Journal of Science Education and Technology
https://doi.org/10.1007/s10956-023-10040-5

Integrative Analysis Using Big Ideas: Energy Transfer and Cellular Respiration

Jonathan T. Shemwell[1] · Daniel K. Capps[2] · Ayca K. Fackler[2] · Carlson H. Coogler[1]

Accepted: 8 March 2023

Abstract
Big ideas in science education are meant to be interpretive frameworks that empower student learning. Unfortunately, outside of the broad conception of scientific evaluation, there are few theoretical explanations of how this might happen. Therefore, we contribute one such explanation, an instructional concept called integrative analysis wherein students use a big idea to interconnect isolated scenarios and enrich their meanings. We illustrate the characteristics and value of integrative analysis within an empirical study of student learning in 9th-grade biology. The study focused on using energy transfer as a big idea for teaching cellular respiration. Fifty-nine students were randomly assigned to one of two instructional conditions. In the "analysis" condition, students processed a set of three manipulatives representing cellular respiration molecules; then, they abstracted the deep energy transfer structure of these manipulatives as a big idea. In the "recognition" condition, students processed the same molecule-manipulatives, but without energy interpretations. Instead, they constructed additional manipulatives using novel materials. Then, students in both conditions received an identical lesson where they used their knowledge of the manipulatives to learn about one cellular respiration process, glycolysis. Specifically, students processed a sequence of three texts describing glycolysis, annotating the texts with either their deep energy transfer structure (analysis condition) or their contextualized knowledge of the manipulatives (recognition condition). A posttest showed that in the analysis condition, this process was significantly integrative as evidenced by analysis students' advantage over recognition students in connecting glycolysis to novel phenomena and generating causal explanations about glycolysis.

Keywords Analysis · Big ideas · Deep structure · Knowledge integration · Energy · Cellular respiration

Restructuring the school science curriculum around big ideas, or fundamental concepts that run through domains, has been a central feature of science education reform for the past two decades and more (Australian Curriculum Assessment and Reporting Authority (ACARA), 2012; College Board, 2011, 2019; Harlen, 2010; NGSS Lead States, 2013; National Research Council (NRC), 2002, 2012; Smith & Girod, 2003; Smith et al., 2006). The main purpose for focusing on big ideas is to constrain an otherwise overbroad curriculum (Schmidt et al., 1997, 2007, see also American Association for the Advancement of Science (AAAS), 1990; NRC, 2002). A second, slightly less prominent purpose is to provide focus points for developing student knowledge over years of schooling (Castro-Faix et al., 2021; NRC, 2012; Plummer et al., 2020; Stevens et al., 2010; Todd & Kenyon, 2016). Yet a third, purpose, easily overlooked and the topic of this article, is to provide crucial theoretical structures for evaluation practices of argumentation and critique (NRC, 2012). Or, framed more broadly, it is to provide necessary ideational structures for critical (i.e., knowledge using) learning processes of synthesis and analysis.

As argued by Mitchell et al. (2017), the role of big ideas in critical learning processes gets overlooked because the application of big ideas to instruction is too easily seen as unproblematic. For example, it is easy to interpret the disciplinary core ideas in US science reform (NGSS Lead States, 2013) as mere statements of ideas to be learned rather than theoretical structures meant to enable analysis and

✉ Daniel K. Capps
dacapps@uga.edu

Jonathan T. Shemwell
jtshemwell@ua.edu

Ayca K. Fackler
aycakfackler@uga.edu

Carlson H. Coogler
chcoogler@crimson.ua.edu

1 The University of Alabama, Tuscaloosa, USA

2 University of Georgia, Athens, USA

synthesis, as intended by the reform (NRC, 2012). Therefore, the aim of the present article is to increase awareness and understanding of how big ideas propel these critical learning processes, particularly analysis. We pursue this purpose by presenting a small-scale classroom experiment involving 59 high school biology students using the big idea of energy transfer to learn about cellular respiration. The experiment engaged students in one particular form of analysis with this big idea, which we call integrative analysis. Briefly, this was the sustained use of energy transfer as a framework for interpreting seemingly isolated scenarios in cellular respiration, thereby interconnecting the scenarios and enriching their meanings. Importantly, and as a caveat, the present study's instantiation of integrative analysis did not take place within an epistemic activity, or even a rich learning activity. Instead, it was embedded in a frankly decontextualized, didactic process in which students used an energy transfer model to interpret expository texts describing cellular respiration. Thus, the evidence developed, while perhaps showing what integrative analysis can do despite a severely lean instructional activity, does not explicitly show what it might do within a richer one. We take up this latter issue conjecturally at the close of the article.

Big Ideas and Integrative Analysis

What Are Big Ideas?

The present article's use of the term "big idea" is traceable to *How People Learn*, a synthesis of constructivist-era research pointing out that experts' knowledge is "organized around core concepts or 'big ideas' that guide their thinking about their domains" (NRC, 2000, p. 36). NRC's (2000) concept of knowledge for thinking is reflected in most authors' accounts of big ideas, for instance, Mitchell et al.'s (2017) description of "a unifying principle that connects and organizes a number of smaller ideas or concepts and multiple experiences" (p. 598) and Wiggins and McTighe's (2005) description of "meaningful patterns that enable one to connect the dots of otherwise fragmented knowledge" (p. 339). In science education especially, big ideas must be sufficiently articulated to illuminate particular domains of study (Mitchell et al., 2017; Windschitl et al., 2020). For this reason, big ideas align more closely to what current US science education standards call disciplinary core ideas than what those standards call crosscutting concepts (NGSS Lead States, 2013; NRC, 2012).

Big Ideas vs. Encompassing Ideas Writing for curriculum and instruction audiences, authors have striven to underscore the importance of distinguishing big ideas, which must have explanatory power for their domains, from encompassing ideas that do not have this power (Mitchell et al., 2017; Wiggins & McTighe, 2005; Windschitl et al., 2020). They point out that encompassing ideas may seem "big" because they allow diverse phenomena to be grouped together (Mitchell et al., 2017), but they add little to understanding in the absence of information specifying what relationships must hold within and across the groups. Thus, Mitchell et al. proposed that big ideas should include "links between content ideas" (p. 608), and Windschitl et al. (2020) stipulated that big ideas should "express relationships, not just facts" (p. 22). Windschitl also required that these relationships be mechanistic (see especially Windschitl et al., 2012), meaning they must inform why events happen (Russ et al., 2008). Windschitl et al. (2020) gave the example of "convection drives plate movement" (p. 22) as a relationship within a big idea that was insufficiently mechanistic, evidently because there was no information about how convection drove plate movement. Framing the need for relationships in big ideas more technically (though only vaguely alluding to their mechanistic aspect), NRC (2000) described big ideas as having "meaningful relations among related elements clustered into related units that are governed by underlying concepts" (p. 38).

Big Ideas as Deep Structures In the present article, we employ the term *deep structures* as an off-the-shelf expression for big ideas that usefully augments the existing plain-language terminology just presented. Specifically, *structure* denotes a set of conceptual constraints (or relationships), and *deep* indicates their invariance across levels of abstraction (i.e., elements, units, and concepts). Thus, a deep structure is an arrangement of conceptual constraints uniting otherwise dissimilar phenomena (Gentner & Markman, 2006; see also Chi & VanLehn, 2012). As an example, Gentner and Kurtz (2005) gave the example of a foot bridge and a dental bridge as having the same conceptual constraints. While the term deep structure is rarely used by instructional theorists to describe big ideas, it does have currency in education broadly (Schwartz et al., 2011), and in science education specifically (Shemwell et al., 2015; Chase et al., 2019; Kuo & Weiman, 2016). Our use of it in the present study is meant to enable a concise theoretical presentation of our focal instructional concept, integrative analysis.

Integrative Analysis as a Role for Big Ideas

Any presentation of one particular role for big ideas must necessarily be framed within an ecology of possible roles for them. This ecology is not easily described, as indeed surprisingly few authors have attempted to theorize about the spectrum of instructional uses of big ideas in the post reform era (for a notable exception, see Mitchell et al., 2017). However, a useful foundation for an ecology was provided by Osborne (2011, 2014) which also appeared in NRC (2012), namely the concept of epistemic "evaluation" (NRC, 2012, p. 46) as the practice of coordinating theory and evidence within scientific practices in the classroom. Osborne specified two distinct directions of coordination, from evidence to theory, and vice versa. Here, we describe these as the synthetic direction (evidence to theory) and the analytic direction (theory to evidence). Switching to noun forms (synthesis, analysis) and using the example from NRC (2012) of developing and using models, developing models would be synthesis; using them would be analysis. And, as is specifically pointed out by NRC (2012), both synthesis and analysis depend crucially on ideas that are generalized relationships in a domain (i.e., big ideas that are deep structures, and not encompassing ideas, see page 31).

Analysis and Synthesis as Critical Learning Processes By abstracting the critical learning processes of analysis and synthesis away from scientific evaluation, we hope to highlight a fact that is perhaps too easily lost among the complexities of the latter. Namely, to yield their true value, critical learning processes must be centered on deep structures, meaning big ideas, not encompassing ideas. Additionally, while the most potent forms of synthesis and analysis are probably epistemic, there is undoubtedly some utility in more traditional, didactic versions of them. From this broader perspective, our use of the term analysis means using a big idea as a deep structural framework for interpreting a phenomenon, especially when "transforming it in some way, in order for the resources of a given theory or conceptual framework to be brought to bear" (Beaney, 2014, Part 6, Paragraph 2). By contrast, synthesis is when students develop deep structure from a number of related phenomena (Chalmers et al., 2017; see also Lesh & Doerr, 2003; Windschitl et al., 2020; Windschitl et al., 2012).

