and ligand (that is where the electron density is distributed) and explains that the oppositely charged entities experience an attractive noncovalent interaction. Furthermore, we designed the task to incorporate two contrasting cases so that students could compare two potential binding sites and explain, using the strength of the charges, which site would bind most strongly and why their selected site preferentially binds the ligand. Then, using responses from the final version, we aimed to characterize the ideas that students used as well as their engagement with CMR to develop the coding scheme.

The development of this task and coding scheme required several iterations, and, in the methods section, we describe the different groups of students whose responses we analyzed during this process. As our analysis of the students' responses drove the iterative redesign of the PL task, we describe key features of each task version and the rationale for our design decisions in the results section. The coding scheme has both analytic and holistic components to it and, while the development of the analytic rubric required several rounds of refinement, in this paper we present only the final coding approach. However, in the Supporting Information S1, we include a detailed discussion of our decisions about several earlier bins that were refined, removed, or combined with others to ultimately reach our final analytic rubric. We have included this discussion in accordance

Table 1. Overview of Administration of Each Version of the PL Task

semester	PL task version	course	responses collected (response rate %)	responses analyzed (task development)	responses analyzed (coding scheme development)
Fall 2018	1	Organic Chemistry 1	44 (75%)	20	
Fall 2018	1	Molecular Biology	94 (68%)	20	
Spring 2019	2	Organic Chemistry 2	74 (74%)	20	
Spring 2019	2	Molecular Biology	313 (74%)	20	
Summer 2019	3	General Chemistry 2	61 (73%)	29	61
Fall 2019	final	Molecular Biology	121 (55%)		60
Fall 2019	final	Organic Chemistry 2	300 (85%)		60

with calls for greater transparency surrounding coding scheme development and application.[2]

Participants

For the development and testing of different versions of this task and coding scheme, we collected and analyzed the responses of students enrolled at a large midwestern public research institution, including participants from Molecular Biology (MB), General Chemistry 2 (GC2), and Organic Chemistry 1 (OC1) and Organic Chemistry 2 (OC2). The majority of these students are familiar with being asked to explain phenomena as the general and organic chemistry courses from fall 2018 and spring 2019 have been transformed, using the three-dimensional learning (3DL) approach,[20,21] and the molecular biology course is also undergoing transformation using the 3DL approach. The fall 2019 OC2 course had not been transformed using the 3DL approach. Unfortunately, due to logistical constraints (e.g., not all courses were offered every semester, existing agreements with instructors for data collection), we could not administer every version of our task to students in all these courses. Students were offered a small amount of extra credit in their course for completing our activity. The number of student responses collected and analyzed is shown in Table 1. All the students in this study consented for their work to be used for research purposes, and their responses were collected and deidentified in accordance with our IRB protocol. The names included in this manuscript are all pseudonyms.

Most student responses were collected using the online assessment system beSocratic, which allows students to draw and write free-form responses.[22] Early versions (1 and 2) were administered to MB students on a hard-copy worksheet. For the final version, all students responded using beSocratic. These activities consist of a series of "slides", on which students can write or draw. In versions 2 and 3, we administered alternate versions (described in the results section) of the activity to test small, specific aspects of the task. Students in these semesters randomly received either the original or alternate version (full student counts presented in Supporting Information S2).

Analysis Guiding Task Development

Our iterative analyses of the students' drawings and explanations guided the development of the PL task. The first author qualitatively analyzed small sets of responses (randomly selected via random number generator) from each course and each version of the task to identify themes and patterns in the students' responses (Table 1). We selected more responses to analyze from version 3 to ensure that we could stop making substantive changes to the task after this version. Due to differences in the assessment medium, the first author was aware of the course background of each student (i.e., if the response came from a biology or chemistry student) during the analysis.

In this analysis, the first author examined the students' entire response (instead of their responses to each individual question) to explore all the resources they used to approach the overall phenomenon. Then, we compared the ideas that students used with our ideal causal mechanistic explanations of the phenomenon. That is, did the students identify and unpack the properties and behaviors of entities a scalar level below to explain the phenomenon. After analyzing the responses from one version of the task, we shared our findings with our larger interdisciplinary team to discuss changes we could make to the scaffolding of the following version of the task to better activate the appropriate resources and elicit CMR. Examining the students' responses also provided us the opportunity to reflect on our evidence statements and determine what would be reasonable for a student to include in their explanation. We repeated this process three times until the task appropriately cued the students to provide a causal mechanistic response. In an effort to be as transparent as possible about our analysis that drove the task refinement, we have provided the full responses from all the students in the Supporting Information S3.

Analysis Guiding Coding Scheme Development

Following version 3, we stopped making substantive changes to the PL task, which is why we used responses from both this version and the final version to guide the coding scheme development. Using these more effective tasks, the first and second authors qualitatively analyzed larger sets of responses from version 3 ($N = 61$) and the final version ($N = 120$) of the PL task (Table 1). By analyzing more responses at this stage, we hoped to compile a more exhaustive list of the ways students responded to our question and use that information to build an analytic rubric. To capture the presence or absence of the ideas in a student's response, we initially used an analytic approach, rather than a holistic one, which would require assigning a single code to each response as a whole. However, once the final analytic rubric was determined, it was used to develop holistic codes based on different combinations of the presence or absence of the key ideas included in the explanations. During development of the analytic rubric, the authors were aware of the course background of the students they were coding to provide context for the ways in which the students answered the task. In this process, the first and second authors independently analyzed the responses in sets of 30, meeting afterward to discuss the ideas students included in their explanations and how those ideas relate to the CMR framework. Drafts of the analytic rubric were continually modified based on discussions between the first and second authors, as well as the input of our larger interdisciplinary team, and the iterative process continued until we felt that we had accounted for all the key ideas relevant to constructing a causal mechanistic explanation, of which there were three. At this point, we developed the holistic scheme characterizing if the student included some, all, or none of the key ideas in the analytic rubric. We recognize the importance of the researchers' roles in decision-making about codes, so we have included a more detailed discussion of the analytic rubric development in Supporting Information S1.

