JOURNAL OF CHEMICAL EDUCATION

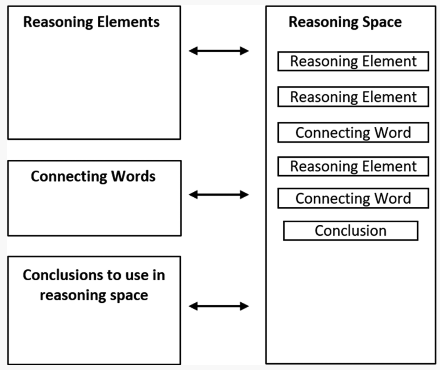
pubs.acs.org/jchemeduc Article

Implementation of Reasoning Chain Construction Tasks to Support Student Explanations in General Chemistry

Megan L. Nagel* and Beth A. Lindsey

Cite This: https://doi.org/10.1021/acs.jchemed.1c00729 Read Online

ABSTRACT: The assessment of student understanding, and of student reasoning skills more broadly, hinges upon the ability to elicit and interpret student-generated explanations. In practice, however, the explanations that students provide in response to traditional prompts do not always reveal as much about student reasoning as teachers and researchers in chemistry education might prefer. In this article, we describe the application of a novel methodology, the *reasoning chain construction task*, applied in a general chemistry course at a large research institution using a well-known question on chemical bonding. In a reasoning chain construction task, students respond to a chemistry question by drawing from a list of reasoning elements (all of which are true) in order to assemble a chain of reasoning in support of a conclusion. Our findings indicate that use of this type of task can lead students to generate richer explanations while not impacting the overall distribution of student conclusions. These explanations provide teachers and researchers with insights into what information students deem helpful and relevant for responding to a prompt that is wholly lacking from traditional free-response implementations of the same question. We interpret our results through the theoretical framework of dual-process theories of reasoning.



KEYWORDS: *First-Year Undergraduate/General, Chemical Education Research, Problem Solving/Decision Making, Covalent Bonding*

FEATURE: Chemical Education Research

■ INTRODUCTION

The ability to use foundational principles and evidence to reason about chemical phenomena is an essential skill for practitioners of chemistry. For experts, this process is natural, as they have both accumulated the necessary knowledge and also organized this knowledge into coherent and interconnected frameworks.[1] For students, however, this knowledge is often fragmented, making it difficult for them to assemble disparate facts into a coherent logical argument in support of an explanation or prediction.[2,3] As such, there is widespread agreement that science education should incorporate, as a staple, the instructional practice of generating causal explanations.[4−7] Beyond learning the practice itself, student participation in the development of scientific explanations is shown to have additional benefits, including increased engagement with the material and improved learning.[8−10] Not only do students benefit from the act of creating these explanations but instructors can also use them to provide insight into how students reason about various phenomena and tailor their instruction appropriately.[11]

Despite the numerous educational benefits associated with students generating explanations, incorporating opportunities for them to do so can present numerous practical challenges. Writing a good question is hard—in particular, structuring questions such that it is clear to students that they are expected to use logical reasoning to explain their answer is nontrivial.[12−14] Evaluating, analyzing, or even drawing meaning from the responses that students provide can also be difficult.[15] The current methods for eliciting explanations or exploring student reasoning patterns generally fall into two broad categories, both of which have affordances and limitations.

The first category of questions for investigating student reasoning involves an open-ended (or free-response) design. These tasks offer students the freedom to express explanations without necessarily constraining their thought process or encouraging a particular reasoning path.[16] Novice learners who have not already been trained in the skill of generating causal explanations, however, will be unlikely to do so spontaneously simply because they are told to "explain your reasoning".[17] The responses may thus ultimately provide limited evidence into the thought patterns of students. Furthermore, the analysis of such responses can be time-consuming, and it may be difficult to observe the emergence of clear response patterns without

Received: July 2, 2021
Revised: October 29, 2021

making assumptions or applying personal interpretation to student statements.

The second question category used to study student reasoning is a forced-choice option. These may be traditional multiple-choice questions or two-tiered questions in which students first select an answer from among the available options and then select an explanation from another list of multiple-choice options.[18−20] This method allows for a very simple categorization and analysis of student responses. Unlike in an open-ended question, students are limited to a small number of possible responses. These questions do not require students to generate their own explanation but rather to select an explanation, increasing the likelihood that the provided choices may cue a particular line of reasoning. The provided explanation options may even include false information, thus potentially providing students with misleading ideas about what the appropriate reasoning could be.

In this paper, we present the reasoning chain construction task, a novel question format recently introduced in the Physics Education Research literature.[21] These tasks provide a new option for eliciting and analyzing student reasoning patterns in chemistry. In a reasoning chain construction task, students are able to freely generate a line of causal reasoning by drawing from a limited suite of accurate information in the form of reasoning statements that are provided to them.

In its general form, the reasoning chain construction task is not domain or topic specific. In this study, however, we apply this methodology to a question involving energy changes associated with chemical bonding. This topic has received significant attention in the chemical education literature, including well-documented student difficulties and research-based frameworks to support student explanations.[22−25] The notion that bond breaking is an exothermic process has been shown to be persistent in numerous contexts.[26] One factor that has been implicated in perpetuating this belief is the common description of energy as being "stored in bonds".[27] While this description of two atoms interacting is not incorrect, it can be misleading, even for students who have had the relevant instruction on electrostatic interactions, forces, and the changes in potential energy that are associated with bonding. This topic is thus one that deserves continued attention. The ability to examine in more detail what knowledge students rely upon to describe energy changes observed during bonding can help refine instruction and could point to potential pitfalls for educators to avoid. Guided by previous work on this topic, we offer an example of the utility of the reasoning chain construction task in the context of chemical bonding. More broadly, we demonstrate that this question type can provide a useful alternative to traditional methods for the purposes of both research and instruction in general chemistry.

Research Questions

Herein, we describe student responses to a question on chemical bonding when presented in a reasoning chain construction task format and compare them to responses from a traditional multiple-choice with free-response format. Because the reasoning chain construction task is specifically designed to explore patterns of students reasoning by providing students with several true statements, or "reasoning elements", from which they must select to support their answer to a question, we also seek to specifically explore the selections made by students in building their reasoning chains. We use dual-process theories of reasoning as a theoretical lens to guide our analyses. Overall, we aim to address the following questions:

1. What differences in causal explanations and overall student performance are observed for students receiving the traditional question format versus the reasoning chain construction task?
2. When reasoning elements are provided in the form of a reasoning chain construction task, how are the elements students select related to their performance on the question?

Theoretical Framework: Dual-Process Theories of Reasoning

Studying student-generated explanations to chemistry questions can provide useful insights for adapting and modifying instruction. How a student makes sense of, or reasons about, their response to a question can be used to inform both the specific feedback provided to that student and future instruction on that topic to other students. As such, this work is motivated by the desire to understand how students are reasoning about questions in introductory chemistry. While the explanation a student provides in response to a task may give some window into their thinking, the process of human reasoning may actually be much more complex than suggested by written responses to questions. Student reasoning patterns are not always consistent with the notion that, when presented with a novel question, a student will begin from first-principles or a content conception and reason logically to reach their conclusion.[28] Dual-process theories of reasoning[29,30] provide a framework emerging from the field of cognitive psychology that has recently been used to help predict and explain student reasoning patterns[31,32] as well as to inform the development of instructional interventions.[33,34] In this section, we provide an overview of these theories.