Integrative vs. Connective Analysis Integrative analysis, our topic here, is the act of using a big idea to interpret scenarios that are initially apprehended in isolation from each other, thereby enriching their meaning and interconnecting them. As shown in Fig. 1, the big idea, as a deep structure, contains multiple elements with defined relationships that are themselves interrelated within a unit, as described by NRC (2000). Within the transactions comprising integration, each scenario is mapped to the deep structure at the element level, as happens in the preparatory, element-to-element structure-mapping within the learning activities in studies of analogical transfer (Gentner, 1983; Gentner et al., 2003; Gick & Holyoak, 1980, 1983). Two transformations occur within this element-level mapping that are meant to be salient in the figure. One is an adjustment of idea contents within each scenario—and potentially a given element of the deep structure—to bring the two into alignment. In the present study, we call this adjustment enriching to emphasize

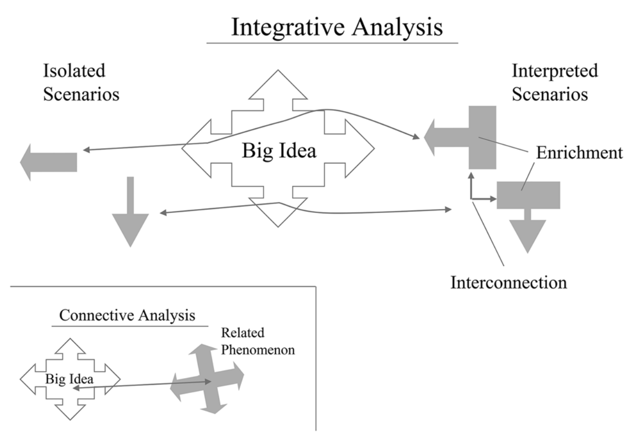


Fig. 1 In integrative analysis, learners initially perceive scenarios as isolated phenomena. In connective analysis, they apprehend them as related phenomena forming a structure

the meaning that is added to the scenario from the deep structural element. The other transformation is to interconnect the formerly isolated scenarios by imposing element-to-element relationships within the big idea as a unit. These two processes together, enriching and interconnecting, comprise the integration aspect of integrative analysis.

Integrative analysis can be contrasted with connective analysis, shown at the bottom left of Fig. 1, with the distinction being how the learner apprehends a set of novel scenarios. With integrative analysis, the learner initially perceives isolated phenomena, while in connective analysis, they perceive related phenomena forming a structure. Connective analysis is therefore a whole-structure to whole-structure mapping, as contrasted with the piecemeal mapping of integrative analysis. As such, connective analysis is what happens when participants of analogical transfer studies solve novel problems by detecting structural similarities between problem scenarios and abstract ideas they have learned (Gick & Holyoak, 1983; Gentner, 2010; Gentner et al., 2003).

As a point of caution, the model of integrative analysis presented here should not be taken as a sequence of operations that proceeds neatly and steadily toward a durable knowledge product. We do not think of enriching and interconnecting as happening that way. Rather we think of them being as what Kellman et al. (2010) called "transactions with structure" (p. 290) that are messy and transient, largely because both parties in the transactions—the given scenario and the big idea—are dynamic, not fixed, entities within student thinking (Gupta et al., 2010; Hammer & Manz, 2019; Sikorski & Hammer, 2017).

Instructional Ideas Reflecting Integrative Analysis

Most of the aspects of integrative analysis just described are present in the instructional literature pertaining to big ideas, but as fragments rather than a coherent whole. Moreover, authors tend to be more suggestive than definitive in describing either integration or analysis with big ideas. For example, NRC (2012) suggested analysis by stating that big ideas should be "organizational structure[s] for the acquisition of new knowledge" (p. 25). However, the authors did not specify how the big idea structure might organize knowledge (i.e., interconnecting, enriching). Mitchell et al. (2017) touched upon interconnecting somewhat more explicitly when writing that a big idea "*connects* and organizes a number of smaller ideas or concepts and multiple experiences" (p. 598, italics added), though the authors did not elaborate on how interconnecting could occur; nor did they address analytical vs. synthetic modes of achieving it. In a similar way, Wiggins and McTighe (2005) wrote that big ideas could "connect the dots of otherwise fragmented knowledge" (p. 339), and that "a big idea is therefore both central to coherent connections in a field of study *and* a conceptual anchor for making facts more understandable and useful" (p. 66, i.e., italics in the original). Thus, they addressed both interconnecting and enriching, though, again, they did not elaborate how these could be achieved. In their case however, analysis with big ideas was at least mentioned as a critical learning process, namely in a statement about big ideas "providing a focusing conceptual lens for any study" (p. 69).

Shades of integrative analysis can be found in the science education literature apart from big ideas, especially under the heading of knowledge integration. In one example, Songer (1989) had students classify materials in terms of insulators or conductors. Potentially, the opposing relationships in the classification scheme were inherited by the materials being classified, which would be a rudimentary example of interconnecting as depicted in Fig. 1. A more sophisticated example occurred in a study by Kali et al. (2003) who had 7th graders use the rock cycle to support observations about rock structures they observed in the field. Students observed a granite outcrop, and later, a sandstone deposit located downhill with granite pieces embedded in it. Using the rock cycle, the students conjectured that erosion, transport, and deposition (i.e., processes in the cycle) had occurred from the granite outcrop to create the sandstone deposit. Thus, students used the rock cycle as a deep structure to both interconnect the two phenomena (i.e., causally and temporally), while also enriching their observations of them with unobservable formation processes.

The Present Study

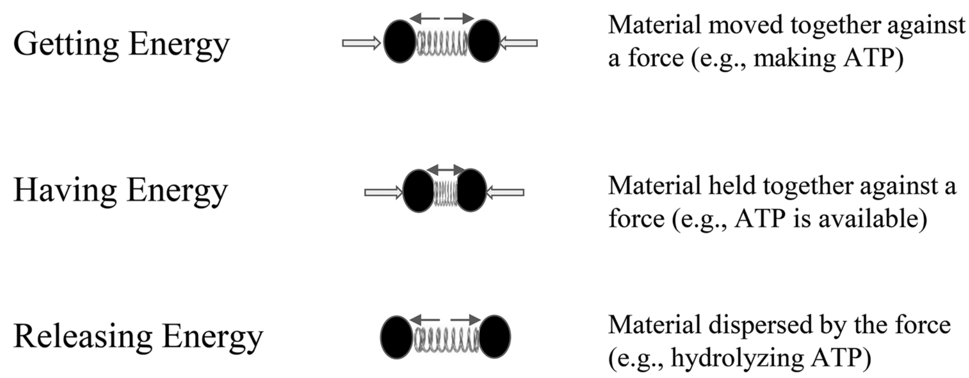
To investigate the utility of integrative analysis, we carried out a classroom-based study of its contribution to learning cellular respiration in 9th grade biology. Archetypically, learning about cellular respiration involves memorizing a set of descriptive (i.e., causeless) narratives of submolecular processes. Predictably, these scenarios contain abundant, unfamiliar information that often overwhelms students (Patro, 2008; Ross et al., 2008; Scholer & Hatton, 2008; Songer & Mintzes, 1994; White, 2016). As Patro (2008) put it, "The abstract nature of the processes, the multitude of details, and the new technical vocabulary make the topic challenging to even the most diligent students" (p. 85).

In the present study, we did not deviate from the archetypical presentation of descriptive narratives of cellular respiration processes. However, we did provide students an alternative to memorization, which was using a big idea to analyze these narratives, and hopefully, achieving a degree of integration through this activity. The big idea was energy transfer, as is specified within current large-scale reforms for this content area (College Board, 2011, 2019; NGSS Lead States, 2013, NRC, 2012). Specifically, it was a model[1] of energy input, storage, and output based on effort against or by a springy resistance. Students constructed this model from their combined observations of hands-on manipulatives of three prominent molecules in cellular respiration: glucose, NADH, and ATP (see Fig. 2). Each of the molecules contained some kind of springy or explosion-like action as it was assembled or broken down to form smaller constituents. Students modeled these processes as being similar to the elastic actions of a spring which could be compressed for energy input (bringing material together), held in compression for energy storage (keeping material together), and let go for energy output (allowing material to disperse). In terms of the anatomy of big ideas (i.e., elements that are interrelated to form a structure, NRC, 2000), the elements of the structure were the three energy processes of getting, having, and releasing energy. The element-level meanings within this structure were intuitive ideas about actions against or by springy resistances (DiSessa, 1993; Kapon & diSessa, 2012), for instance, effort against the resistance for "getting" energy. The key inter-element relationships, leveraging intuitive thinking both about springy resistances and consistent with intuitive ideas within the substance metaphor for energy (Close & Scherr, 2015; Duit, 1987; Scherr et al., 2012; Swackhamer, 2005), were that getting energy (i.e., effort against a spring) naturally leads to having it (i.e., a spring compressed), having energy portends releasing it (i.e., ready to uncompress), and releasing energy (i.e., a violent action) can supply the needed effort for getting it.

Procedurally, the study was a small-scale experiment with two instructional conditions, which we called analysis and recognition (see Fig. 3). The instruction for both

[1] We followed Windschitl et al.'s (2012) preference for students to understand big ideas as models. This approach affords big ideas epistemological status with students (i.e., imperfect, fitting distinct purposes, subject to revision) while also providing tools with which to develop and communicate big ideas.

Fig. 2 The energy model abstracted the deep structure of three molecule-manipulatives



conditions included a preparatory lesson, which was different for each condition, followed by a content lesson, which was the same for both. In the analysis preparatory lesson, students used the three molecule-manipulatives to learn the energy transfer model just described. In the recognition preparatory lesson, students used the same manipulatives as the analysis group, but they did not abstract an energy model as a deep structure. Instead, they constructed new molecule-manipulatives to learn more about each molecule's unique characteristics.

In the content lesson, students used their knowledge from the preparatory lesson to interpret a set of narratives describing glycolysis, a key process within cellular respiration. There were three narratives in all, each in a different textual format: a comic strip, a diagram, and a page-length reading. All three spanned the entire glycolysis process, and all featured the molecules that had been introduced as manipulatives in the preparatory lesson. Students' task was to annotate each narrative by finding these molecules and making margin notes denoting their actions. Analysis students did this by writing down which of the three energy transfer elements in their model was occurring (i.e., getting, having, or releasing energy). Recognition students did this by writing short phrases saying what a given molecule was doing. Thus, both processes were supposed to be similarly active. However, while the former process was meant to be analysis, with the attendant enriching and interconnecting from the deep structure depicted in Fig. 1, the latter process was meant to be recognition (of familiar structures and functions), and thus without the integration conferred by a deep structure.

During the content lesson, we measured the completeness and accuracy of students' annotation, purportedly as an index of their success in analysis or recognition. Afterward, we measured the extent of integration using a posttest (see Fig. 3, center and right). Two types of items on the test focused on integration. The first type, called the linking items, checked whether students could connect glycolysis processes structurally to phenomena outside of cellular respiration. The capability of analysis students to do this, as indicated by an advantage over recognition students, would indicate that the energy structure they had putatively used for "analysis" was a deep structure as intended and not a set of inert energy labels. The second type, called the causal explanation items, asked students to speculate why events in glycolysis and other cellular respiration processes might have occurred. Importantly, the speculations we hoped they would generate had not been explicitly presented in the content lesson but could be inferred if students integrated information using the big idea—if they enriched narrative events with causal meanings from the elements within the energy structure, or if they interconnected the events with causal relationships between the elements. As a caveat, and as described in measures and coding, causal meanings in the model elements were also present in the manipulatives. Thus, it was crucial to observe whether the analysis students did better than the recognition students on this measure. Accordingly, the research questions were:

1. How successfully did students analyze (or recognize) events within the descriptive narratives of glycolysis as they annotated texts during the content lesson?
2. Were analysis students analyzing with a deep structure, as evidenced by an advantage over recognition students in linking glycolysis processes to phenomena outside of cellular respiration?