Once the analytic rubric and holistic scheme had been finalized, the first and second authors used this approach to code the 181 responses used for coding scheme development. To determine inter-rater reliability we calculated Cohen's kappa for our initial coding.[23] For the analytic bins, Cohen's kappa ranged from 0.859 to 0.945, and for the holistic codes Cohen's kappa was 0.873. All of the Cohen's kappa values were greater than 0.8 suggesting high (almost perfect) agreement.[24]

RESULTS AND DISCUSSION

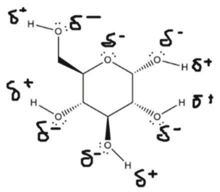
PL Task Development

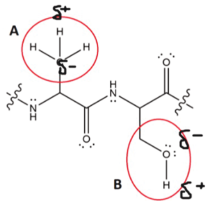
In this section, we discuss the key features of each version of the PL task and our analysis of the responses which guided the task revisions. Boxes 1, 2, and 3 show abbreviated versions of each task in the body of the manuscript, and images of each full task version can be found in Supporting Information S4.

PL Task Version 1. In the initial version of the PL task, we chose to use a glucose molecule as the ligand and asked students a series of questions addressing how and why such a molecule interacts with a hypothetical peptide chain that had both polar

Box 1. The response from Isabella to PL task version 1

[Below], a glucose molecule is shown. Please draw any appropriate partial charges in this molecule.





Why did you draw the charges where you did in the glucose molecule? Hint: think about the role of the subatomic particles.

I placed partial negative charges around the oxygen molecules because they are highly electronegative, and partial positives on the H's because they are bonded to the oxtgen [sic] atoms.

Here is part of a peptide showing the two amino acids A and B. Please draw any appropriate partial charges in the side chains (circled) of each amino acid.

[Drawing shown in black rectangle above]

Why did you draw the partial charges where you did [on the amino acids]?

The top partial positive charge is spread out among the hydrogens, because they are attached to a carbon which is bigger and therefore more electronegative. On B, the partial negative is on the oxygen and the partial positive is on the Hydrogen. because the oxygen is highly electronegative.

Which amino acid would glucose more strongly interact with?

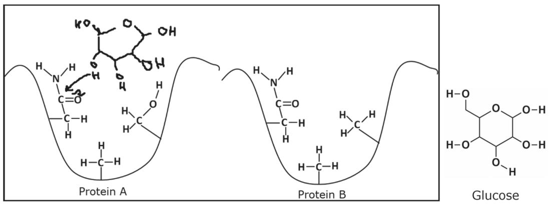
B

Why does the glucose interact more strongly with that amino acid?

B because glucose can form hydrogen bonds with the oxygen.

Box 2. The response from Katrina to PL task version 2

Pick the binding site you think is most likely to bind glucose (shown [below]) and draw a possible way glucose could bind in the binding site.



Explain why the protein you chose has the better glucose binding site and how the differences in the site cause this difference in binding. Include specific references to the figures.

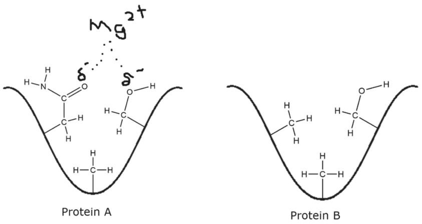
Protein A has a better glucose binding site because it has an additional OH group. This additional OH group can be added onto the glucose ring so that the OH group can help the biological system because O is necessary in these systems.

Use your drawing to help explain what causes glucose to bind to the protein.

Glucose binds to the protein because the OH groups are able to attack at the carbonyl and break the double bonded O. This creates a single bonded O and an OH group to the C. This helps for the ring to break and glucose to attach to the protein.

Box 3. The response from Conor to PL task version 3 alternate

The drawings below represent binding sites in two different versions of a protein showing only the atoms in relevant amino acid side chains. Consider a positively charged magnesium ion (Mg^{2+}). Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



Explain why the protein you chose has the better magnesium binding site and how the structural differences in the site cause this difference in binding. Include specific references to the figures.

The magnesium ion has a + 2 charge and will be attracted to the 2 slightly negative oxygen molecules that are present in protein A. The negative charge provided by a single oxygen molecule in protein B is not as strong as the charge in protein A.

Explain what causes the magnesium ion to bind to the protein making specific references to your drawing of the magnesium ion in the binding site.

The magnesium ion will form an ionic-dipole interaction with the slightly negative oxygen molecules in protein A. Electrostatic forces will keep the magnesium ion at the binding site until there in [sic] an introduction of energy into the system.

and nonpolar regions (Box 1). This protein−ligand pair was chosen because glucose can form multiple hydrogen bonds and is introduced in most introductory MB courses, so students are familiar with its structure. To activate students' resources related to the electrostatic nature of the interaction, we first asked students to draw the partial charges present on a glucose molecule and explain why they drew the charges in those locations. We also included a hint asking students to consider the role of the subatomic particles to lead them to think about the entities a scalar level below. This process was then repeated for a hypothetical peptide chain featuring an alanine and serine amino acid. Then, students were asked to predict which amino acid would interact most strongly with glucose and explain why the interaction was stronger (Box 1).