Dual-process theories model human cognition by proposing two separate processes by which all reasoning and decision-making occurs. The first, Process 1 or the "heuristic process", is engaged for quick and intuitive decision making, such as responding to "what is 2 + 2?". The second, known as Process 2 or the "analytic process", is more slow and effortful and may be engaged for more complex tasks, such as calculating 17×84. One example of a dual-process theory, the heuristic-analytic theory of reasoning put forward by Evans,[29] describes the interactions between these two processes. When reasoning about a novel situation, Process 1 will immediately put forth a mental model (the "first available mental model" or "intuitive model"). Evans explains that the heuristic process will generate one mental model at a time (the singularity principle). This model may suggest a particular response. The first-available mental model put forward by Process 1 will be heavily influenced by a student's prior knowledge and beliefs and also by contextual cues (the relevance principle). For instance, the presence or absence of a highly salient, irrelevant feature in the representation of the task at hand can influence the response generated by Process 1.[35] The task of the analytic process is to evaluate this model by ascertaining whether or not it is satisfactory for the task at hand (the satisficing principle). If the analytic process detects a reason to reject the initial model, Process 1 will put forth a new mental model and the cycle will repeat.

As a matter of practicality for minimizing cognitive load and freeing working memory, much of adult decision-making engages only Process 1. In many cases, the analytic process does not engage at all. Even if it does engage, the analytic process is also subject to several reasoning biases, such as a confirmation bias and other reasoning shortcuts. It will tend to accept the response generated by the heuristic process unless a "red flag" is raised, meaning that the analytic process detects a compelling reason to reject the intuitive mental model. In many cases, no such "red flag" is perceived; thus, the response suggested by the first-available model generated by Process 1 tends to be the default response.

In summary, the dual-process theories of human reasoning suggest that content knowledge alone may not be sufficient for formal reasoning, as the intuitive appeal of an incorrect response can be difficult to overcome. In the work presented here, we use these theories not as theories of learning (although they can be used to inform the design of instructional interventions)[34] but rather to provide insights into how students construct explanations in the moment. These theories thus represent a framework through which student answer patterns can be predicted and understood.[36] Reasoning chain construction tasks, such as the ones described in this work, can be implemented without this background in dual-process theories; however, viewing student responses through this lens can help pinpoint mechanisms that contribute to incorrect answers and reasoning difficulties.

■ METHOD

Reasoning Chain Construction Tasks

In this work, we adapt a new methodology developed in the field of physics education research called a *reasoning chain construction task*, or "chaining task".[21] In a chaining task, students are presented a question and several conclusion options (similar to the answer options in a multiple-choice question). In these tasks, however, students are also provided a series of statements, referred to as reasoning elements or "tiles", which are related to the question being asked. These statements can include features of the problem itself, chemical principles, or comparisons between quantities. In all cases, the statements presented contain only true information, and the set of available statements contains all the information necessary to generate an argument in support of the correct conclusion. Not all the statements presented on the tiles, however, are necessarily relevant or useful for the task at hand. To complete a chaining task, students select and order the reasoning elements to support a conclusion. A list of conjunctions, or "connecting words" ("so", "but", "because", "and", "therefore"), is also provided to help students clarify their argument. Finally, blank tiles are available if a student wishes to generate their own reasoning element to include in the explanation.

The chaining task is implemented using the "pick/group/rank" feature in Qualtrics.[37] The spaces available to students in our chaining tasks are shown in Figure 1. Students simply click and drag their chosen reasoning elements, conjunctions, and conclusions into a blank space called the "Reasoning Space". Elements within this space can be placed in any order, and elements can be moved into and out of the "Reasoning Space" as students work on building their explanation. There is no constraint on the number of tiles that students can move into the Reasoning Space—the box expands as elements are added. Furthermore, a student may place a conclusion tile into the reasoning space at any time while answering the question. Thus, it is possible for students to first determine a conclusion and then support it with reasoning, or they may initially develop a line of reasoning that then leads to a particular conclusion.

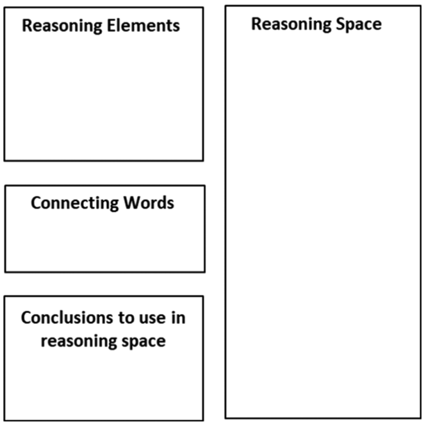


Figure 1. Basic format of a chaining question. The reasoning elements, conclusions, and connecting words spaces are prepopulated. Any of the contents of these other sections can be moved into, out of, and rearranged within the initially vacant "Reasoning Space".

Reasoning Chain Task Question Selection

The task used in our research involves the question, "When a chemical bond is "broken" between two atoms (for example two H atoms) is energy absorbed or released?"[38] (we refer to this question as "the bond-energy question"). The reasoning tiles provided to students in the chaining versions of the task are closely aligned with the tenets of the Nahum's "bottom-up" framework, which focuses on atomic-level Coulombic interactions, forces, and stability, for understanding chemical bonding.[39] Prior to the administration of the online chaining task, the statements included on the reasoning tiles were validated using a series of interviews. In these interviews, students were provided with the reasoning tiles on slips of paper and were asked to use these statements to construct a response. Students then manually constructed their argument, while discussing the decisions they were making. This was done to ensure that the text provided to students was being interpreted as intended and to verify that the tiles provided were sufficient for most students to answer the question.

Population and Sample

Data were collected for this study during the spring 2019 semester from students enrolled in General Chemistry I at a large public institution located in the mid-Atlantic region of the United States. The course is designed to primarily serve those students pursing a degree in science or engineering. All students enrolled in three sections of the course were invited to participate in the study ($N = 640$). The questions were made available to students on an assignment during the 14th week of a 15 week semester. The assignment was hosted on Qualtrics, and students accessed it through a link on the course management system. Students received a small amount of course credit (<1% of the overall grade) for completing the assignment. While the three sections were taught by different instructors, the same text, homework, recitations materials, and exams were used in all sections. As such, for the purposes of this study, the students were treated as a single population. All research activities were deemed "exempt" by the Institutional Review Board, and every participant involved in the study explicitly provided informed consent. Of the 512 students who

https://doi.org/10.1021/acs.jchemed.1c00729
J. Chem. Educ. XXXX, XXX, XXX−XXX

completed the assignment, 463 consented to have their results included, representing 90% of all those who completed the assignment and 72% of all students enrolled in the course.

Task Design

Three different versions of the assignment were randomly distributed to students (Table 1). Each version consisted of two separate tasks, each presented on a separate page of the online assignment (see the Supporting Information). Every participant received the same first task. This initial task was written in the chaining format, and it was intended to be a simple chemistry question that would provide participants some experience in using the chaining format. Results from this task were not analyzed as it was intended to serve solely as a warm-up activity. Students were then randomly directed to one of three different versions of the research task (Version A, Version B, or Version C) using Qualtrics randomization features. In each version, students were asked to respond to the bond-energy question, but the details of how students were asked to respond to this question differed between the three versions, as described below and summarized in Table 1. In all cases, students could not go back to a previous page within the assignment nor could they continue to the next page if they had not submitted a response to each question. In Version A, the bond-energy question was presented in a multiple-choice format with a free-response explanation. Students were asked to select their answer from among the provided answer options

Table 1. Description of Assignment Versions That Were Randomly Distributed to Students[a]

Assignment Version	Description
Version A (N = 157)	Multiple choice questions with explanation given as free response
Version B (N = 150)	Chaining task including reasoning elements
Version C (N = 154)	Chaining task including reasoning elements plus an additional tile, "A system of interacting charges can be considered to store energy."

[a]Numbers in in parentheses reflect the number of students in each version who consented to be included in the study.