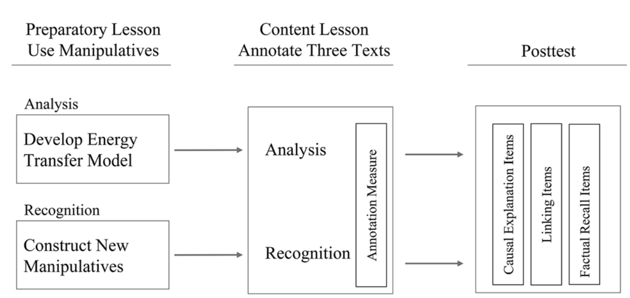


Fig. 3 Overview of the study showing the two instructional conditions along with the activities and measures

3. Was analysis integrative, as evidenced by an advantage for analysis students over recognition students in generating causal explanations?

To bring the answers to these questions into sharper relief, and as a partial check on whether the instructional contrast was robust, the posttest also included a third type of item, called factual recall, which measured students' knowledge of glycolysis processes described in the content lesson (see Fig. 3, right). We reasoned that recognition students would be able to learn and remember this descriptive information at least as well as analysis students because it involved familiar actions of the focal molecules from the preparatory lesson. Thus, we expected recognition students to do well on these items, but not so well on the linking and explanation items that required deeper, more structured interpretations.

Finally, and as should be clear by now, neither group of students in the study was in anything like an agentive or epistemic position within the content lesson. Students were not asking what they wanted to figure out, or how to figure it out (Schwarz et al., 2017); nor were they evaluating what should count as knowledge (Ford, 2008). On the other hand, the analysis students were (or were meant to be) in a strong position as learners. Specifically, they were equipped with a useful framework with which to analyze and integrate material which might otherwise come across as descriptive facts to memorize. It is this position, apart from other considerations, that the study that follows was meant to highlight.

Method

As indicated by the preceding overview, the overall design of the study had two mutually supporting components. The first and more salient component was the instructional contrast in which a total of 59 students, through random assignment, received either the analysis preparatory lesson or recognition preparatory lesson. This contrast was meant to isolate the anticipated effect of analysis (i.e., integration via enriching or interconnecting via a deep structure) from incidental effects of the instruction, most obviously the opportunity to learn from the molecule-manipulatives. The second design component, less conspicuous but just as important as the instructional contrast, was the use of measures meant to be diagnostic of integration via analysis. These indicated the extent to which "analysis" took place during learning, whether this process was genuine (i.e., done with a deep structure), and whether integration occurred as a result.

Context and Participants

The study took place in a public high school serving a rural community in the southeastern United States. The community was located about 40 minutes from a medium-sized town. The student population of the school was 95% white, typical of rural areas in the region. The 59 participants comprised all of the biology students taught by one of the school's two biology teachers and were distributed across two class periods. Both periods were honors-level biology, and most of the students were in the 9th grade. Two of the students were absent from school on the day of the posttest, both from the analysis condition. As a result, scores for 57 students overall, 28 analysis and 29 recognition, are reported on in the results.

The students' teacher was one of the two instructors for the study. Identifying as an African American female, she had 3 years of experience teaching biology, Earth science, and physical science. The other instructor was a university professor and an author of the present study who identified as a white male. He was a former National Board Certified science teacher experienced in a range of pedagogies and student populations. Two researchers aided the instructors by helping to manage time, collect student work, and facilitate small group learning.

Procedures

Assignment to Condition

Individual students in each class period were randomly assigned to either the analysis or recognition condition. This procedure created separate class halves for each class period, each with its own instructor. To minimize the possibility of non-equivalent groups due to fluke randomization (a concern arising from having only 59 students in the study), equivalence was checked within each class period using Pre-ACT scores. The criterion was that neither mean nor median scores could be more than one point apart. A coin flip assigned the halves to a condition, and a second coin flip assigned instructors to the class halves under the constraint that each instructor taught one analysis class half and one recognition class half. Thus, instructors were counterbalanced across conditions.

The Molecule-Manipulatives

The molecule-manipulatives that both conditions used in their preparatory lesson, namely ATP, NADH, and glucose, are depicted in Fig. 4. Each manipulative was analogous to

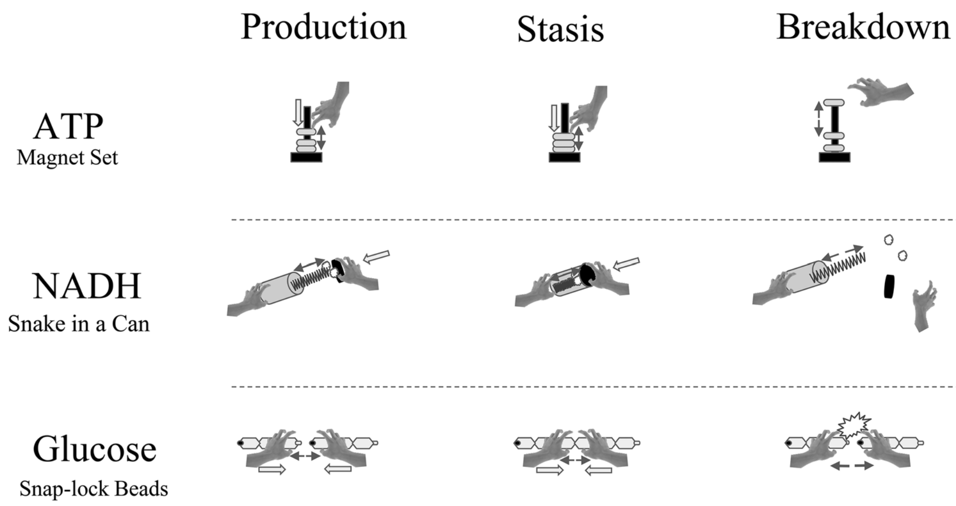


Fig. 4 The molecule-manipulatives used in both instructional conditions

its respective molecule on two levels, material and energy. On the material level, a given manipulative represented the distinctive structural configurations and transformations of its referent. On the energy level, it represented conservative forces involved in the structural transformations.[2] The ATP manipulative was an upright "adenosine" post onto which three repelling ring-magnet "phosphates" could be forced. To produce ATP, students would start with a two-phosphate configuration (ADP) and press down a third phosphate against the repelling force of the magnet. To reverse this process (i.e., to return to ADP), students would allow the magnet to push back up, recovering the effort they had put in. The NADH manipulative was a novelty snake-in-a-can. The can represented NAD + ; the spring represented the action of chemical bonds within NAD + ; the lid represented H- (hydrogen); and two cotton balls represented electrons. To produce NADH, the spring and cotton balls had to be stuffed into the "NAD + " can (i.e., exerting effort) before the "hydrogen" lid was placed over, to hold everything in. To reverse this process, the lid was removed, and the spring and electrons jumped out of the can (i.e., recovering effort). Finally, the glucose manipulative comprised six plastic beads, each representing a carbon atom, that were demonstrably pushed together to form glucose (exerting effort). When these were pulled into smaller pieces to signify breaking glucose down to form smaller constituents, a pull-string firework tied between the beads made a small pop (i.e., recovering effort).

The Preparatory Lessons

Both conditions' preparatory lesson lasted 50 minutes. During the lessons, both conditions spent 25 minutes processing the manipulatives and 25 minutes either modeling the energy structure of the manipulatives (analysis condition) or constructing an additional set of manipulatives (recognition condition).

Both Conditions: Processing the Manipulatives Students worked in small groups to process the molecule-manipulatives in sequence, starting with glucose and then going on to ATP and NADH. For each molecule, groups got a bin of disassembled materials along with an "energy story" instruction sheet that directed them to use the materials to act out the production, stasis, and breakdown of the molecule. The instruction sheets (see Figure S1) gave parallel procedures for acting out these processes with two key differences between conditions. The first was that the descriptions of processes in the analysis condition drew students' attention to resistive forces in the manipulatives, while corresponding descriptions in the recognition condition ignored these forces. Using ATP production as an example, the analysis instructions said, "it is hard to attach a third phosphate onto ADP," thus drawing attention to the repelling force of the magnets in the manipulative, while the recognition instructions said, "To make ATP, a third phosphate has to be bonded to ADP," thus ignoring the force. The second difference was the use of covering vocabulary (i.e., labels) for production, stasis, and breakdown processes. In the analysis instructions, energy terms were used as labels, specifically

[2] A major issue with this energy aspect of the manipulatives—which also extended to the model that students constructed from them—was that it incorrectly represented the effort to form molecules as pushing particles together against repulsive forces between them. From a chemistry standpoint, this is exactly wrong, as the forces involved are attractive, so it takes energy to pull particles away from each other. See the limitations in "Discussion" for additional explanation of this issue and how it limits the conclusions of the study.

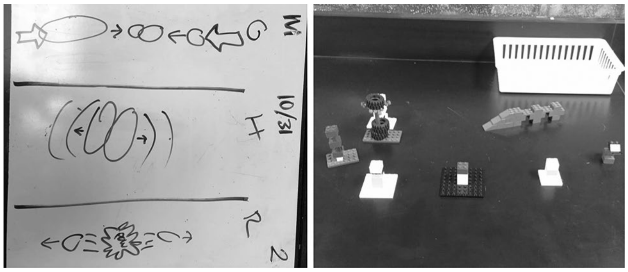


Fig. 5 An analysis group's general model (left) abstracted the energy structure of all of the molecules. On the right, a collection of recognition groups' additional manipulatives were faithful to the unique structure of a given molecule. Here, ATP was represented as having an adenosine base with three detachable phosphates

getting, having, and releasing energy. In the recognition condition, no labels were used. For example, at a point when analysis students used the ATP manipulative to show ATP breakdown as "releasing energy," recognition students used it to show that "the third phosphate can be removed from ATP. This is how ATP powers the cell."