As we directly asked the students about the location of the charges, it was not surprising that all the students mentioned charge somewhere in their response. However, only about half of the students (OC1, $N = 10$, 50%; MB, $N = 9$, 45%) explicitly leveraged the charges to explain the attraction between their selected amino acid and glucose. One potential issue is that some students (OC1, $N = 6$, 30%; MB, $N = 8$, 40%) only named the attractive force (i.e., hydrogen bonding) without explicitly linking the interaction to the charges present. For example, Isabella (Box 1) assigned the appropriate partial charges to the alcohol functional group in serine and glucose but explained that glucose would interact more strongly with side chain B *"because glucose can form hydrogen bonds with the oxygen"*. Without explicitly connecting these ideas, it is unclear if this student understood the role of charges in hydrogen bonding, especially since they also drew partial positive and negative charges on the alanine side chain. This made it difficult to determine if the students truly understood the electrostatic nature of this interaction or if they had simply memorized that hydrogen bonds occur between hydroxyl groups—a strong possibility as many studies have identified noncovalent interactions as a difficult idea for students to understand.[25−29] These findings led us to believe that the task overemphasized the role of charges, which may have led students to discuss charges because we asked them to and not because they viewed those ideas as relevant in this context.

PL Task Version 2. To more appropriately cue ideas related to electrostatics, in PL task version 2 we removed the questions which asked about the locations of the partial charges in each molecule. Instead, we asked the students to draw glucose binding to one of two potential binding sites and explain (1) why their selected site had the better binding site and (2) what caused glucose to bind to the site (Box 2). By reducing the scaffolding, we could see if binding alone was enough to activate students' resources related to electrostatics. We also tested the order of the two explanation questions, creating an alternate PL task version 2 (see Supporting Information S4) in which we asked for the cause of the protein−glucose binding before we asked why glucose preferentially binds to one site.

We also changed the hypothetical binding sites in the task to include 3 amino acids, with both sites containing at least one polar amino acid (asparagine). As both sites could now interact with glucose, students would have to compare the relative strengths of the interactions with each site to determine which site would preferentially bind glucose. We hoped this would lead more students to discuss the relationship between the magnitude of the charge and the strength of the interaction rather than solely stating that one site could interact with glucose and the other could not.

We found that the ordering of the two explanation questions did not impact the students' responses; instead, the broader changes to the task had a more notable impact. Specifically, the reduction in scaffolding related to electrostatics caused fewer students (OC2, $N = 10$, 50%; MB, $N = 1$, 5%) to include charge in their explanations. Without appropriate activation from the task, the students' course enrollment (OC2 or MB) appeared to determine which resources were activated. For example, the majority of OC2 students ($N = 14$, 70%) approached this phenomenon as a reaction (instead of an interaction) in which the oxygen in glucose acted as a nucleophile to attack the carbonyl carbon in asparagine (see Katrina's response in Box 2). While not appropriate in this context, trying to predict how different species react with one another is a reasonable strategy for an organic chemistry course, in which reactivity and reactions are strongly emphasized.

Unlike the OC2 students, the MB students addressed the noncovalent binding of glucose to the protein. However, only one MB student discussed the role of charges in the binding. A few more MB students ($N = 5$, 25%) used polarity instead of charge to describe the properties of certain groups. The majority, however, ($N = 13$, 65%) named the noncovalent interaction, typically hydrogen bonding, without leveraging charge or polarity. Even though students identified that more hydrogen bonds formed with protein A, they did not explain *why* the atoms formed these interactions, so it is unclear if the students understood the electrostatic nature of this interaction or if they used a memorized heuristic (e.g., hydroxyl groups form hydrogen bonds). We highlight two students who discussed hydrogen bonding in Supporting Information S5. It is possible that some MB students were using resources related to electrostatics when reasoning through noncovalent interactions; however, without the appropriate framing for the task, students may have felt that simply identifying the presence of hydrogen bonding was a sufficient answer. These themes in the OC2 and MB student responses suggest that the task (1) did not activate the appropriate resources related to electrostatic ideas necessary to explain this phenomenon or (2) did not cue the students to include those ideas in their explanations.

PL Task Version 3. In the third version of this task, we tried to activate relevant electrostatic resources in a different way by replacing glucose with a positively charged magnesium ion (Mg^{2+}), a biologically relevant metal ion (Box 3). By changing the ligand to Mg^{2+} students could no longer use the ligand as a nucleophile and, also, it would no longer be sufficient for students simply to identify "hydrogen bonding" as the reason for binding. As we found no impact from the order of the explanation questions, in version 3 we used the order of the questions in the original PL task version 2 and arranged the questions so the students could see both questions at the same time. In this iteration, we tested an alternate version of the activity where we replaced the asparagine, instead of serine, with alanine in protein B to identify which combination of amino acids best elicited causal mechanistic responses (see both versions in Supporting Information S4).

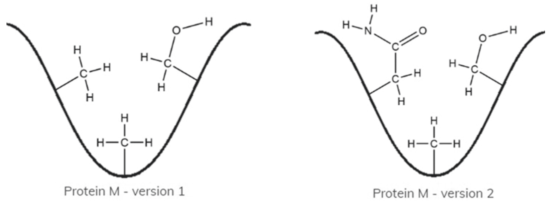
As with version 2, we found no difference between the original and alternate versions of the activity (which varied the amino acids in protein B); instead, the change of ligand appeared to have the largest impact to the students' responses. With this version, about half ($N = 16$, 55%) of the students explicitly identified the role of charge in the binding of Mg^{2+} to the protein, and 34% ($N = 10$) of the students both identified the role of charge *and* used the strength of the charge to explain the preferential binding (example in Box 3). Additionally, only a few students ($N = 5$, 17%) named the noncovalent interaction without including evidence of the role of electrostatics. Other students provided evidence that was not explicitly electrostatic in nature, such as the size of the groups in each protein or the number of bonds present in each site, to explain the binding of the ligand.

On the basis of these responses, we felt we had reached an appropriate amount of scaffolding. By switching from glucose to Mg^{2+}, we appeared to activate the resources related to electrostatics for those students who saw electrostatics as relevant to protein−ligand binding and encouraged them to include those ideas in their explanation. This change also avoided the unresolved issue of whether students were using the idea of hydrogen bonding appropriately, and at the same time, reduced the scaffolding so that not all students felt forced to discuss charge.