Reasoning Tiles
A bond represents a potential energy minimum.
The potential energy of a system of two attracting particles increases as their separation increases.
The distance between two atoms must be increased to "break a bond."
$U_e = k\frac{q_1 q_2}{d}$
$U_g = mgh$
A system of interacting charges can be considered to store energy.
The attractive force between two atoms decreases as their separation increases.
Energy is required to move an object to a less stable position.
Bond length represents a distance between atoms where stability is at a maximum.
Bonded atoms are most stable when their attractive forces balance their repulsive forces.
In a bond, $F_{net} = 0$
In an endothermic process, heat is absorbed.
In an exothermic process, heat is released.

Connecting Words
and
but
because
so
therefore

Required Conclusion
Energy is absorbed.
Energy is released.
Energy is both absorbed and released.
Energy is neither absorbed nor released.
There is not enough information to tell.

Figure 2. Reasoning tiles provided for chaining versions of the task. The bolded statement was included only in Version C. Students determined which of these tiles they moved into the vacant Reasoning Space.

https://doi.org/10.1021/acs.jchemed.1c00729
J. Chem. Educ. XXXX, XXX, XXX−XXX

(see "Required Conclusion" Figure 2). Students were then asked to explain their reasoning for the answer option they had chosen in an accompanying text box which allowed for unlimited text.

In Versions B and C, the bond-energy question was presented in the form of a chaining task. The allowed conclusions for each task were the same as in Version A. Students were asked to construct the explanation for their response using the items shown in the "Reasoning Tiles" box of Figure 2. The only difference between Versions B and C of the task was a single reasoning tile; Version C included a tile that stated "A system of interacting charges can be considered to store energy", which was absent in version B. We included this tile as we felt that it was a common description of potential energy that might be appealing to students in building their reasoning chains, and we were interested in observing whether there were notable differences in student reasoning chains when it was present. For the chaining versions of the task, once students had generated their reasoning chain, they indicated that their chain was complete, and they were then prompted to confirm that a conclusion tile had been included in their reasoning chain. Students were instructed to include a single conclusion tile (and only one) in the reasoning space; however, this could not be enforced within the chaining format. Thus, a small number of students included no conclusion or included multiple conclusion tiles. Only those students who placed a single conclusion tile in their reasoning space were included in the analysis.

The correct conclusion to the question is that when a bond is broken between two atoms, energy is absorbed. There are many ways of justifying this conclusion, but based on the tiles that are provided in the chaining version of the task, a reasonable correct chain could be

- Bond length represents a distance between atoms where stability is at a maximum.
- The distance between two atoms must be increased to "break a bond".
- But
- Energy is required to move an object to a less stable position.
- Therefore
- Energy is absorbed.

This reasoning chain requires only three reasoning elements, two connecting words, and a conclusion. Yet, it demonstrates an understanding of *why* breaking a bond is an endothermic process. This chain is not merely a memorized heuristic, but rather, it draws on the relationship between stability and energy as a means of justifying the correct conclusion.

Data Analysis

In order to effectively evaluate the differences between causal explanations that students generate in the chaining format and those that they generate in more traditional question formats (research question 1), the open-ended responses collected in Version A of the assignment were coded according to their alignment with the reasoning elements presented in Version C of the assignment. During the coding process, it was noted that students were occasionally using an argument about bond dissociation energy that was not included among the available reasoning tiles. Therefore, an additional code for any mention of bond dissociation energy was also included but ultimately only two responses received this code. Two different researchers coded the answer responses independently. Any discrepancies in the applied codes were discussed until there was 100% agreement.

In Version A, the MCwFR version, students were required to choose a single conclusion option before they could proceed. Unlike Version A, the chaining task design in Versions B and C does not require students to choose a single conclusion element. Instead, students were instructed to select a single conclusion element (and only one) and place it in the reasoning space, along with reasoning tiles that support the answer. Given the design of the question, the selection of a single conclusion could not be enforced. Of the students in this data set, 22 (7%) who received a chaining assignment (Version B or C) did not select a conclusion element or even a single reasoning element. This evidence suggests that these students had little interaction with the task beyond just clicking through to the end. As a result, this group of students is not included in the analysis. An additional 29 students (10%) did not provide a conclusion element but did move at least one reasoning element into the reasoning space. Eight students (2%) selected every tile and moved it into the reasoning space. Finally, two students (0.6%) selected more than one conclusion element. Removing all these problematic cases from Versions B and C leaves 116 and 127 participants, respectively, across those two versions of the assignment (80% of consented respondents in those two versions). The remaining analyses are based on this set of student responses.

■ RESULTS

Across all three versions, the majority of students (69%) were able to correctly identify that breaking a bond requires an absorption of energy. The breakdown by specific answer and version is presented in Table 2. Across both chaining versions, the correct conclusion was selected by 173 students (71%) and an incorrect response was chosen by 70 respondents (29%). A χ^2 test of independence shows that there is no significant association between any of the versions and selection of the correct or incorrect conclusion, χ^2 (2, N = 400) = 0.5773, p = 0.749. Overall, the proportion of students answering this question correctly is larger than reported in previous studies employing a similar question.[38,40]

Table 2. Student Conclusion Selection for the Bond-Energy Question, Based on Assignment Version

Provided Conclusions	Version A (n = 157); %	Version B (n = 116); %	Version C (n = 127); %
Energy is absorbed	68	72	70
Energy is released	26	21	23
Energy is both absorbed and released	6	3	6
Energy is neither absorbed nor released	0	2	1
There is not enough information to tell	0	2	0

For Version A, a large majority of students did not include any substantive reasoning to support their selected conclusion, even though all students had entered text, as required, in the text box. In many cases, students simply restated information provided in the question or reworded the selected conclusion. This pattern was observed regardless of whether students answered with a correct or incorrect conclusion. Examples of characteristic responses are given in Table 3.

https://doi.org/10.1021/acs.jchemed.1c00729
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 3. Common Reasoning Responses for the Free-Response Bond-Energy Question

Reasoning Used to Support Answer	Percent of Students	Representative Examples
None	71	Energy needs to be absorbed because energy needs to be put into a molecule to break its chemical bonds.
		Energy is absorbed because the molecule needs energy to break its bond.
In an endothermic process, heat is absorbed.	13	It takes energy to break bonds and breaking bonds is an endothermic reaction so energy would be absorbed
In an exothermic process, heat is released.	6	When a bond is broken it is exothermic

For those students who did provide some additional support for their answer, it was essentially a definition of the terms endothermic or exothermic (13%). Very few students used ideas of stability (<3%), potential energy (<3%), or even the change in interatomic distance (<2%) to justify why energy is absorbed during bond breaking. It should also be noted that, while the statements from the reasoning tiles were initially employed to guide the coding process, they were ultimately applied very generously. For example, any explanation that included the word "exothermic" was coded as using "In an exothermic process, heat is released."