Analysis Condition: Constructing a General Model For the final 25 minutes of the lesson, analysis students worked in small groups to construct a model of the deep energy structure of glucose, ATP, and NADH. Model construction followed a synthesis procedure described in Capps and Shemwell (2020) that combined abstracting processes for learning deep structures (Chase et al., 2019; Gick & Holyoak, 1980, 1983; Gentner et al., 2003; Schwartz et al., 2011) with representational practices of explanatory modeling (Cheng & Brown, 2015; Clement, 2008; Clement & Steinberg, 2002). The key synthetic step involved placing visual representations of the molecule-manipulatives around a whiteboard and following instructions to "draw a general model that stands for how each of the molecules can get, have, and release energy." The typical result was an abstract representation of material that could be forced together, held together, and let apart (see Fig. 5, left). After groups drew their initial models, the instructor collected the whiteboards at the front of the room and indicated contrasts, especially contrasts in portrayals of the restoring forces in the energy processes. Then, students improved their models, mostly focusing on showing the restoring forces as vividly as possible. Finally, students generated and practiced sound words and embodied actions to represent the three energy processes, again focusing on vivid representation of energy transfer via restoring forces. Examples of sound words were "ergh" for energy input, "buzz" for energy storage, and "boom!" for energy output. The embodied versions of these processes were straining muscles to force one's hands together against a springy resistance (energy input), holding the hands against the resistance (energy storage), and letting the hands come apart (energy output).

Recognition Condition: Constructing Additional Manipulatives Parallel to constructing models in the analysis condition, recognition students used the final 25 minutes of the preparatory lesson to work in small groups to construct a new set of manipulatives for each of the three energy carriers. To do this, they were given novel materials that afforded analogies to the material level of the analogs, but not the force level. These materials were Lego blocks for ATP and glucose, and craft materials for NADH. Students were told that their new set of manipulatives should "look and act" like their referents, meaning they should mimic their material configurations as well as the transformations of production, stasis, and breakdown (but without instruction or affordances for representing forces involved). The resulting manipulatives generally represented these configurations and transformations with fidelity equal to that of the original manipulatives, with the most salient variation between groups coming from the way the medium was used, for instance, the size and orientation of Lego blocks, or the craft item selected (see Fig. 5, right). After the students finished constructing the first molecule-manipulative, the instructor brought all groups' versions to the front of the room and went over how they looked and acted like the focal molecule despite surface-level differences from other groups' manipulatives. Then, students constructed the next molecule's manipulative, again followed by instructor commentary, and so on until all three had been completed.

The Content Lesson

The content lesson on glycolysis lasted 50 minutes and was the same for both conditions. To standardize the instruction, the mode of learning was annotating a sequence of three texts providing narrative descriptions of glycolysis. The first text was a comic strip, the second was a diagram, and the third was a page-length reading. Students spent about 10 minutes annotating each text. They worked in small groups for the first two texts and alone for the third, which comprised the annotation measure reported on in the results (see Fig. 6). For all three texts, annotating consisted of finding the energy carriers students had learned about in the preparatory lesson, and for each carrier, making a margin note to say what it was doing. After all students finished a given text, the instructor spent 5 minutes going over the annotations by calling on students to give their answers and making corrections.

Analysis students annotated the texts using their energy labels or sound words from the model (i.e., getting energy, or "err!", having energy, "buzz," or releasing energy, "boom").

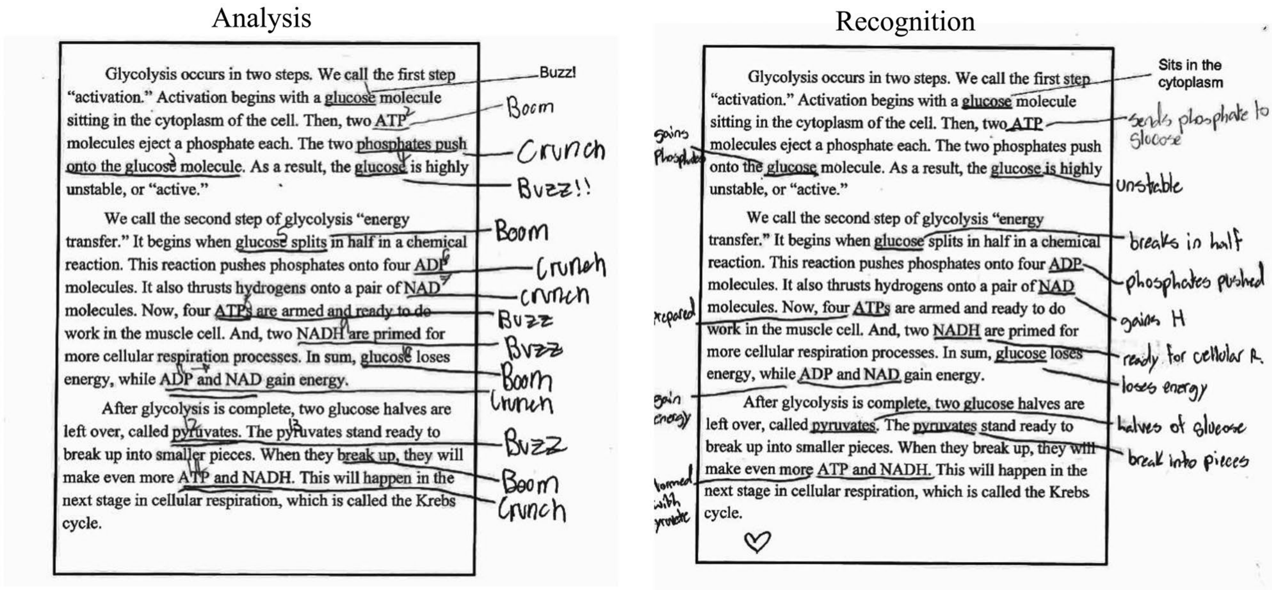


Fig. 6 The annotation measure occurred during the content lesson. Analysis annotations (left) applied the energy model. Recognition annotations (right) were more locally descriptive

Thus, their process was to interpret each event in the texts by assigning it to the appropriate element in their model structure. This process would be analysis to the extent that the energy terms were backed by meanings and relationships, and it would be integrative analysis to the extent that the model was a deep structure by which students enriched and/or interconnected events as they interpreted them (i.e., as in Fig. 1). The recognition students annotated the texts with short phrases of their own construction using their knowledge of the unique actions possible for each carrier. Thus, their process was in some ways more generative than the modeling process (i.e., they had to decide what to say), but it was potentially less interpretive. Nevertheless, the recognition process was meant to be meaningful, as events in the text should have been reminiscent of students' experiences with the actions of the manipulatives.

Measures and Coding

The measurement tools were the annotation measure during the content lesson and the posttest after the lesson. The annotation measure, embedded in the third of the three texts presented in the content lesson, indexed students' ability to find available information in the text and accurately annotate it. Sample papers from both conditions are shown in Fig. 6. A review of the analysis annotation shows that these student's annotations were constrained to be generalized and transformative such that a given energy process accommodated a range of scenarios. For example, the "having energy" annotation included glucose being highly unstable, ATP being armed, NADH being primed, and pyruvates being left over and standing ready (Fig. 6 left). Recognition annotations of these same events were descriptive and more unique to each particular context, which was the intention of the instructional contrast.

The posttest, which students took the day after the content lesson, had 15 items and took 20 minutes to complete. The test included 2 linking items, 6 causal explanation items, and 7 factual recall items. Table 1 provides an example of each item type, and the entire test is provided in the Supplementary Materials. Internal consistency, alpha, for the test was 0.69.

The linking items were meant to be diagnostic of deep structural knowledge by checking whether students could link glycolysis events structurally to scenarios in novel contexts. As the example item in Table 1 shows, students could do this either by interpreting both contexts in terms of energy transfer (characteristic of analysis students) or by interpreting both in terms of dynamic actions they had experienced in the molecule-manipulatives (characteristic of recognition students).[3] Either way, successful linking required transformation, a key aspect of mapping to a deep structure (Gentner, 2010). In the example shown in Table 1, the applied and restoring forces of the novel system, the rubber band, were in the opposite directions of those in students' energy model and the molecule-manipulatives, so linking glycolysis events to this system required students

[3] Correct student responses did not have to be characteristic of their condition.

Table 1 Three types of posttest items

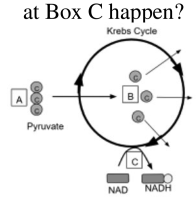
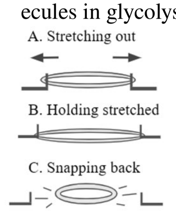
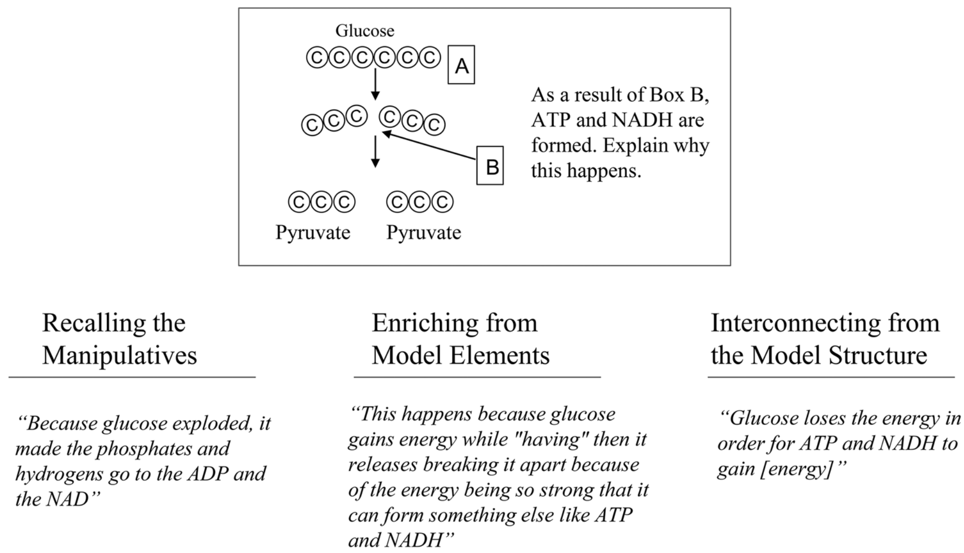
	Demand	Example item	Criterion for correct	Correct response		Incorrect response	
				Analysis	Recognition	Analysis	Recognition
Factual recall (7 items)	Describe events taught in the glycolysis lesson	The first phase of glycolysis is called activation. What happens in this phase?	Facts are consistent with those of the lesson	*The ATPs send their third phosphate to bond to the glucose*	*In activation the phosphates attach themselves to glucose and as a result glucose becomes unstable*	*The molecules start moving*	*Carbon atoms bond together*
Causal explanation (6 items)	Infer why one event would cause another (glycolysis and novel domains)	What makes the event at Box C happen? 	At least one cause is given for the scenario	*The releasing energy of the pyruvate gave the H energy to move to the NAD, making NADH*	*The glucose [pyruvate] splitting up and thrusting everything back*	*Because NAD is coming over to NADH*	*Hydrogen bonds to NAD*
Linking (2 items)	Detect structural similarities between novel phenomena and those of glycolysis	How are these pictures of a rubber band similar to what happens to molecules in glycolysis? 	Stated similarities to glycolysis accurately map to the scenario structure	A. *This is where the glucose is coming together, even though it is not going the right way compared to glycolysis, it is still getting energy* B. *This is when the glucose is holding energy* C. *Again, even though it is the wrong way, this sort of has a "boom" type of effect, so it is releasing*	A. *Phosphates bonding to glucose, making it unstable* B. *Glucose is ready to split* C. *Glucose is split*	A. *Phosphates stretch over to glucose* B. *The glucose splitting up making pyruvates* C. *The molecules go back to start the process over again*	A. *Glucose is formed* B. *NAD and NADH is made* C. *Everything starts over*

Fig. 7 By design, there were three sources of inference for answering the causal explanation items, two of which depended on integrating information into the energy structure



to detect the underlying similarity within the two contexts while ignoring these surface-level differences.