PL Task Final Version. Having struck (we believe) a good balance of scaffolding, the final version of the PL task (Box 4) has the same major features as version 3. In addition, we included two lessons learned from the original and alternate versions of the second and third iterations of this activity. First, we used the ordering of the explanation questions from the alternate version 2, in which the students explained what causes the protein−ligand binding before answering why their selected site preferentially binds Mg^{2+}. Although the ordering did not result in major differences in student responses, we used this order because it aligns better with the CMR framework since the student identifies and unpacks the properties and behaviors of

Box 4. The final version of the PL task

The drawings below represent binding sites in two different versions of protein M showing only the atoms in relevant amino acid side chains. Consider a positively charged magnesium ion (Mg^{2+}). Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



In the space below, explain what causes the magnesium ion to bind to the protein making specific references to your drawing.

[space for student response]

Explain why the protein you chose has the better magnesium binding site and how the structural differences in the site cause this difference in binding.

[space for student response]

the entities first (in explaining what causes the protein−ligand binding) before then using those properties and behaviors to explain why one site better binds the ligand. Second, we decided to use the distractor protein amino acids from the alternate PL task version 3 featuring two alanine residues and a serine residue. We chose this version because both binding sites contain the amino acid serine, which has a hydroxyl functional group. By having this functional group in both sites, we hoped to further reduce the number of students focusing on the presence or absence of hydrogen bonding as the sole reason for the binding of Mg^{2+}.

PL Coding Scheme Development

In developing the coding scheme, we leaned on work done by Jescovitch et al. in which a combination of analytic and holistic approaches were used to characterize explanations of a phenomenon.[30] Our goal was to characterize students' responses holistically, but, in order to do so, we first developed an analytic rubric to capture the presence or absence of specific ideas in the students' explanations. From the analytic rubric, we identified three key ideas which, when taken together, represent a fully causal mechanistic explanation. These ideas, while related, are not dependent on each other, and thus, we can use the analytic rubric as a means of identifying the combinations of conceptual pieces in students' reasoning. Thus, rather than attempting to make sense of the entire response all at once to assign a single holistic code, we could use the more structured analytic rubric to characterize the response by capturing the presence or absence of each of these three ideas. The resulting combinations of ideas in the analytic rubric can then be used to form a holistic scheme consisting of three codes: noncausal mechanistic (CM), partially CM, or fully CM. When a response provides evidence of an understanding of all three ideas in the analytic rubric, it is given the holistic code fully CM. In previous studies, some of the authors have characterized explanations about chemical phenomena as causal, mechanistic, or causal mechanistic;[17,31−33] however, the scheme we developed for this task emerged from this set of student responses and discussions with our interdisciplinary team, resulting in an approach that is distinct from (and should not be compared to) previous publications.

Developing the analytic rubric. We developed the analytic rubric based on emergent themes from our analysis of the students' responses and deductive themes based on the CMR framework.[1] That is, we leaned on the CMR framework and the principles of ECD to define what ideas are required for a fully CM explanation of Mg^{2+} binding preferentially to a protein site, but also tailored our rubric to be reflective of the knowledge pieces that students provided in their responses.

In the initial stages of development, we tried to capture as many ideas as possible with a range of analytic "bins". For example, does the student mention oxygen? Do they say oxygen is negative? Do they mention electrons? Do they identify a noncovalent interaction? Do they identify polar groups? During discussions between the first and second authors, both the bins themselves and the options for each bin were frequently refined, eventually resulting in our final analytic rubric which, after its use, can then be used to code responses holistically through the lens of CMR. Several decisions, such as whether to combine, refine, or remove conceptual bins, were made in this process (see Supporting Information S1); however, we want to be clear that while the resulting rubric is one way that we feel appropriately captures the ideas relevant to CMR in the context of protein−ligand binding, one might envisage others that are also effective.

Here, we present our analytic rubric and how it can be used to code responses holistically in terms of CMR, followed by a description of, and example responses for, each of these codes. As discussed earlier, a fully CM explanation for this phenomenon involves identifying the (partially) negative atoms in the binding sites, their attraction to Mg^{2+}, and the idea that one site is more negative and therefore more strongly attracts the Mg^{2+}, causing the preferential binding. This fully CM explanation can be deconstructed into three key ideas which make up our analytic rubric: an understanding of (1) the attraction between oppositely charged species, (2) the negative or polar nature of atoms and amino acids in the binding sites, and (3) the larger number of negative entities in one version, causing a stronger attraction to Mg^{2+}. These three ideas make up the analytic rubric specific for this task, but they also encompass the more general ideas laid out in the CMR framework.[1] We use the analytic rubric as the first step of characterizing a response by designating "yes" or "no" to each category depending on the presence or absence of that particular idea in the response. For example, a response may explain the attraction of oppositely charged species ("yes" for the first idea), but not identify a lower-level charged entity ("no" for the second idea), nor compare the magnitude of the charge between sites ("no" for the third idea). In another response, the combination may be "no" "yes" "no", for each respective bin. Once we have identified the combination of relevant ideas the response includes (based on this analytic rubric), it is used to assign a holistic CM code. During the process of finalizing this rubric, we encountered several "edge cases" or responses that fell right on the line between receiving a "yes" or a "no". We discuss these edge cases in Supporting Information S1.