For each version, the count of explanatory ideas provided by the student was quantified and is referred to as the reasoning "chain length". For Version A, the chain length was determined using the coding scheme developed for the short answer responses. On the basis of this method, the average reasoning chain for Version A was calculated to be 0.35 (std dev 0.58) elements long. For Versions B and C, the reasoning chain length is a tally of the reasoning tiles moved by each student into the reasoning space (not including their conclusion or any connecting word tiles). Figure 3 illustrates four representative final reasoning spaces. The average number of reasoning tiles moved into the reasoning space was found to be 2.78 (std dev 2.19) and 2.47 (std dev 1.72) for versions B and C, respectively. A full distribution of reasoning chain lengths for

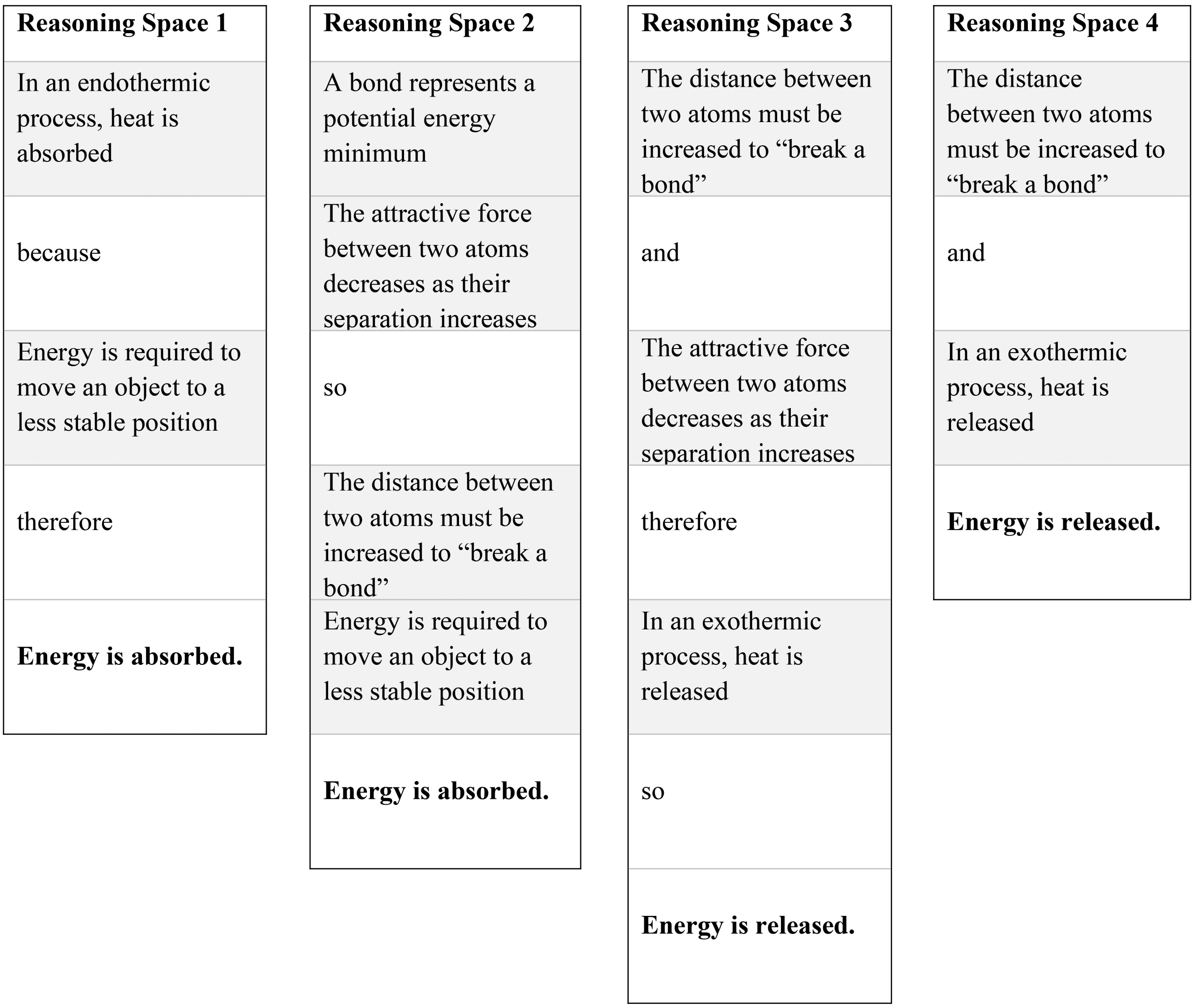


Figure 3. Final "reasoning space" selections from four student responses. Reasoning tiles are represented in gray boxes, connecting words are in white boxes, and the selected conclusions is in bold. Reasoning Spaces 1 and 2 represent chains built to support a correct conclusion, and Reasoning Spaces 3 and 4 are chains built to support an incorrect conclusion.

https://doi.org/10.1021/acs.jchemed.1c00729
J. Chem. Educ. XXXX, XXX, XXX−XXX

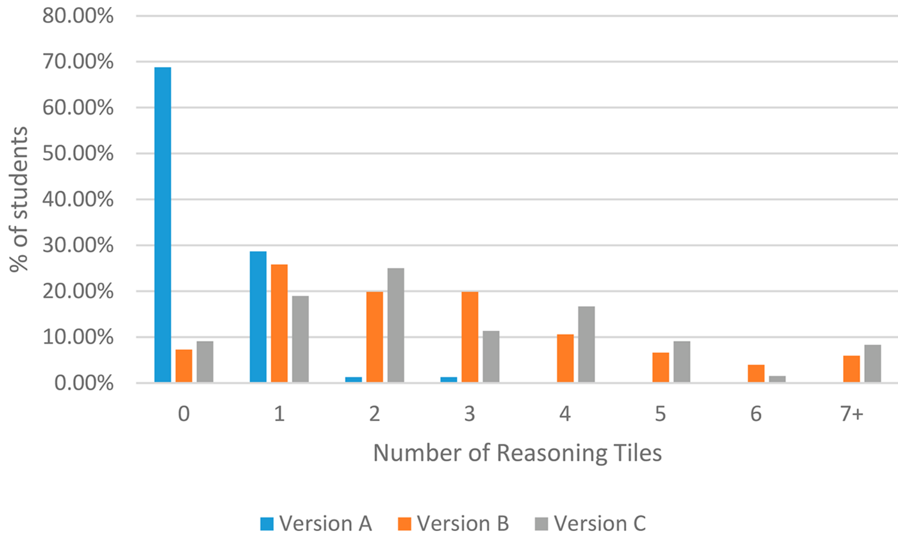


Figure 4. Frequency of reasoning chain length as a function of assignment version.

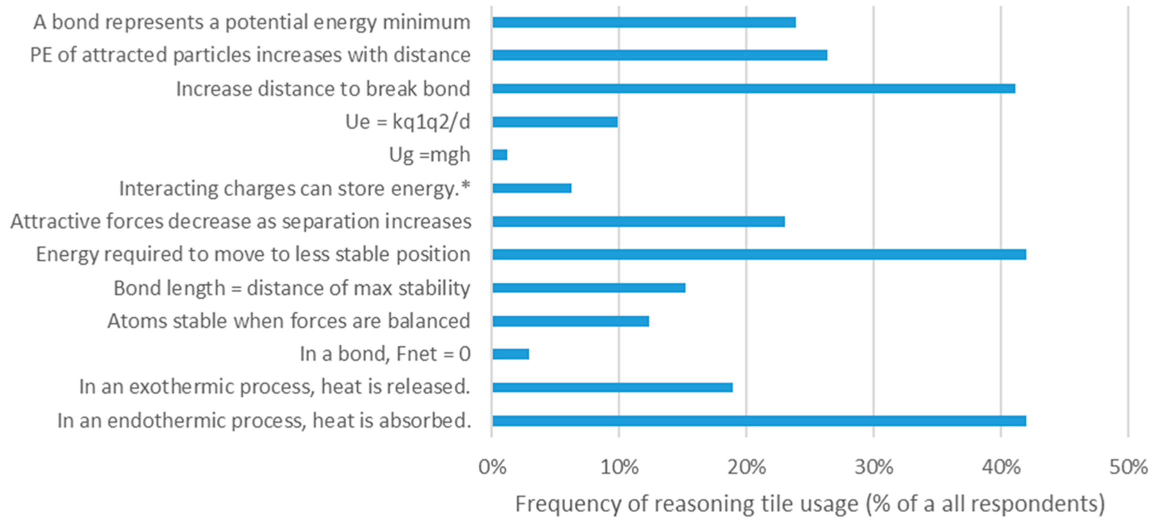


Figure 5. Percentage of students who selected each reasoning tile for use in their reasoning chain (*calculated from only those students who received this tile in Version C).

each version is shown in Figure 4. An independent-samples Kruskal–Wallis test (a nonparametric version of ANOVA used to test for differences between more than two groups in the case of a noncontinuous dependent variable) confirms that there is a statistical difference with a large effect size (measured by η^2) between the chain lengths used by students across the three versions ($H(2) = 186$; $p < 0.001$, $\eta^2 = 0.463$). Posthoc testing indicates that those students who answered the question in Version A provided less reasoning support for their answers than those students who received the chaining version of the questions. There is no statistical difference between the chain lengths provided by students responding to the two chaining versions.