The causal explanation items asked students to speculate about why events in cellular respiration and related processes might occur. Some of the items featured glycolysis processes taught in the content lesson and others featured processes to which students had not yet been exposed (e.g., Krebs Cycle; Calvin cycle). Regardless, at no time in the content lesson were students explicitly taught the causes of the events in the items. Nevertheless, these causes could be inferred, and as illustrated by Fig. 7, there were three sources of inference. The first, which was supported by both instructional conditions, was to interpret the scenario in terms of the springy or explosive actions of the molecule-manipulatives (Fig. 7, left). The second source, supported only by the analysis condition, was to interpret the scenario in terms of one or more element-level meanings within the energy transfer structure (see Fig. 7, center). The third, also supported only by the analysis condition, was to impose a causal relationship between elements in the energy transfer structure, for instance, the relationship that an energy gain must be supplied by an energy release (Fig. 7, right). Thus, even though all three sources of inference could lead to correct responses on this item type, only one source did not involve integration as depicted in Fig. 1.

Finally, the factual recall items checked whether recognition instruction was robust. These items asked students to describe events presented in the glycolysis lesson without regard to their causes (i.e., description, not explanation).

Coding Criteria

For the posttest, all items were coded as correct or incorrect. Table 1 shows the acceptance criterion and example responses for each item type. For the text annotation measure, the coding structure was slightly more complex. First, it divided the text into 12 annotatable phrases, which allowed coding how many phrases were annotated. Second, a general criterion of accuracy to the text was used to code whether or not each annotation was accurate (i.e., faithful to the text). As examples, the annotations "errr" (getting energy) and "ATP is formed" were accurate annotations of the phrase "This reaction puts phosphates onto four ADP molecules." Responses of "boom" or "ATP gives up its phosphate" were inaccurate. Finally, a third level of coding was applied, but only within the recognition condition because it involved a response characteristic that was not possible in the analysis condition. This was whether or not a given annotation was a copy of the text. The criterion for copying was having the same major words as the original text, including subsets of these words and variations on them. As an example, for the phrase "Glucose splits in half," the annotations "splits in half" and "glucose splitting" were copies, while "breaks in two" was not.

Coding Procedure

The procedure for coding both annotation and test item responses began with transcribing all of the responses to a spreadsheet. Then, two researchers independently coded a random sample of 25% of the data for a given test item or annotation phrase. If agreement was less than 90%, they adjudicated differences and repeated the procedure until the 90% threshold was met. Afterward, a single researcher coded the remaining responses for the item or phrase. Agreement on posttest items ranged from 90 to 100%, and Cohen's kappa, *K*, ranged from 0.71 to 1.00. Agreement on accuracy of annotation and copying also ranged from 90 to 100%, and *K* ranged from 0.89 to 1.00 and 0.65 to 1.00 respectively.

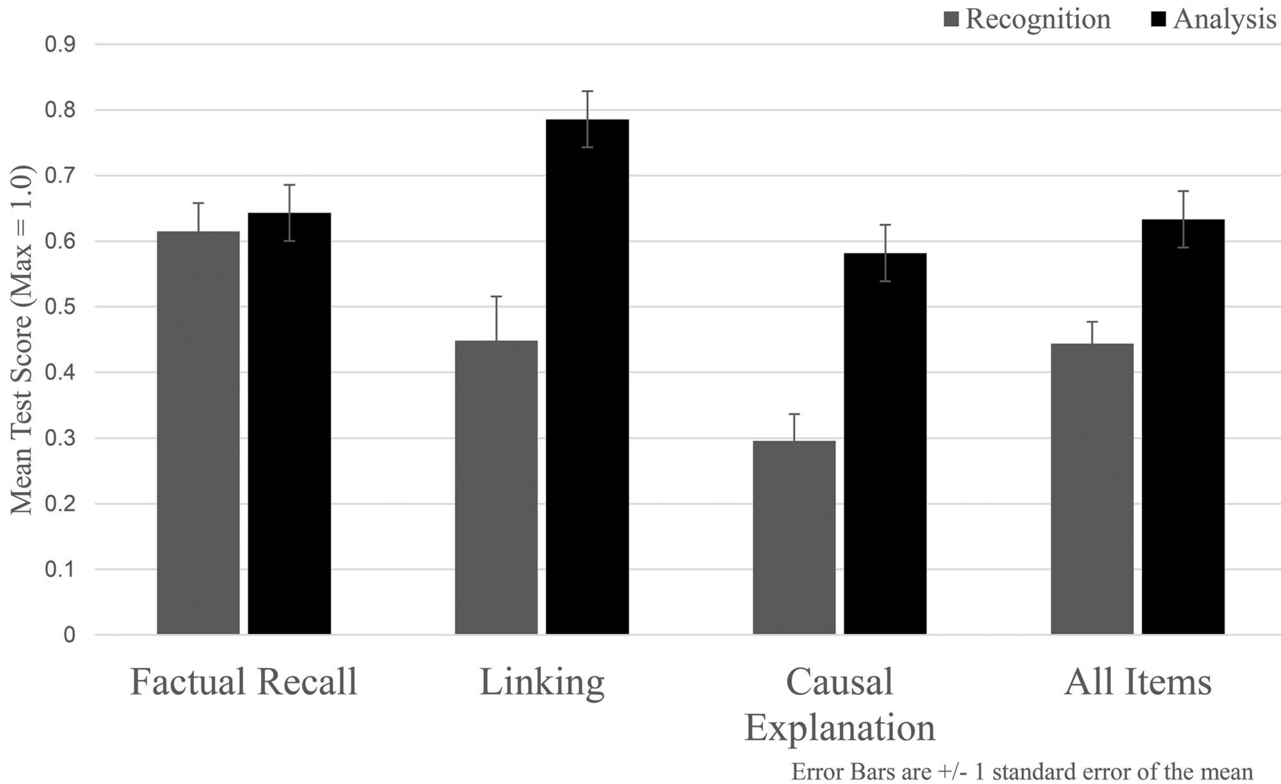


Fig. 8 Analysis and recognition students had identical posttest scores for factual recall, but analysis students had higher scores for linking, causal explanation, and the overall test

Results

Figure 8 compares the posttest scores (i.e., mean proportion of items correct) for the analysis and recognition conditions. Subscores are shown for factual recall, linking, and causal explanation followed by the overall test scores. On the overall test, analysis students scored higher than recognition students with an effect size, *d*, of 1.06 standard deviations. Looking at the subscores, analysis students outscored recognition students on linking and causal explanation, with large effects in both cases. For linking, the effect size was 1.04 standard deviations, $M_{\text{analysis}} = 0.786$, S.D. = 0.286; $M_{\text{recognition}} = 0.448$, S.D. = 0.362. For causal explanation, the effect size was 1.27 standard deviations, $M_{\text{analysis}} = 0.582$, S.D. = 0.226; $M_{\text{recognition}} = 0.296$, S.D. = 0.222. On factual recall, by contrast, the two conditions did equally well, indicating that neither group had an advantage in learning the material aspects of the molecules via the molecule-manipulatives. The overall pattern of results was statistically significant in a MANOVA in which the dependent variables were the means for the three measurement dimensions (factual recall, linking, causal explanation), and the independent variable was instructional condition (analysis, recognition). There was a statistically significant effect of instructional condition for the omnibus test, $F(3, 53) = 11.0$, $p < 0.001$, and also for the univariate tests of causal explanation and linking, $F(1, 55) = 23.1$, $p < 0.001$, and $F(1, 55) = 15.2$, $p < 0.001$ respectively. For factual recall, there was not a statistically significant difference, $F(1, 55) = 0.212$, $p = 0.647$, consistent with the two conditions' nearly identical means on this subscore. Thus, among two groups of students who were equally well prepared to remember facts about processes in the content lesson, one group was better prepared to explain why these processes occurred and to link them structurally to phenomena in other domains.

The annotation scores of the two groups are displayed in Table 2 as the mean number of correct and incorrect annotations made out of 12 opportunities. The table shows that

Table 2 Means (standard deviations) for text annotation out of 12 opportunities

Metric mean	Analysis	Recognition	*p*-value[a]
Total annotated	10.1** (1.79)	6.68 (2.67)	< .001
Total correct	9.25** (2.09)	5.79 (2.94)	< .001
Accuracy (Total correct/ Total annotated)	.911 (.093)	.816 (.258)	.093

[a]Difference in means, independent samples *t*-test

analysis students annotated more opportunities than recognition students, while both groups annotated with high accuracy. The difference in mean correct annotations was 1.36 standard deviations. A comparison of mean correct annotations at the beginning, middle, and end of the text revealed that the advantage of the analysis condition remained roughly the same throughout the measure. Thus, while both groups persisted in processing the text, the analysis students were more efficient at finding opportunities to annotate as they went along. Finally, there was a positive Pearson correlation between annotation score (i.e., mean correct annotations) and overall posttest score, indicating a common source of doing well at annotation and doing well on the test. The correlation was highest when both conditions were combined, $r = 0.715$, $p < 0.001$, showing that some of the covariance was driven by group differences. With the conditions separated, the correlations were moderate but still sizable, $r = 0.505$, $p = 0.012$ for analysis, $r = 0.625$, $p < 0.001$ for recognition. The larger correlation in the recognition group could reflect a restriction in range in the analysis group's annotation scores, which were close to the ceiling level.