The Holistic Scheme. As noted, a response could show the presence or absence of any of the three key ideas based on the evidence provided. To this end, leaving the categories independent of each other both (1) allows us to identify where students may be struggling to construct a complete CM response and (2) provides a more structured method of coding. To characterize each student's explanation, we used the analytic rubric to keep track of how many ideas (of the three that we

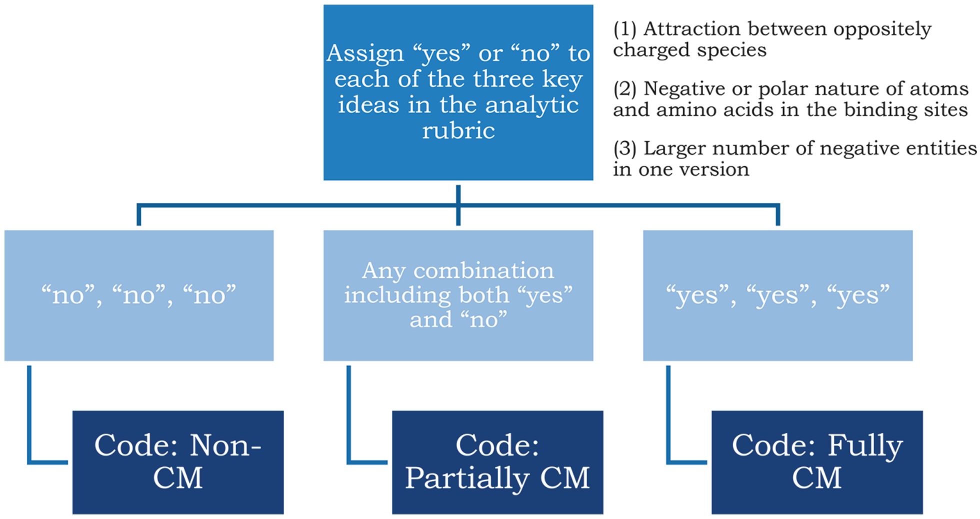


Figure 2. Process of using the analytic rubric to assign a holistic code for each response.

defined) they included, ultimately resulting in three holistic codes: non-CM, partially CM, and fully CM (Figure 2). Table 2 and the following subsections provide examples and more detailed descriptions of the different codes.

Non-CM. A non-CM explanation is one that does not provide evidence of any of the key ideas in the analytic rubric. For example, Claudia wrote,

> *"I believe proten [sic] B has the better binding site. Protein A already has a lot binded to it unlike the left structure on protein B. So i believe it will bind to that structure to sort of mimic protein A's structure. Im [sic] not totally sure about this. However, I believe it will choose the arealeast [sic] with the least amount of bonds already in order to create more and equal it out."*

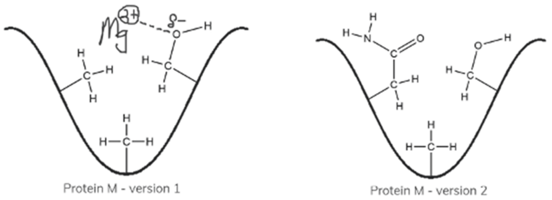
Claudia focused on a physical aspect of the binding site, noting that the site they chose had "the least amount of bonds". Although it is an interesting heuristic, identifying the number of bonds in each site is not productive in explaining this phenomenon and does not provide evidence of an understanding of electrostatics, so this receives "no" for each idea captured in the analytic rubric and, therefore, is coded as non-CM.

Partially CM. A partially CM response includes any combination of "yes" and "no" for the ideas in the analytic rubric (i.e., it has one or two of those ideas, but not all three). For example, consider Wayne's response in Box 5, which was given the combination "yes" "yes" "no" from the analytic rubric. Wayne provided an appropriate explanation of the role of electrostatics in causing Mg^{2+} to bind; however, rather than using electrostatics to explain the preferential binding, they invoked an additional resource saying,

> *"Version 1 has more space around the primary alcohol on the rightmost amino acid, allowing for the Magnesium to have an easier time binding to the site."*

Again, this other resource (shape/accessibility) may be a useful heuristic; however, it is not what causes the preferential binding and leads Wayne to the incorrect answer.

Box 5. Wayne's drawing and text responses.



In the space below, explain what causes the magnesium ion to bind to the protein making specific references to your drawing.

Version 1 has more space around the primary alcohol on the rightmost amino acid, allowing for the Magnesium to have an easier time binding to the site. Version 2 has the amide group on the rightmost amino acid instead of the methyl group in Version 1, which is much larger and blocks some of the alcohol's surface area.

Explain why the protein you chose has the better magnesium binding site and how the structural differences in the site cause this difference in binding.

The alcohol has a partially negative charge on the oxygen due to the oxygen being more electronegative and pulling the bonds' electrons closer to it. Since the Magnesium has a positive charge, it is attracted to the partial negative charge of the oxygen.

Fully CM. Responses are coded as fully CM explanations for this phenomenon by providing evidence of all three key ideas in the rubric. For example, Triston wrote

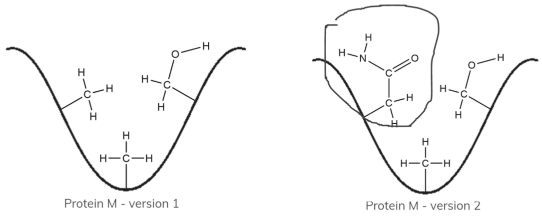
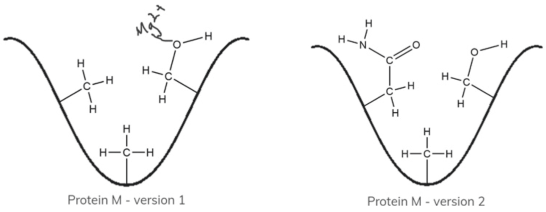
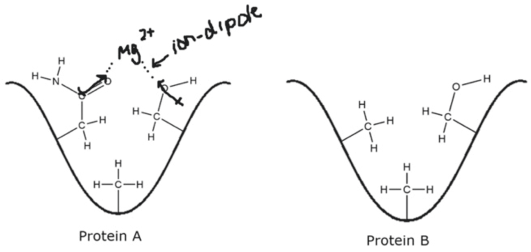
> *"Protein 2 has a better magnesium binding cite [sic] because there is a larger partial negative charge due to having two oxygens present. Protein 1 only has one oxygen to give a partial negative charge which will not attract nor stabilize the Mg ion as well. The protein 2 that I chose has the better binding cite [sic] because it has more partial negative (the two oxygen containing molecules) that attract the positive charge on the Mg 2+ ion."*

Triston leveraged electrostatics to not only explain the cause of Mg^{2+} binding, but also to explain why the version they chose is the better binding site.