The average number of reasoning tiles selected by students who chose the correct conclusion was 2.91 ($n = 159$, SD = 1.85). The average reasoning chain length for students selecting an incorrect conclusion was 2.92 ($n = 60$, SD = 1.85). A two-tailed t-test with equal variances indicates that there is no difference between the average number of tiles selected on the basis of the correctness of the conclusion selected ($t(217) = 1.97$, $p = 0.968$).

For Versions B and C, we also explored how often each reasoning tile was used. The frequency with which each tile was added to the reasoning space is shown in Figure 5. Overall, three reasoning tiles emerged as being used more than the others. These tiles were "Energy is required to move an object to a less stable position" ($n = 102$, used in 42% of chains), "In an endothermic process, heat is absorbed" ($n = 102$, 42%), and "The distance between two atoms must be increased to break a bond" ($n = 100$, 41%). The next most commonly used reasoning element, "The potential energy of a system of two attracting particles increases as their separation increases", was selected by only 26% of respondents ($n = 64$).

With few exceptions, students used tiles with the same frequency regardless of whether the conclusion they selected was correct or incorrect. Figure 6 illustrates the frequency with which each tile was used in support of both correct and incorrect conclusions. A Fisher's Exact test was performed for each reasoning tile to investigate whether the use of a particular tile was associated with the selection of the correct conclusion. The Bonferroni correction was applied to reduce the likelihood of obtaining a spurious significant result with

https://doi.org/10.1021/acs.jchemed.1c00729
J. Chem. Educ. XXXX, XXX, XXX−XXX

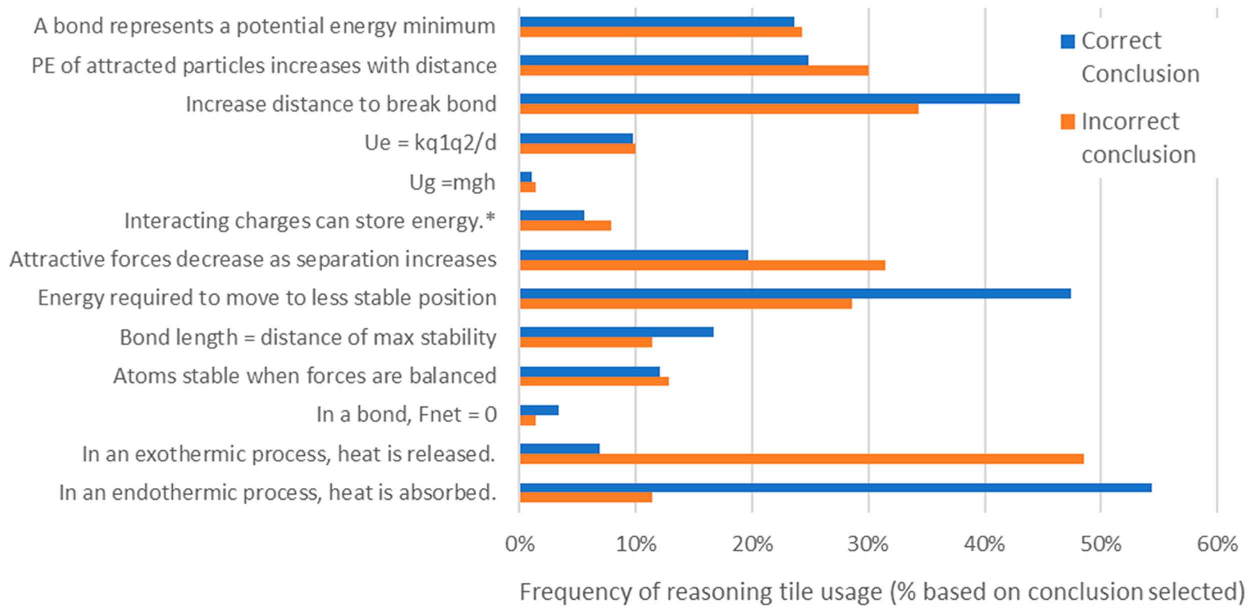


Figure 6. Percentage of students using each reasoning tile on the basis of whether a correct or incorrect conclusion was selected (*calculated from only those students who received this tile in Version C).

multiple tests. As there were 13 tiles (i.e., 13 tests were conducted), $p < 0.05/13 = 0.0038$ was used as the significance threshold. Only two tiles show an apparent difference in the frequency of selection, based on the chosen conclusion tile: "In an exothermic process, heat is released" and "In an endothermic process, heat is absorbed". The "exothermic process" tile was used by only 7% of students selecting a correct conclusion, while 49% of students with an incorrect conclusion included this tile in their reasoning chain ($p < 0.000\,01$, OR = 12.67). In contrast, the "endothermic process" tile was used by 54% of students with a correct conclusion and only 11% of student with an incorrect conclusion ($p < 0.000\,01$, OR = 9.22). In both cases, these differences represent a large effect size, with students who used the tile giving the logically related conclusion. A third tile, "Energy is required to move an object to a less stable position", appeared to show disproportionate use between students giving correct and incorrect conclusions. Nearly half of students who selected the correct conclusion also selected this reasoning element (47%), whereas less than one-third (29%) of students with an incorrect conclusion used this tile, but this difference did not quite rise to the level of statistical significance once the Bonferroni correction was applied ($p = 0.0095$, OR = 1.56).

■ DISCUSSION

In this work, we set out to answer two research questions. We organized our discussion section in response to these questions.

Research Question 1: What Differences in Overall Student Performance and Causal Explanations Are Observed for Students Receiving the Traditional Question Format versus the Reasoning Chain Construction Task?

No differences in response distributions were observed between groups of students receiving different versions of the assignment. The presence of accurate information in the chaining task does not result in an overall change in answer distribution. Even though students responding to the chaining task versions of the question (Versions B and C) were presented all the information necessary to support an accurate response, the fraction of students responding correctly was virtually identical across the three assignment versions. This finding is consistent with expectations from the dual-process theories of reasoning and reinforces similar results from chaining tasks reported in the physics literature.[21] In the heuristic-analytic theory of reasoning as presented by Evans,[29] the first available mental model can be highly resistant to change, especially when it is rooted in a commonly held belief such as the notion that bond breaking is an exothermic process.[26] According to the singularity principle, students will tend to formulate a single hypothesis and only abandon this belief if they encounter evidence that forces them to consider an alternative.[41] In the case of a chaining task, simply reading correct and relevant information does not appear to support additional engagement of the analytical process in a way that allows students to reconsider their initial response. Students may not be reflecting carefully on the provided statements or recognizing that some conflict with their first-available mental model. Previous results in PER demonstrated that student response patterns were only altered when a chaining task included information that explicitly refuted the common incorrect response.[42] Results reported here, with a chaining task that made no specific effort to refute the common incorrect model, are wholly consistent with that finding. As a result, simply casting chemistry questions as reasoning chain construction tasks does not appear to substantially alter overall response patterns.