Recall that we also coded for whether annotations were copies of phrases in the text, which was possible in the recognition group but not the analysis group. We found that 36% of recognition annotations were copied phrases. Looked at positively, this finding indicates that, most of the time, recognition students made efforts to annotate meaningfully using their own constructions and succeeded in these efforts. Looked at negatively, there was a good deal of nonconstructive annotation in the recognition group. By contrast, analysis students, who were constrained to make interpretive annotations (i.e., map the scenario to the model), could not have done this correctly without a thought process that was at least more constructive than copying the text.

Discussion

The results were that analysis students were successful at annotating the text, and they had attendant positive outcomes on the posttest. In the first half of the discussion that follows, these results are considered with respect to the research questions, namely whether these students' annotation of the texts was analysis, whether this process used a deep structure, and whether it was integrative. The second half of the discussion offers a reflection on the significance of the answers to these questions for instruction as well as an examination of limitations of the study and opportunities for improvement.

Was Text Annotation Analysis?

The first research question was how successful students were at analysis (or recognition) as a way of processing the descriptive narratives presented in the content lesson, and the measure of success was how completely and accurately students annotated the third of the three narratives. While the accuracy in both groups was comparable, analysis students annotated more of the information (i.e., they were more efficient than recognition students), and they did so without the fallback of copying phrases that was common in the recognition group. Thus, analysis students seemed to be successful. But was their annotation process actually analysis? On the surface it was, because they assigned events in glycolysis to particular elements in their energy transfer structure, as depicted in Fig. 1. On the other hand, it is unclear how much students actually thought of events they annotated in terms of abstract energy meanings in the structure, as should occur in analysis. Greater clarity on this point is bound up in the answers to the second and third research questions, which we will discuss next. However, there are two minor indicators of meaningful process in the annotation evidence alone that bear mentioning. The first was the sure-footedness with which the students deployed their energy structure. Analysis students mapped information from the text accurately to this structure, while meeting the significant cognitive demand of decoding the text (i.e., efficiently), and under the requirement for transformation in which each component of the energy structure had to encompass a variety of scenarios (see "Measures and Coding" for details). Accuracy and efficiency, we reason, would be unlikely under transformation if mapping the information was not meaningful. The second indicator was the positive correlation between annotation score and test score. Given that the test was relatively difficult, such a correlation would be unlikely if the annotation process was trivial. Thus, annotation was at least a substantial cognitive process.

While we had hoped that analysis students would annotate at least as much of the text as recognition students, they actually annotated more. Speculatively, this occurred because mapping to an energy structure students knew was an easier process than generating short descriptions from scratch. Indeed, the challenge posed by rephrasing the text may have been why recognition students sometimes copied it word for word. As a further contributor to difficulty, recognition students wrote more words than analysis students, a short phrase compared to a one-word label. In this sense, recognition students lacked the advantage, not only of a strong organizing framework, but also of a semantically dense representation (Buxton et al., 2019; Fang, 2006; Maton, 2013). Thus, the two groups' annotations cannot be interpreted as an apples-to-apples comparison. Analysis students were analyzing the descriptive narratives (we argue); recognition students were not. Given this fact, the data cannot be used to claim that using a big idea to annotate text, or more broadly, to make sense of information, would be more efficient than an otherwise similar process that did not use a big idea.

Were Students Analyzing with a Deep Structure?

The second research question concerned whether the set of energy transfer labels students used to annotate the text (i.e., getting, having, releasing energy) comprised a deep structure. The linking items on the posttest checked for this by measuring whether students could use either the model structure (analysis group) or the underlying structure of the manipulatives (recognition group) to detect structural similarities between glycolytic processes and processes in other domains. Importantly, these structural similarities were not present at the surface level of the items. For instance, the rubber band item depicted in Table 1 presented resistances and material movements that were in the opposite directions of those that students experienced in the molecule-manipulatives and/or built into their energy model. Therefore, students had to look past these surface-level differences to detect the similarities that lay beneath them (e.g., the need for effort). Since analysis students were successful in this (i.e., in comparison to recognition students), our inference is that the energy structure they used for annotating the text was a deep structure, meaning a set of conceptual constraints that apply under transformation to a range of situations (Gentner & Markman, 2006). Thus, the conditions were in place for analysis that would be integrative.

Was the Process Integrative?

The final research question concerned whether the transformations of enrichment and/or interconnection shown in Fig. 1 occurred, the measure of which was students' ability to speculate about causes of events in glycolysis and in related phenomena (e.g., Krebs cycle, Calvin cycle). These causes were not taught in the content lesson, but they could be inferred from one of three sources as shown in Fig. 7: the actions of the molecule-manipulatives, element meanings of the energy transfer structure (i.e., for enriching), element-to-element relations in the energy transfer structure (i.e., for interconnecting). Recognition students had access to the first of these sources, while analysis students had access to all three. Since analysis students were able to generate causal explanations more often than recognition students on the posttest, our by-hypothesis explanation is that some combination of enrichment or interconnection must have been operative, either during the text annotation or afterward, when students answered these items. A review of the sample item responses in Table 1 and Fig. 7 illustrates how they could do this using both meanings within elements in the energy structure (e.g., the powder keg nature of having energy) and between elements in the structure (e.g., the fact that having energy portends releasing energy).

Of course, and especially reasoning in isolation from all the other evidence, integration is far from the only possible reason for the analysis students' advantage on the causal explanation items. As just one possibility, the cognitive efficiency of energy terminology, which the analysis group had but the recognition group did not, probably made it easier for analysis students to interpret the questions and write down explanations. However, when considered with other evidence, particularly the evidence from the linking items that suggest that analysis students were using a deep structure, the argument for integration takes on a degree of triangulation. Nevertheless, our claim of integration via enriching and interconnecting would be stronger if it were supported by more direct evidence of whether these processes were occurring when students answered the items (e.g., evidence from think-aloud protocols).

The Value of Integration for Instruction

In asking and answering questions about whether students in the present study engaged in integrative analysis with a big idea, it is easy to neglect a more fundamental question lying beneath them. Namely, why is integrative analysis worth knowing about? In light of the present study's findings, there are three immediate answers to this question. The first is that integrative analysis comprises a use for big ideas that practitioners may not be acquainted with, namely as a tool with which students can make sense of scientific narratives that may otherwise come across as a parade of disjointed facts and descriptions. Thus, knowing about integrative analysis should help practitioners faced with teaching these kinds of narratives, for example, narratives for cellular respiration and photosynthesis, use big ideas more purposefully. The second answer is that integrative analysis can be incremental to teaching routines that practitioners may already use. Indeed, it may be transformative for routines that may otherwise be ineffective. The present study is a good example of this. The base instruction, annotating descriptive texts, was crudely didactic and therefore a questionable instructional process. Yet, integrative analysis rendered this process significantly useful as evidenced by both learning process and learning outcome advantages in the analysis group. The third answer is that integrative analysis with big ideas, particularly as defined in this article wherein the big idea is a deep structure that enriches and interconnects information, is a powerful learning process. This should be evident almost by definition, as others have pointed out (NRC, 2000, 2012).

However, the present study's finding that a short intervention helped students generate causal explanations and connect their learning to phenomena in other domains should help to underscore this important point.

Toward Richer Applications of Integrative Analysis

When considered separately from the modeling activity that preceded it, the use to which integrative analysis was put in the present study, annotating descriptive texts, did not constitute an exemplary learning activity. Although students had a useful knowledge tool (i.e., the big idea), and they used it for a critical learning process (i.e., analysis), there was no context for what they were doing, no authentic purpose for it, and no scope for how to do it. This fact raises the question of whether integrative analysis could be employed within richer learning processes, especially in ways that would empower students to direct the course of their learning (Schwarz et al., 2017) and to shape knowledge and knowledge-building practices (Damşa et al., 2010; Ford, 2008; Stroupe, 2014). Conjecturally, we think this would be possible by meeting four instructional requirements. The first would be to focus the analysis on an engaging phenomenon or anchoring event rather than an abstract process (Symeonidis & Schwarz, 2016; Windschitl et al., 2012). Keeping to the present context of cellular respiration, an example might be a scenario comparing an animal like a cheetah noted for high effort in spurts, to an animal like a wolf, characterized by lower level but more sustained effort. Within this scenario, integrative analysis could be a useful tool for investigating why the two animals differed. The second requirement would be to afford students multiple forms of representation within integrative analysis, as is done in modeling practice to support dispersion in student engagement with phenomena (Gericke et al., 2013; Harrison & Treagust, 2000). As an example, in a newer version of our project, students have the freedom to use energy representations at three levels of abstraction: rubber band model sticks, energy words, and numerical quantities. Working in small groups, students choose which of these representations they will use at various points of analysis, and different choices raise different sorts of questions about the phenomena. Thus, students have a degree of control over what to explain, and how to explain it, and different groups end up shedding light on differing aspects of the phenomenon under study. This scenario points to the third requirement, which would be to support differences in points of view within integrative analysis, and to engage students with them through talk and dialog so as to rehearse, refine, and internalize meaning (Mortimer & Scott, 2003), especially meaning that is fragmented and inconsistently activated (Sikorski & Hammer, 2017). As just outlined, fruitful dialogic opportunities could arise from both the process and products of analysis. The final requirement, with the goal of making integrative analysis an epistemic practice, would be to support students in problematizing model fit as they conduct their analysis. In our current project, this has occurred when students have encountered a phenomenon that fits their concept of energy transfer in the abstract, but does not conform to their existing representational tools. Students' spontaneous response has been to repurpose the tools, taking a variety of approaches that lead to valuable points of comparison, both within small group activities and afterward, in whole-class discussion.

Energy as a Big Idea for Analysis in Cellular Energetics

As a big idea, energy transfer has great potential as a tool for learning in the biology domains of cellular respiration and photosynthesis. However, while prominent curriculum standards documents have acted to tap this potential (e.g., ACARA, 2012; College Board, 2011; NGSS Lead States, 2013), some of these reforms seem to have undercut their efforts by employing concepts of energy transfer that come closer to encompassing ideas than true big ideas. As discussed at the front of this article, true big ideas are abstract structures consisting of meaningful elements with specified relationships between them (NRC, 2000)--what others have characterized as providing links between ideas (Mitchell et al., 2017) or expressing relationships (Windschitl et al., 2020). For energy transfer, the required elements and relationships were articulated nicely by Swackhamer (2005) as being "changes that energy transfers may cause in the interacting systems. No energy transfer can occur without at least two changes, at least one in the giver and at least one in the receiver" (p. 26). Unfortunately, neither of the two most prominent curriculum standards in the USA, both of which have explicitly organized cellular respiration and photosynthesis around energy transfer as a purported big idea, fully meets Swackhamer's requirement for defining energy transfer relationships. We are referring to the Advanced Placement (AP) Biology standards (College Board, 2011, 2019) and the Next Generation State Standards (NGSS Lead States, 2013). For the AP Biology standards, the problem is especially obvious, as the standards organize the curriculum around energy transfer as a big idea, but they do not venture any definition of what energy transfer is. Concomitantly, while energy transfer is salient in summary curriculum statements, it is conspicuously absent, or at best incidental to, detailed specifications of what students should know. This scenario is emblematic of an organizing concept that is an encompassing idea, not a big idea.