Some students discuss the lack of polar groups in one version or the presence of hydrophobic/nonpolar groups, making it a poor binding site for Mg^{2+} (see Supporting Information S1 for our decisions regarding responses that discuss polarity instead of charge). For example, Jordan said

Table 2. Examples of Student Engagement in Causal Mechanistic Reasoning

Table 2. Examples of Student Engagement in Causal Mechanistic Reasoning

Student Text Response	Student Drawing	Type of explanation
"This protein has the better magnesium binding site because its easier to bind and the structural differences in the site causes a difference because it makes the structure longer. The extra carbon helps the magnesium ion to bind to the protein" -Teddy		Non-causal mechanistic
"I picked version one to bind Mg2+ to O. This is because O is usually negatively charged, so there is a high chance that it will bind to Mg which is positively charged. Oxygen is negatively charged. According to Coulomb's law, opposite charges attract. This is why I think Mg2+ will bind with O" - Simone		Partially causal mechanistic
"Protein A has a better binding site because the there are two polar molecules with partial negative charges compared to only one polar molecule in Protein B. Intermolecular forces bind the magnesium ion to Protein A. The positively charged magnesium ion interacts with the partial negative oxygens in the two polar molecules within the binding site through ion-dipole interactions." -Lois		Fully causal mechanistic

> *"Version 2 has the better magnesium binding site because there are uneven distributions of charges (C = O, O−H) in some of the side chains that allow for the positively-charged magnesium ion to bind with the negative charges of the oxygen atoms. Version 1 has very hydrophobic side chains (C−H) so they would shy away from the charged ion."*

Jordan did not explicitly say that one site is more negative than the other, but they noted that version 2 has negative charges to bind Mg^{2+}, while version 1 has hydrophobic groups which "*would shy away from the charged ion.*" While this response does contain an extra idea that is not strictly correct (i.e., the idea that hydrophobic groups are repelled by the ion), we believe Jordan did enough in comparing the charged nature of the two sites. Ultimately, the students constructing fully CM responses link their ideas about electrostatics to the phenomenon of Mg^{2+} binding preferentially to one of the sites because it is more negative and, therefore, more strongly attracts the ion.

■ LIMITATIONS

This study was designed and carried out primarily in the context of chemistry and biology courses which emphasize the importance of constructing causal mechanistic explanations for phenomena and give the students the opportunity to do so on assessments. It is probable that in another setting where the curriculum does not focus on these types of explanations that student responses to the task would be different. Additionally, we recognize that the number of responses analyzed for task versions 1 and 2 are somewhat low ($N = 40$), and it may be that analyzing more responses would have resulted in a different final task. However, this report is intended to show how we compromised to find a process that was feasible in a reasonable amount of time. That being said, the overall task and coding scheme design still took almost two years.

CONCLUSIONS AND FUTURE WORK

This paper describes the process by which we developed (1) a task to elicit causal mechanistic explanations of how species bind in a simplified biological system and (2) an associated coding scheme to make sense of those explanations. To do this, we drew upon literature related to evidence-centered design, the resources perspective, scaffolding, and causal mechanistic reasoning. Designing tasks to elicit rich explanations about how and why phenomena occur is difficult; we hope that the lessons we learned during this process can help other instructors and researchers to design effective assessments to collect evidence about what students know and can do.

In accordance with ECD, we first determined what we want students to know about this phenomenon and what work products we would accept as evidence of such an understanding.[8] These evidence statements allowed us to identify which resources the task must activate (e.g., electrostatics), which then influenced the design of the initial task. We iteratively refined the task three times, guided by our analysis of the students' responses, to better activate productive resources in the students' minds by modifying the scaffolding. This required us to reflect, carefully considering: (1) what resources students used to construct an explanation of the phenomenon, (2) what aspects of the current task activated those resources, and (3) how we might modify the scaffolding of the task to activate the appropriate resources. This process was not trivial, and the iterative design of the task was crucial to the development process, ensuring that the responses elicited most accurately reflected the students' understanding.

In comparing iterations, we found that seemingly small changes to the task (i.e., switching the ligand from a glucose molecule to a Mg^{2+} ion) dramatically changed how the students responded to the task. This highlights how aspects of the task can activate a particular set of resources, either productive or unproductive, that the student then uses to reason through the phenomenon. In our case, we found that the association between alcohol functional groups and hydrogen bonding is so strong that, in order to elicit explicit explanations of electrostatics, we had to change the ligand. Species such as glucose are still important molecules however, and cannot be avoided forever, so additional research is needed to explore how students understand hydrogen-bonding interactions.

In addition to activating the appropriate resources, we found that the framing of the task impacted what students chose to include in their written explanations. For example, we wanted students to explain *how* and *why* the entities interacted with one another instead of just describing *that* they interacted with one another, so we modified the task to cue the students to provide such information. However, we needed to include just enough scaffolding to (1) activate the productive resources necessary for CMR and (2) provide enough information about which ideas we expected students to include in their response, without overspecifying.