The additional tile present in Versions C, "A system of interacting charges can be considered to store energy", had no noticeable impact on student response patterns and presented negligible differences in the reasoning chains that students constructed. We anticipated that students might be drawn to this statement that implicitly relates potential energy to "stored" energy.[40,43] This tile, however, was included by only 6% of students who were served Version C of the task, making it one of the least-used tiles in that version. Furthermore, it was included at a similar frequency in support of both correct and incorrect conclusions. These results indicate that, when

https://doi.org/10.1021/acs.jchemed.1c00729
J. Chem. Educ. XXXX, XXX, XXX−XXX

available in a chaining task format, students do not find the idea that energy is "stored" to be a particularly useful piece of information.

There was an important difference observed between students who responded to chaining versions of the task versus the MCwFR version. Those with the chaining tasks provided substantially more and richer reasoning than students who generated their own free-response explanation to the multiple-choice task. Free-response questions are generally considered to be a tool for gaining greater insight into student thinking and understanding. However, as documented in Table 3 and Figure 4, the explanations that students generated in their responses to Version A were very limited.

These results are again consistent with an interpretation of our data rooted in dual-process theories. According to these theories, the instructions provided to students can influence the level to which their analytic process becomes engaged.[44] The instructions to generate a chain of reasoning would be expected to engage the analytic process differently than the instruction to simply justify their response. The number of reasoning tiles selected for the chaining task versions of the question (see Figure 4) confirms that students are making a deliberate effort to engage with the task as instructed. The reasoning chain construction task provides a structured framework from which students can build an argument, allowing them the opportunity to discriminate between the different pieces of information available. However, as noted above, this framework does not in and of itself appear to assist the analytic process in overriding an incorrect default model presented by the intuitive process.

There may be several reasons why students do not spontaneously provide much information in response to this question in the free-response format. One reason is that the prompt itself may not be structured in a way that naturally invites students to provide a causal link between an atomic level phenomenon (bond breaking) and a macroscopic observation (heat transfer).[11] Alternately, the limited information provided by students in their short answer responses might suggest that students rely on memorized heuristics rather than responding by drawing on a coherent framework of chemical thermodynamics.[45] The latter interpretation would be consistent with Bodner's observation that knowledge does not equate to understanding.[2] Because most students are unlikely to have received targeted instruction and feedback on developing coherent causal explanations, from these responses alone, it is difficult to pinpoint whether students do not possess the relevant knowledge to make the argument or if they do not know how to construct their existing knowledge into a reasonable response.

The fact that students responding to chaining versions constructed explanations using substantially more information suggests that the reasoning tiles do provide students with language and structure to build a more robust response (even though the presence of reasoning tiles does not alter the overall conclusion patterns). Moreover, student tile usage on the chaining task provides information about student reasoning that is not readily available in their free-response answers, both in terms of which arguments students find appealing and which they do not. For instance, stability was not an idea that students employed in their open-ended responses. The tile, "Energy is required to move an object to a less stable position", however, is one of the most used tiles, included in reasoning chains by 42% of students. This demonstrates that arguments based on stability are in fact very appealing to students. As previously noted, the tile was used by 47% of students who gave a correct response to the task but by only 29% of students giving an incorrect response. The difference between these falls just short of statistical significance when the Bonferroni correction is applied but is suggestive of an idea that students appear to find in coherence with the conclusion that bond breaking is a process that requires energy. Conversely, the equation for calculating potential energy for a near earth system, "U_g=mgh" is the least-used reasoning tile. This shows that even though near-Earth gravitation ideas are frequently used to motivate ideas about the energetics of chemical bonding,[24] even when made available to them, these ideas are not appealing to students when they are reasoning about energy and bonding. An analysis of student tile usage in the chaining tasks provides information that is lacking from their free-response explanations. As discussed below, dual-process theories can provide us a mechanism for understanding *why* specific pieces of information are appealing to students in supporting a particular conclusion.

Research Question 2: When Reasoning Elements Are Provided, Is There a Relationship between the Elements Students Select and Their Performance on the Question?

The majority of reasoning tiles were used equally in support of both correct and incorrect answers, demonstrating that students will draw on correct information to support incorrect conclusions. Furthermore, the average chain length is approximately the same for chains supporting both correct and incorrect conclusions, suggesting that students are able to perceive relevance for their conclusion in the provided tiles, regardless of which conclusion they are supporting. For instance, the tile "the distance between two atoms must be increased to 'break a bond'", was one of the most commonly used tiles overall. This suggests that students are comfortable with the idea that atoms increase in interatomic distance as a bond breaks, but knowing this fact alone, does not give any indication of an understanding of the energy dynamics of the situation.

Dual-process theories can provide insights into how students may be engaging with reasoning tiles to support their responses. When presented with a task, a student's intuitive process immediately and automatically serves them with a model leading to a particular conclusion. However, even if the analytic process is engaged to evaluate this model, its engagement may be subject to confirmation bias, and may lead them to interpret the available pieces of information in ways that support their initial mental model. For example, the tile stating the equation for calculating electrostatic potential energy was also used at similar rates for all conclusions. Using this equation requires an understanding of negative potential energies. It has previously been shown that students frequently neglect the sign of this equation, and draw upon it to support the idea that the potential energy of a system of attracting particles will decrease as the distance between them increases.[46] Thus, this statement can be reasonably employed for supporting a range of thermodynamic conclusions, and its usage to support both correct and incorrect conclusions would be aligned with dual-process theories.

There are two reasoning tiles that provide an exception and were used preferentially in support of specific conclusions. The "endothermic process" reasoning tile was used primarily by students who gave the conclusion that energy is absorbed.

https://doi.org/10.1021/acs.jchemed.1c00729
J. Chem. Educ. XXXX, XXX, XXX−XXX

Likewise, the "exothermic process" reasoning tile was used disproportionately with the conclusion that energy is released. Given the specificity of these statements (which explicitly use the words "heat is absorbed" and "heat is released", respectively) and their similarity to the conclusions, the observed correlation is unsurprising and provides evidence that students were engaging with the task as intended.

Limitations

Reasoning chain construction tasks present some clear advantages for studying student reasoning patterns. However, the methodology does have some shortcomings. While the majority of students engaged appropriately with the task, by entering some reasoning support and a single conclusion, the flexibility in how chains could be constructed resulted in a number of responses that were not included for analysis. In a multiple-choice with free-response version of the task, students were forced to select a single conclusion, and thus, these ambiguities were avoided. A small amount of additional instruction on how to engage with the task could be provided to reduce the number of problematic cases in the chaining format.

There is a significant cognitive load associated with processing and completing the chaining format tasks. Thus, even students who have a solid understanding of chemical bonding might not produce high-quality chains simply because the format of the task is too demanding, leading us to underestimate student understanding of the topic at hand. In contrast, data from multiple choice instruments may lead to overestimation of student understanding by providing a significant crutch for students in the form of the available answer choices.[47] The fact that students produce chains of substantially higher quality on the chaining versions of the task, however, suggests that this cognitive load is not a significant impairment to student performance, and the similarity of overall answer distribution between the survey versions similarly suggests that the differences in cognitive load are minimal.

Another limitation is that the data presented in this work only addresses the final chain that students assembled in their reasoning space. They do not provide insights into the process by which students arrived at their conclusion or how they selected reasoning tiles to include in their chain (notably, this information is also lacking in the standard multiple choice and free response question types). A limited number of interviews conducted with reasoning statements on index cards were used to validate the reasoning tiles and provide some insights into the dynamics of student behavior when presented with reasoning statements. However, additional interviews in which students think aloud while interacting with the actual task (preferably while also collecting eye tracking data for the students) would be helpful in understanding the dynamics of student reasoning in the moment.