For the NGSS (NGSS Lead States, 2013), the lack of specification of relationships is more subtle. In this case, the standards do incorporate energy transfer relationships within their core ideas for cellular energetics, but to us at least, these relationships only partially meet Swackhamer's requirements to specify system changes by which energy changes. The core idea for cellular respiration exemplifies this situation:

> HS-LS1-7. Use a model to illustrate that cellular respiration is a chemical process whereby the bonds of food molecules and oxygen molecules are broken and bonds in new compounds are formed resulting in a net transfer of energy (p. 82).

Here, there is a defined change in the interacting molecular systems (formation of new compounds when food molecules break down) by which changes in energy occur. On closer examination, however, there is no specification of how energy changes in relation to these molecular system changes, or vice versa. Thus, the standard does not define an energy transfer structure, but rather a material configuration structure in which the fact of energy transfer is entailed. Put another way, the standard is not an idea *about energy* that will enable students to think of molecular processes in terms of its abstract structure.

Stepping back to consider both AP and NGSS standards together, if our critique is even partially valid, then there is much opportunity for improving them. More fundamentally, the fact of this opportunity would indicate that broadly held understandings of what big ideas should do for instruction, and how they should be constructed to do this, are far from fully developed in current educational practice.

Limitations of the Energy Model Used in the Study

Having pointed out shortfalls in others' concepts of energy transfer, we also need to recognize our own. Ours was an intuitive energy model in which energy was input through effort to bring submolecular particles together against a springy resistance, was stored by holding them together against the resistance, and was released by letting the particles be dispersed by the resistance. While this model went some distance toward meeting Swackhamer's requirement to explain changes in the interacting material systems by which changes in energy occurred, it pulled for thinking of energy storage "in" chemical bonds of individual molecules,

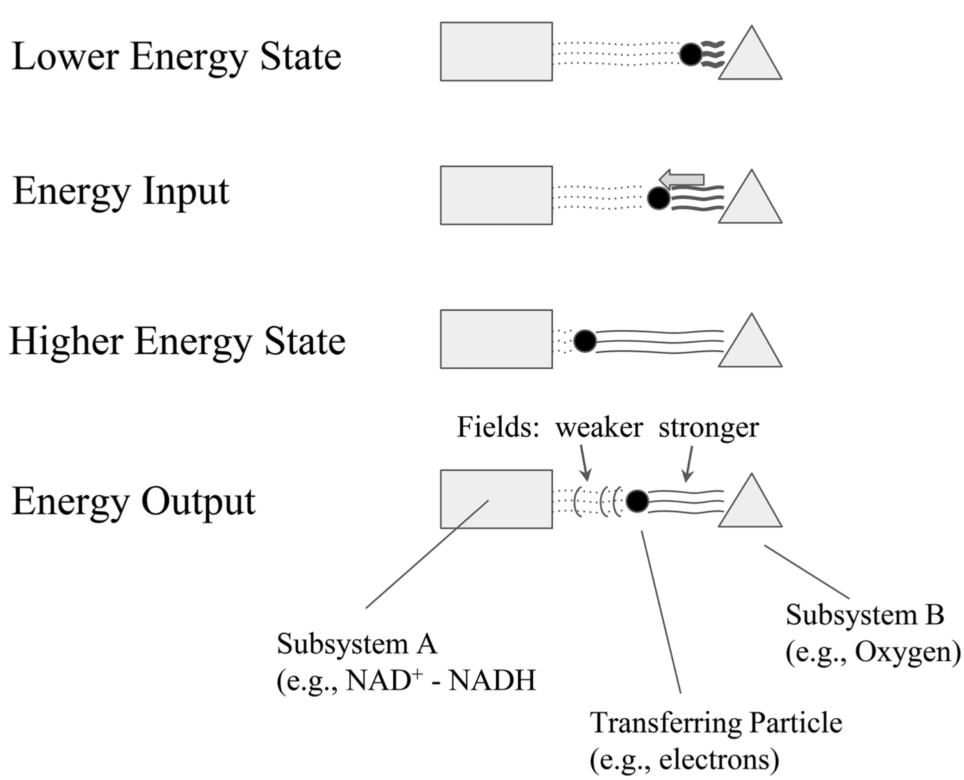


Fig. 9 The energy model used in the present study (left) located energy within molecules. An improved version (right) would locate energy in the electro-magnetic fields of systems of molecules

instead of in systems of molecules with bonds at higher and lower negative energies. While the energy-in-bonds concept is not uncommon in biology teaching as a useful shortcut (Cooper & Klymkowsky, 2013; Dreyfuss et al., 2014), our use of it was more egregious than most because the incorrect force relationships implied by energy in bonds were made explicit in our energy model (i.e., repulsive forces). Thus, many would say that we installed a misconception (Barker & Millar, 2000; Boo, 1998; Cooper & Klymkowsky, 2013; Novick, 1976; Storey, 1992).[4] This fact limits what the present study does and does not contribute to practice. Namely, the study does contribute a concept of teaching with big ideas, integrative analysis, but it does not contribute an unproblematic way to use this concept with the featured big idea, energy transfer.

Figure 9 shows a possible improvement on the present study's energy model that we have developed within our current project to correct the direction of forces through which energy changes occur (i.e., represent them as attractive). Here, potential energy is attributed to distortions of electromagnetic fields that occur when particles are forced to move against the fields (Fortus et al., 2019; Nordine et al., 2018). When paired with macroscopic experiences of electric and magnetic fields (Dreyfus et al., 2014), a model like this should help students intuit that energy must be input to pull atoms away from molecules (breaking bonds), and energy is output when molecules are formed (making bonds). Additionally, in the US system at least, this kind of model would accord with an explicit field energy concept in the disciplinary core ideas for physical science and the crosscutting concept for energy (NRC, 2012).

On the other hand, a model like that of Fig. 9 would continue to have shortcomings for teaching about energy in cellular systems, some of them serious. For example, kinetic energy, arguably the central figure in energy transfer, would be at best crudely represented, and this would lead to uncomfortable omissions and distortions in teachers' and students' accounts of energy transfer. One way around this and other problems would be to model biochemical energy indirectly, by modeling physical systems that students could use to think analogically and abstractly about energy in biochemical systems.[5] As an example of this approach, Dreyfus et al. (2014) taught college students a mechanical energy well model and supported them to think of biomolecules as having an analogous energy structure. There is much wisdom in this indirect approach. Still, there comes the inevitable point when students ask, "What are we saying happens in this biomolecule?", and a teacher taking the indirect approach would find it difficult to give a straight answer. Consequently, and bearing in mind that no model will be perfect (Box, 1976; Giere, 2004; Hesse, 2008; Mäki, 2011), we continue to favor a more direct modeling strategy. Admittedly, this sets a difficult path going forward because it involves negotiating disciplinary norms and priorities that have only begun to be charted (Cooper, 2020; Redish & Cooke, 2013; Redish et al., 2014). Further complicating these negotiations will be the fact that the best ideas for teaching science often differ from the best ideas for doing science (Loughran et al., 2006).

Conclusion

Voicing a concern that should probably be more widely acknowledged, Martin and Nock (2018) observed that there is a significant gap between interpreting a curriculum that is organized around big ideas (in their case NGSS Lead States, 2013), and having a vision of teaching and learning in which big ideas play an integrative role:

> When unpacking the NGSS with teachers, we continue to observe a strong bias toward a mindset of coverage of the lists of performance expectations and evidence statements, with much greater uncertainty about how to implement integration of ideas (pp. 2134–2135).

The authors also pointed out that practical strategies to support teaching with big ideas are in short supply. Taking these concerns at face value, it seems unlikely that mere directives for teachers to cluster content around big ideas will achieve curriculum reformers' goal of making big ideas function as tools for understanding. Rather, instructional designers and teachers need a workable instructional theory that explains precisely what uses of big ideas are possible and how to support them. In the present article, we have proposed a start on such a theory by defining integrative analysis as one important use of big ideas and illustrating the value of this use for science instruction. We hope this contribution will inspire and support a further investment in understanding teaching and learning with big ideas.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10956-023-10040-5.

Acknowledgements We wish to acknowledge team members Eric Kirk who helped with the initial development of instructional materials and Michael Elgin Leary who supported in data collection and analysis. We are also grateful for the commentary of two anonymous reviewers who helped us understand our work more clearly.

Funding This material is based on research supported by the National Science Foundation under Grant Numbers 1720996, 2010334, and 2010223.

[4] Researchers classify this way of thinking as a misconception based on evidence of its persistence through instruction, though some are circumspect about why this is so (Teichert & Stacy, 2002).

[5] We want to thank an anonymous reviewer for pointing out this possibility.sssss.

Data Availability The data sets for the current study are not publicly available due to the fact that they are part of research in progress. However, they will be made available from the corresponding author on reasonable request.

Declarations

Ethical Approval This research was conducted under the approval of Institutional Review Boards at The University of Alabama (IRB #17-OR-415-R1 and IRB #20–08-328) and the University of Georgia (IRB #00004700 and IRB #00002425).

Informed Consent Informed consent was sought for all participants for participation in the project, and for their de-identified data to be used for publication. The research was performed in accordance with the ethical standards as laid out in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Competing Interest The authors declare no competing interests.

References

American Association for the Advancement of Science. (1990). *Science for all Americans*. Oxford University Press.

Australian Curriculum Assessment and Reporting Authority [ACARA]. (2012). Shape of the Australian science curriculum. Retrieved May 15, 2021, from https://www.australiancurriculum.edu.au/

Barker, V., & Millar, R. (2000). Students' reasoning about basic chemical thermodynamics and chemical bonding: What changes occur during a context-based post-16 chemistry course? *International Journal of Science Education, 22*(11), 1171–1200.

Beaney, M. (2014). Analysis. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2021 Edition). https://plato.stanford.edu/archives/sum2021/entries/analysis/

Boo, H. K. (1998). Students' understandings of chemical bonds and the energetics of chemical reactions. *Journal of Research in Science Teaching, 35*(5), 569–581.

Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association, 71*(356), 791–799.