To design this task, we had to balance multiple factors. For example, instead of probing all the ideas relevant to protein—ligand binding (such as shape or entropy), we chose to ask students about a simplified version of this phenomenon centered around the role of electrostatic attraction. We hope in the future we can ask students, particularly in upper-level courses, about more complex phenomena, but we know students struggle to understand the electrostatic basis of these noncovalent interactions.[17,25−27] These ideas are hard, and it would not be productive to overwhelm the students for the sake of scientific accuracy. At the same time, we were also cognizant of the potential to overcue students, which may encourage more rote learning behaviors.[34] How then can we provide just the right amount of support to students? It is not easy to strike this balance, but we argue that iterative design and being responsive to what our students say and do should play an important role in this process.

Once the final task version was determined, we began developing a coding scheme including both analytic and holistic components which can be used to characterize responses. We found that beginning with an analytic rubric allowed for more structure and detail in the initial stages of characterizing the data. The final analytic rubric, which was the result of several iterations, consists of three key ideas which were distilled both from our ideal causal mechanistic explanation as well as the ideas students included in their responses. From these three key ideas, holistic codes emerged based on evidence of the number of key ideas in the response. Thus, we could characterize responses as non-CM, partially CM, or fully CM in this context of protein—ligand binding. This process of designing a coding approach with both analytic and holistic aspects was insightful, and the approach will be used to characterize chemistry and biology students' responses to the PL task in forthcoming publications.

Constructing and evaluating tasks in this way is philosophically different than many approaches to designing assessments, in which students may not be provided with scaffolded cues. Our goal is to determine what it is that students know and can do, rather than letting them rely on memorization or heuristics which may not be backed by a robust understanding of the underlying concepts. Designing and using such tasks is crucial if we are to help students to go beyond rote memorization of isolated facts; we need to both understand how students engage in sophisticated forms of reasoning (like CMR) and provide opportunities on assessments for students to use those types of reasoning. Furthermore, by including questions which invoke CMR on assessments, we send a message to the students that understanding *how* and *why* phenomena occur is important and valued.[35]

This work is a step toward our larger goal of supporting students' interdisciplinary understanding of science and their ability to use CMR. While in this paper we share a developed task and coding scheme to assess explanations of protein—ligand binding, our team has also developed materials in the contexts of protein structure—function and phenotypic variation which we report out in future publications.[36] In forthcoming studies, we plan to use these tasks and coding schemes to explore how undergraduate students respond at different time points in their science degree programs and in the context of different disciplines. By using these tasks, we hope to gain a better understanding of how students may or may not be engaging in CMR and to develop more effective ways to support CMR both within and across disciplines, with the ultimate goal of providing opportunities for students to make meaningful connections between chemistry and biology and develop more powerful reasoning strategies in science.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.1c00959.

Description of the analytic rubric development process, details on the alternate task administration, the full responses used to develop the PL task, screenshots of each version of the activity, and examples of students discussing hydrogen bonding (PDF, DOCX)

AUTHOR INFORMATION

Corresponding Author

Keenan Noyes − *Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States;* orcid.org/0000-0002-8587-1694; Email: noyeskee@msu.edu

Authors

Clare G. Carlson − *Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States;* orcid.org/0000-0003-1778-8407

Jon R. Stoltzfus − *Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, United States*

Christina V. Schwarz − *College of Education, Michigan State University, East Lansing, Michigan 48824, United States*

Tammy M. Long − *Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824, United States*

Melanie M. Cooper − *Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States;* orcid.org/0000-0002-7050-8649

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jchemed.1c00959

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are very appreciative of all the past and present members of our interdisciplinary team and Juli Uhl for providing thoughtful feedback on this project over the years. We would also like to thank all the students who participated in this study for taking the time to complete our activity and entrusting us with their responses. This project would not have been possible without them. This work is supported by the National Science Foundation under DUE 1725521.

REFERENCES

(1) Krist, C.; Schwarz, C. V.; Reiser, B. J. Identifying Essential Epistemic Heuristics for Guiding Mechanistic Reasoning in Science Learning. *Journal of the Learning Sciences* **2019**, *28* (2), 160−205.

(2) Hammer, D.; Berland, L. K. Confusing Claims for Data: A Critique of Common Practices for Presenting Qualitative Research on Learning. *Journal of the Learning Sciences* **2014**, *23* (1), 37−46.

(3) Russ, R. S.; Scherr, R. E.; Hammer, D.; Mikeska, J. Recognizing Mechanistic Reasoning in Student Scientific Inquiry: A Framework for Discourse Analysis Developed from Philosophy of Science. *Science Education* **2008**, *92* (3), 499−525.

(4) Hammer, D. Student Resources for Learning Introductory Physics. *American Journal of Physics* **2000**, *68* (S1), S52−S59.

(5) Hammer, D.; Elby, A.; Scherr, R. E.; Redish, E. F. Resources, Framing, and Transfer. In *Transfer of learning from a modern multidisciplinary perspective*; IAP: Greenwich, CT, 2005; Chapter 3, pp 89−120.

(6) Smith, J. P. I.; diSessa, A. A.; Roschelle, J. Misconceptions Reconceived: A Constructivist Analysis of Knowledge in Transition. *Journal of the Learning Sciences* **1994**, *3* (2), 115−163.

(7) National Research Council. *Knowing What Students Know: The Science and Design of Educational Assessment*; National Academies Press: Washington, DC, 2001.

(8) Mislevy, R. J.; Almond, R. G.; Lukas, J. F. A Brief Introduction to Evidence-Centered Design. *ETS Research Report Series* **2003**, *2003* (1), i−29.

(9) Wood, D.; Bruner, J. S.; Ross, G. The Role of Tutoring in Problem Solving*. *Journal of Child Psychology and Psychiatry* **1976**, *17* (2), 89−100.

(10) Vygotsky, L. S. *Mind in Society: The Development of Higher Psychological Processes*; Harvard University Press, 1980.

(11) Bruner, J. S. *Actual Minds, Possible Worlds*; Harvard University Press: Cambridge, Mass, 1986.