Finally, we believe that the data generated by the reasoning chain construction tasks can be of interest to researchers using a wide variety of theoretical frameworks. For example, while we have favored the use of dual-process theories of reasoning, the resources framework, which proposes that knowledge is assembled from the activation of particular cognitive resources "in the moment", could similarly be applied to interpret results from tasks such as these.[48] Thus, we cannot use this work to claim that dual-process theories are the only "correct" way in which to interpret student reasoning patterns or that it is the only theoretical lens with which these data are consistent. We find, however, that dual-process theories provide a particularly helpful framework for thinking about our data because, as illustrated in our Discussion section, these theories provide mechanistic predictions and explanations about when and why certain cognitive resources are activated, and how mental models are evaluated and endorsed or rejected.

Implications for Teaching and Research

The reasoning chain construction task presents a method for collecting data on student reasoning across a wide variety of questions and topics. Data collected via this tool can provide much deeper insights into what information students find relevant and useful for a given question than is easily obtained from written free-response data. While not all reasoning chains are built with a logical flow of information, they do all provide evidence as to what information students deem relevant for answering the question and give an indication of the ideas, both productive and unproductive, that students may hold about a single topic.[49] The chaining construction task allows for this simultaneous expression of ideas in a way that multiple choice instruments alone cannot and with more detail and less ambiguity than is necessarily included in a free-response answer. Additionally, unlike free-response data that must be hand-coded by researchers, the analysis of student responses to a reasoning chain construction task is easily automated, allowing reasoning patterns to be quickly established for large numbers of respondents.

The design of a reasoning chain construction task can easily be altered to test the relevance of information made available to students. Students can randomly be served different versions of a chaining task to test whether the presence or absence of specific information alters student response patterns. The ability to identify student response patterns and test the uptake or effects of individual reasoning statements can help elucidate unproductive reasoning pathways and may help with the development of instructional interventions.

In the current work, the reasoning chain construction task only has been used only as a data collection tool. Ultimately, we hope that these tasks may serve as a teaching tool as well. However, to date we have collected no data as to the utility of these tasks as an instructional aid. In the current work, a correct answer and an example of an appropriate reasoning chain were provided to students after the assignment due date. As students complete the task, no feedback was provided. The inclusion of automated feedback based on the conclusion selected by a student, the presence or absence of specific tiles in a student's final chain, or a combination of the two, may provide a powerful tool for helping students develop the skills to better consider and articulate the reasoning they use in response to a chemistry question.

■ CONCLUSION

In this work, we illustrate the use of a novel question type, the reasoning chain construction task, in the context of a question on chemical bonding known to be difficult for many students. When presented with reasoning tiles containing correct information in the chaining-format, student conclusion patterns remained consistent with those from a multiple choice with free-response format. These findings are consistent with a dual-process theories interpretation of student reasoning, which suggests that students will be unlikely to abandon their first-available mental model in the absence of

https://doi.org/10.1021/acs.jchemed.1c00729
J. Chem. Educ. XXXX, XXX, XXX−XXX

information that explicitly refutes that model. The reasoning tiles made available to students drew upon an established framework, Nahum's bottom-up framework, for understanding chemical bonding. Students selected many tiles equally in support of both correct and incorrect conclusions.

One significant advantage of the reasoning chain construction task revealed in this study is the depth of reasoning support provided by students. The structure of the task appeared to help students formulate support for their conclusions, both correct and incorrect, in a way not typically observed in a multiple choice with free response question format. Students were able to engage with the information provided in a way that allowed them to build a link between their initial belief-based response and reasons for this belief. The additional information provided in student reasoning chains, for both correct and incorrect responses, could be of use in designing instructional interventions that more accurately align with information that students find relevant in response to a particular question.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.1c00729.

Warm-up task as provided to every student (PDF, DOCX)

■ AUTHOR INFORMATION

Corresponding Author

Megan L. Nagel − *Department of Chemistry, Penn State Greater Allegheny, McKeesport, Pennsylvania 15132, United States;* orcid.org/0000-0002-8707-6635; Email: mlm458@psu.edu

Author

Beth A. Lindsey − *Department of Physics, Penn State Greater Allegheny, McKeesport, Pennsylvania 15132, United States;* orcid.org/0000-0002-5017-7440

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jchemed.1c00729

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are extremely grateful to Mary Jo Bojan for allowing access to her course for the duration of this study. This material is based upon work supported by the National Science Foundation under Grants Nos. DUE-1821390, DUE-1821123, DUE-1821400, DUE-1821511, and DUE-1821561.

■ REFERENCES

(1) Council, N. R. *How People Learn Brain, Mind, Experience, and School*; The National Academies Press: Washington, DC, 2000.

(2) Bodner, G. M. I Have Found You an Argument: The Conceptual Knowledge of Beginning Chemistry Graduate Students. *J. Chem. Educ.* **1991**, *68* (5), 385.

(3) Cooper, M. M.; Corley, L. M.; Underwood, S. M. An Investigation of College Chemistry Students' Understanding of Structure-Property Relationships. *J. Res. Sci. Teach.* **2013**, *50* (6), 699−721.

(4) Cooper, M. M. Why Ask Why? *J. Chem. Educ.* **2015**, *92* (8), 1273−1279.

(5) *A Framework for K−12 Science Education: Practices. Crosscutting Concepts, and Core Ideas*; National Research Council: Washington, DC, 2012.

(6) Russ, R. S.; Coffey, J. E.; Hammer, D.; Hutchison, P. Making Classroom Assessment More Accountable to Scientific Reasoning: A Case for Attending to Mechanistic Thinking. *Sci. Educ.* **2009**, *93* (5), 875−891.

(7) Sevian, H.; Talanquer, V. Rethinking Chemistry: A Learning Progression on Chemical Thinking. *Chem. Educ. Res. Pract.* **2014**, *15* (1), 10−23.

(8) Berland, L. K.; Reiser, B. J. Making Sense of Argumentation and Explanation. *Sci. Educ.* **2009**, *93* (1), 26−55.

(9) Chi, M. T. H.; Wylie, R. The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educ. Psychol.* **2014**, *49* (4), 219−243.

(10) McNeill, K. L.; Lizotte, D. J.; Krajcik, J.; Marx, R. W. Supporting Students' Construction of Scientific Explanations by Fading Scaffolds in Instructional Materials. *J. Learn. Sci.* **2006**, *15* (2), 153−191.

(11) Pellegrino, J. W. *Knowing What Students Know*; National Academies Press: Washington, D.C., 2001; pp 48−52.

(12) Sandoval, W. A.; Millwood, K. A. The Quality of Students' Use of Evidence in Written Scientific Explanations. *Cogn. Instr.* **2005**, *23* (1), 23−55.

(13) Moon, A.; Stanford, C.; Cole, R.; Towns, M. The Nature of Students' Chemical Reasoning Employed in Scientific Argumentation in Physical Chemistry. *Chem. Educ. Res. Pract.* **2016**, *17* (2), 353−364.

(14) Cooper, M. M.; Underwood, S. M.; Hilley, C. Z.; Klymkowsky, M. W. Development and Assessment of a Molecular Structure and Properties Learning Progression. *J. Chem. Educ.* **2012**, *89* (11), 1351−1357.

(15) Haudek, K. C.; Prevost, L. B.; Moscarella, R. A.; Merrill, J.; Urban-Lurain, M. What Are They Thinking? Automated Analysis of Student Writing about Acid-Base Chemistry in Introductory Biology. *CBE Life Sci. Educ.* **2012**, *11* (3), 283−293.

(16) Birenbaum, M.; Tatsuoka, K. Open-Ended Versus Multiple-Choice Response Formats - It Does Make A Difference for Diagnotic Purposes. *Appl. Psychol. Meas.* **1987**, *11* (4), 385−395.

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

(18) Chandrasegaran, A. L.; Treagust, D. F.; Mocerino, M. The Development of a Two-Tier Multiple-Choice Diagnostic Instrument for Evaluating Secondary School Students' Ability to Describe and Explain Chemical Reactions Using Multiple Levels of Representation. *Chem. Educ. Res. Pract.* **2007**, *8* (3), 293−307.

(19) Tuysuz, C. Development of Two-Tier Diagnostic Instrument and Assess Students' Understanding in Chemistry. *Sci. Res. Essays* **2009**, *4* (6), 626−631.

(20) Cloonan, C. A.; Hutchinson, J. S. A Chemistry Concept Reasoning Test. *Chem. Educ. Res. Pract.* **2011**, *12* (2), 205−209.

(21) Speirs, J. C.; Ferm, W. N.; Stetzer, M. R.; Lindsey, B. A. Probing Student Ability to Construct Reasoning Chains: A New Methodology. *2016 Physics Education Research Conference*, Sacramento, CA, July 20−21, 2016; pp 328−331.

(22) Venkataraman, B. Emphasizing the Significance of Electrostatic Interactions in Chemical Bonding. *J. Chem. Educ.* **2017**, *94* (3), 296−303.

(23) Barker, V.; Millar, R. Students' Reasoning about Basic Chemical Thermodynamics and Chemical Bonding: What Changes Occur during a Context-Based Post-16 Chemistry Course? *Int. J. Sci. Educ.* **2000**, *22* (11), 1171−1200.

(24) Lindsey, B. A.; Nagel, M. L.; Savani, B. N. Leveraging Understanding of Energy from Physics to Overcome Unproductive Intuitions in Chemistry. *Phys. Rev. Phys. Educ. Res.* **2019**, *15* (1), 10120.

(25) Becker, N.; Cooper, M. M. College Chemistry Students' Understanding of Potential Energy in the Context of Atomic-Molecular Interactions. *J. Res. Sci. Teach.* **2014**, *51* (6), 789−808.
(26) Galley, W. C. Exothermic Bond Breaking: A Persistent Misconception. *J. Chem. Educ.* **2004**, *81* (4), 523−525.
(27) Boo, H. K. Students' Understandings of Chemical Bonds and the Energetics of Chemical Reactions. *J. Res. Sci. Teach.* **1998**, *35* (5), 569−581.
(28) Kryjevskaia, M.; Stetzer, M. R.; Grosz, N. Answer First: Applying the Heuristic-Analytic Theory of Reasoning to Examine Student Intuitive Thinking in the Context of Physics. *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **2014**, *10* (2), 020109.
(29) Evans, J. S. B. T. The Heuristic-Analytic Theory of Reasoning: Extension and Evaluation. *Psychon. Bull. Rev.* **2006**, *13* (3), 378−395.
(30) Kahneman, D. *Thinking, Fast and Slow*; Farrar, Strauss, & Giroux: New York, 2011; Vol. *153*.
(31) Kryjevskaia, M.; Stetzer, M. R.; Grosz, N. Answer First: Applying the Heuristic-Analytic Theory of Reasoning to Examine Student Intuitive Thinking in the Context of Physics. *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **2014**, *10* (2), 1−12.
(32) Kryjevskaia, M.; Grosz, N. *Examining Students Reasoning in Physics Through the Lens of the Dual Process Theories of Reasoning: The Context of Forces and Newton's Laws*; Springer International Publishing, 2020.
(33) Gette, C. R.; Kryjevskaia, M.; Stetzer, M. R.; Heron, P. R. L. Probing Student Reasoning Approaches through the Lens of Dual-Process Theories: A Case Study in Buoyancy. *Phys. Rev. Phys. Educ. Res.* **2018**, *14* (1), 10113.
(34) Kryjevskaia, M.; Stetzer, M. R.; Lindsey, B. A.; McInerny, A.; Heron, P. R. L.; Boudreaux, A. Designing Research-Based Instructional Materials That Leverage Dual-Process Theories of Reasoning: Insights from Testing One Specific, Theory-Driven Intervention. *Phys. Rev. Phys. Educ. Res.* **2020**, *16* (2), 20140.
(35) Heckler, A. F. *The Ubiquitous Patterns of Incorrect Answers to Science Questions: The Role of Automatic, Bottom-up Processes*; Elsevier Inc., 2011; Vol. *55*.
(36) Kryjevskaia, M.; Heron, P. R. L.; Heckler, A. F. Intuitive or Rational? Students and Experts Need to Be Both. *Physics Today*; American Institute of Physics, 2021; pp 28−34.
(37) Qualtrics. Provo, UT 2021.
(38) Teichert, M. A.; Stacy, A. M. Promoting Understanding of Chemical Bonding and Spontaneity through Student Explanation and Integration of Ideas. *J. Res. Sci. Teach.* **2002**, *39* (6), 464−496.
(39) Levy Nahum, T.; Mamlok-Naaman, R.; Hofstein, A.; Taber, K. S. Teaching and Learning the Concept of Chemical Bonding. *Stud. Sci. Educ.* **2010**, *46* (2), 179−207.
(40) Nagel, M. L.; Lindsey, B. A. Student Use of Energy Concepts from Physics in Chemistry Courses. *Chem. Educ. Res. Pract.* **2015**, *16* (1), 67−81.
(41) Evans, J. S. B. T. Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annu. Rev. Psychol.* **2008**, *59*, 255.
(42) Speirs, J. C. Exploring and Supporting Student Reasoning in Physics by Leveraging Dual-Process Theories of Reasoning and Decision-Making. *New Methodologies for Examining and Supporting Student Reasoning in Physics*. Doctorate Thesis, University of Maine, Orono, Maine, 2019.
(43) Lindsey, B. a. Student Reasoning about Electrostatic and Gravitational Potential Energy: An Exploratory Study with Interdisciplinary Consequences. *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **2014**, *10* (1), 013101.
(44) Evans, J. S. B. T.; Stanovich, K. E. Dual-Process Theories of Higher Cognition. *Perspect. Psychol. Sci.* **2013**, *8* (3), 223−241.
(45) Talanquer, V. Chemistry Education: Ten Heuristics to Tame. *J. Chem. Educ.* **2014**, *91* (8), 1091−1097.
(46) Dreyfus, B. W.; Geller, B.; Gouvea, J.; Sawtelle, V.; Turpen, C.; Redish, E. F. Negative Energy: Why Interdisciplinary Physics Requires Multiple Ontologies. *Proceedings of the 2013 Physics Education Research Conference*, Portland, OR, July 17−18, 2013.
(47) Pride, T. O.; Vokos, S.; McDermott, L. C. The Challenge of Matching Learning Assessments to Teaching Goals: An Example from the Work-Energy and Impulse-Momentum Theorems. *Am. J. Phys.* **1998**, *66* (2), 147−157.
(48) Hammer, D.; Elby, A. Tapping Epistemological Resources for Learning Physics. *J. Learn. Sci.* **2003**, *12* (1), 53−90.
(49) Nehm, R. H.; Schonfeld, I. S. Measuring Knowledge of Natural Selection: A Comparison of the CINS, an Open-Response Instrument, and an Oral Interview. *J. Res. Sci. Teach.* **2008**, *45* (10), 1131−1160.

https://doi.org/10.1021/acs.jchemed.1c00729
J. Chem. Educ. XXXX, XXX, XXX−XXX