(12) Kang, H.; Thompson, J.; Windschitl, M. Creating Opportunities for Students to Show What They Know: The Role of Scaffolding in Assessment Tasks. *Science Education* **2014**, *98* (4), 674−704.

(13) Caspari, I.; Kranz, D.; Graulich, N. Resolving the Complexity of Organic Chemistry Students' Reasoning through the Lens of a Mechanistic Framework. *Chemistry Education Research and Practice* **2018**, *19* (4), 1117−1141.

(14) Graulich, N.; Caspari, I. Designing a Scaffold for Mechanistic Reasoning in Organic Chemistry. *Chemistry Teacher International* **2021**, *3* (1), 19−30.

(15) Graulich, N.; Schween, M. Concept-Oriented Task Design: Making Purposeful Case Comparisons in Organic Chemistry. *J. Chem. Educ.* **2018**, *95* (3), 376−383.

(16) Bishop, B. A.; Anderson, C. W. Student Conceptions of Natural Selection and Its Role in Evolution. *Journal of Research in Science Teaching* **1990**, *27* (5), 415−427.

(17) Noyes, K.; Cooper, M. M. Investigating Student Understanding of London Dispersion Forces: A Longitudinal Study. *J. Chem. Educ.* **2019**, *96* (9), 1821−1832.

(18) Jin, H.; Anderson, C. W. *Developing Assessments for a Learning Progression on Carbon-Transforming Processes in Socio-Ecological Systems*; Brill, 2012; pp 149−181.

(19) Brandriet, A.; Rupp, C. A.; Lazenby, K.; Becker, N. M. Evaluating Students' Abilities to Construct Mathematical Models from Data Using Latent Class Analysis. *Chem. Educ. Res. Pract.* **2018**, *19* (1), 375−391.

(20) Cooper, M. M.; Stowe, R. L.; Crandell, O. M.; Klymkowsky, M. W. Organic Chemistry, Life, the Universe and Everything (OCLUE): A Transformed Organic Chemistry Curriculum. *J. Chem. Educ.* **2019**, *96* (9), 1858−1872.

(21) Cooper, M.; Klymkowsky, M. Chemistry, Life, the Universe, and Everything: A New Approach to General Chemistry, and a Model for Curriculum Reform. *J. Chem. Educ.* **2013**, *90* (9), 1116−1122.

(22) Bryfczynski, S. BeSocratic: An Intelligent Tutoring System for the Recognition, Evaluation, and Analysis of Free-Form Student Input. Ph.D. Dissertation, Clemson University, 2012.

(23) Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* **1960**, *20* (1), 37−46.

(24) Landis, J. R.; Koch, G. G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33* (1), 159−174.

(25) Cooper, M. M.; Corley, L. M.; Underwood, S. M. An Investigation of College Chemistry Students' Understanding of Structure−Property Relationships. *Journal of Research in Science Teaching* **2013**, *50* (6), 699−721.

(26) Cooper, M. M.; Williams, L. C.; Underwood, S. M. Student Understanding of Intermolecular Forces: A Multimodal Study. *J. Chem. Educ.* **2015**, *92* (8), 1288−1298.

(27) Williams, L. C.; Underwood, S. M.; Klymkowsky, M. W.; Cooper, M. M. Are Noncovalent Interactions an Achilles Heel in Chemistry Education? A Comparison of Instructional Approaches. *J. Chem. Educ.* **2015**, *92* (12), 1979−1987.

(28) Henderleiter, J.; Smart, R.; Anderson, J.; Elian, O. How Do Organic Chemistry Students Understand and Apply Hydrogen Bonding? *J. Chem. Educ.* **2001**, *78* (8), 1126.

(29) Peterson, R. F.; Treagust, D. F.; Garnett, P. Development and Application of a Diagnostic Instrument to Evaluate Grade-11 and −12 Students' Concepts of Covalent Bonding and Structure Following a

Course of Instruction. *Journal of Research in Science Teaching* **1989**, *26* (4), 301−314.

(30) Jescovitch, L. N.; Scott, E. E.; Cerchiara, J. A.; Merrill, J.; Urban-Lurain, M.; Doherty, J. H.; Haudek, K. C. Comparison of Machine Learning Performance Using Analytic and Holistic Coding Approaches Across Constructed Response Assessments Aligned to a Science Learning Progression. *J. Sci. Educ Technol.* **2021**, *30* (2), 150−167.

(31) Becker, N.; Noyes, K.; Cooper, M. Characterizing Students' Mechanistic Reasoning about London Dispersion Forces. *J. Chem. Educ.* **2016**, *93* (10), 1713−1724.

(32) Cooper, M. M.; Kouyoumdjian, H.; Underwood, S. M. Investigating Students' Reasoning about Acid−Base Reactions. *J. Chem. Educ.* **2016**, *93* (10), 1703−1712.

(33) Crandell, O. M.; Kouyoumdjian, H.; Underwood, S. M.; Cooper, M. M. Reasoning about Reactions in Organic Chemistry: Starting It in General Chemistry. *J. Chem. Educ.* **2019**, *96* (2), 213−226.

(34) Torrance, H. Assessment as Learning? How the Use of Explicit Learning Objectives, Assessment Criteria and Feedback in Post-secondary Education and Training Can Come to Dominate Learning. *Assessment in Education: Principles, Policy & Practice* **2007**, *14* (3), 281−294.

(35) Shepard, L. A. The Role of Assessment in a Learning Culture. *Educational Researcher* **2000**, *29* (7), 4−14.

(36) Schwarz, C.; Cooper, M.; Long, T.; Trujillo, C.; Noyes, K.; de Lima, J.; Kesh, J.; Stolzfus, J. Mechanistic Explanations Across Undergraduate Chemistry and Biology Courses. In *ICLS 2020 Proceedings*; International Society of the Learning Sciences: Nashville, TN, USA, 2020; pp 625−628.