Buxton, C., Harman, R., Cardozo-Gaibisso, L., Jiang, L., Bui, K., & Allexsaht-Snider, M. (2019). Understanding science and language connections: New approaches to assessment with bilingual learners. *Research in Science Education, 49*(4), 977–988.

Capps, D. K., & Shemwell, J. T. (2020). Moving beyond the model as a copy problem: Investigating the utility of teaching about structure-preserving transformations in the model-referent relationship. *International Journal of Science Education, 42*(12), 2008–2031.

Castro-Faix, M., Duncan, R. G., & Choi, J. (2021). Data-driven refinements of a genetics learning progression. *Journal of Research in Science Teaching, 58*(1), 3–39.

Chalmers, C., Carter, M. L., Cooper, T., & Nason, R. (2017). Implementing "big ideas" to advance the teaching and learning of science, technology, engineering, and mathematics (STEM). *International Journal of Science and Mathematics Education, 15*(1), 25–43.

Chase, C. C., Malkiewich, L., & Kumar, S. A. (2019). Learning to notice science concepts in engineering activities and transfer situations. *Science Education, 103*(2), 440–471.

Cheng, M. F., & Brown, D. E. (2015). The role of scientific modeling criteria in advancing students' explanatory ideas of magnetism. *Journal of Research in Science Teaching, 52*(8), 1053–1081.

Chi, M. T., & VanLehn, K. A. (2012). Seeing deep structure from the interactions of surface features. *Educational Psychologist, 47*(3), 177–188.

Clement, J. (2008). The role of explanatory models in teaching for conceptual change. *International Handbook of Research on Conceptual Change, 1*, 417–452.

Clement, J. J., & Steinberg, M. S. (2002). Step-wise evolution of mental models of electric circuits: A "learning-aloud" case study. *The Journal of the Learning Sciences, 11*(4), 389–452.

Close, H. G., & Scherr, R. E. (2015). Enacting Conceptual Metaphor through Blending: Learning activities embodying the substance metaphor for energy. *International Journal of Science Education, 37*(5–6), 839–866.

College Board. (2011). *AP biology curriculum Framework 2012–2013*. Retrieved from https://secure-media.collegeboard.org/digitalServices/pdf/ap/10b_2727_AP_Biology_CF_WEB_110128.pdf

College Board. (2019). *AP biology course and exam description effective Fall 2020*. Retrieved from https://apcentral.collegeboard.org/media/pdf/ap-biology-course-and-exam-description.pdf

Cooper, M. M. (2020). The Crosscutting Concepts: Critical Component or "Third Wheel" of Three-Dimensional Learning? *Journal of Chemical Education, 97*(4), 903–909.

Cooper, M. M., & Klymkowsky, M. W. (2013). The trouble with chemical energy: Why understanding bond energies requires an interdisciplinary systems approach. *CBE—Life Sciences Education*, *12*(2), 306–312.

Damşa, C. I., Kirschner, P. A., Andriessen, J. E., Erkens, G., & Sins, P. H. (2010). Shared epistemic agency: An empirical study of an emergent construct. *The Journal of the Learning Sciences, 19*(2), 143–186.

DiSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10*(2–3), 105–225.

Dreyfus, B. W., Gouvea, J., Geller, B. D., Sawtelle, V., Turpen, C., & Redish, E. F. (2014). Chemical energy in an introductory physics course for the life sciences. *American Journal of Physics, 82*(5), 403–411.

Duit, R. (1987). Should energy be illustrated as something quasi-material? *International Journal of Science Education, 9*(2), 139–145.

Fang, Z. (2006). The language demands of science reading in middle school. *International Journal of Science Education, 28*(5), 491–520.

Ford, M. (2008). Disciplinary authority and accountability in scientific practice and learning. *Science Education, 92*(3), 404–423.

Fortus, D., Kubsch, M., Bielik, T., Krajcik, J., Lehavi, Y., Neumann, K., & Touitou, I. (2019). Systems, transfer, and fields: Evaluating a new approach to energy instruction. *Journal of Research in Science Teaching*, *56*(10), 1341–1361.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*(2), 155–170.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34*(5), 752–775.

Gentner, D., & Kurtz, K. (2005). Learning and using relational categories. In W. K. Ahn, R. L. , Love, B.C., Markman, A.B., & Wolff, P.W. (Eds.), Categorization inside and outside the lab. Washington, DC: American Psychological Association.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95*(2), 393–408.

Gentner, D., & Markman, A. B. (2006). Defining structural similarity. *The Journal of Cognitive Science, 6*(1), 1–20.

Gericke, N., Hagberg, M., & Jorde, D. (2013). Upper secondary students' understanding of the use of multiple models in biology textbooks—The importance of conceptual variation and incommensurability. *Research in Science Education*, 43, 755–780.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*(3), 306–355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1–38.

Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science, 71*(5), 742–752.

Gupta, A., Hammer, D., & Redish, E. F. (2010). The case for dynamic models of learners' ontologies in physics. *The Journal of the Learning Sciences, 19*(3), 285–321.

Hammer, D., & Manz, E. (2019). Odd ideas about learning science: A response to Osborne. *Science Education, 103*(5), 1289–1293.

Harlen, W. (Ed.). (2010). *Principles and big ideas of science education*. Association for Science Education.

Harrison, A. G., & Treagust, D. F. (2000). Learning about atoms, molecules, and chemical bonds: A case study of multiple-model use in grade 11 chemistry. *Science Education, 84*(3), 352–381.

Hesse, M. (2008). Models and analogies. In W. H. Newton (Ed.), *A companion to the philosophy of science* (pp. 299–307). Blackwell Publishers Ltd.

Kali, Y., Orion, N., & Eylon, B. S. (2003). Effect of knowledge integration activities on students' perception of the Earth's crust as a cyclic system. *Journal of Research in Science Teaching, 40*(6), 545–565.

Kapon, S., & diSessa, A. A. (2012). Reasoning through instructional analogies. *Cognition and Instruction, 30*(3), 261–310.

Kellman, P. J., Massey, C. M., & Son, J. Y. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science, 2*(2), 285–305.

Kuo, E., & Wieman, C. E. (2016). Toward instructional design principles: Inducing Faraday's law with contrasting cases. *Physical Review Physics Education Research, 12*(1), 010128.

Loughran, J. J., Berry, A., & Mulhall, P. (2006). *Understanding and developing science teachers' pedagogical content knowledge*. Brill.

Lesh, R. A., & Doerr, H. M. (2003). *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching*. Routledge.

Mäki, U. (2011). Models and the locus of their truth. *Synthese, 180*(1), 47–63.

Martin, J. D., & Nock, K. A. (2018). A Nonlinear, "Sticky" Web of Study for Chemistry: A Graphical Curricular Tool for Teaching and Learning Chemistry Built upon the Interconnection of Core Chemical Principles. *Journal of Chemical Education, 95*(12), 2134–2140.

Maton, K. (2013). Making semantic waves: A key to cumulative knowledge-building. *Linguistics and Education, 24*(1), 8–22.

Mitchell, I., Keast, S., Panizzon, D., & Mitchell, J. (2017). Using 'big ideas' to enhance teaching and student learning. *Teachers and Teaching, 23*(5), 596–610.

Mortimer, E., & Scott, P. (2003). *Meaning Making In Secondary Science Classrooms*. McGraw-Hill Education.

National Research Council. (2000). How people learn: Brain, mind, experience, and school: Expanded edition.

National Research Council. (2002). *Learning and Understanding: Improving Advanced Study of Mathematics and Science in U.S. High Schools: Report of the Content Panel for Biology.* The National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States*. The National Academies Press.

Nordine, J., Fortus, D., Lehavi, Y., Neumann, K., & Krajcik, J. (2018). Modelling energy transfers between systems to support energy knowledge in use. *Studies in Science Education, 54*(2), 177–206.

Novick, S. (1976). No energy storage in chemical bonds. *Journal of Biological Education, 10*(3), 116–118.

Osborne, J. (2011). Science teaching methods: A rationale for practices. *School Science Review, 93*(343), 93–103.

Osborne, J. (2014). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education, 25*(2), 177–196.

Patro, E. T. (2008). Teaching aerobic cell respiration using the 5 Es. *The American Biology Teacher, 70*(2), 85–87.

Plummer, J. D., Palma, C., Rubin, K., Flarend, A., Ong, Y. S., Ghent, C., & Furman, T. (2020). Evaluating a learning progression for the solar system: Progress along gravity and dynamical properties dimensions. *Science Education*, *104*(3), 530–554.

Redish, E. F., Bauer, C., Carleton, K. L., Cooke, T. J., Cooper, M., Crouch, C. H., & Klymkowsky, M. W. (2014). NEXUS/Physics: An interdisciplinary repurposing of physics for biologists. *American Journal of Physics*, *82*(5), 368–377.

Redish, E. F., & Cooke, T. J. (2013). Learning each other's ropes: negotiating interdisciplinary authenticity. *CBE—Life Sciences Education*, *12*(2), 175–186.

Ross, P. M., Tronson, D. A., & Ritchie, R. J. (2008). Increasing conceptual understanding of glycolysis & the Krebs cycle using role-play. *The American Biology Teacher, 70*(3), 163–169.

Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education, 92*(3), 499–525.

Scherr, R. E., Close, H. G., McKagan, S. B., & Vokos, S. (2012). Representing energy. I. Representing a substance ontology for energy. *Physical Review Special Topics-Physics Education Research*, *8*(2), 020114.

Schmidt, W. H., McKnight, C. C., & Raizen, S. (Eds.). (2007). *A splintered vision: An investigation of US science and mathematics education* (Vol. 3). Springer Science & Business Media.

Schmidt, W. H., McKnight, C. C., Valverde, G., Houang, R. T., & Wiley, D. E. (Eds.). (1997). *Many visions, many aims: A cross-national investigation of curricular intentions in school mathematics* (Vol. 1). Springer Science & Business Media.

Scholer, A. M., & Hatton, M. (2008). An evaluation of the efficacy of a laboratory exercise on cellular respiration. *Journal of College Science Teaching, 38*(1), 40.

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103*(4), 759–775.

Schwarz, C. V., Passmore, C., & Reiser, B. J. (2017). *Helping students make sense of the world using next generation science and engineering practices*. NSTA Press.

Shemwell, J. T., Chase, C. C., & Schwartz, D. L. (2015). Seeking the general explanation: A test of inductive activities for learning and transfer. *Journal of Research in Science Teaching, 52*(1), 58–83.

Sikorski, T. R., & Hammer, D. (2017). Looking for coherence in science curriculum. *Science Education, 101*(6), 929–943.

Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). Implications of Research on Children's Learning for Standards and Assessment: A Proposed Learning Progression for Matter and the Atomic-Molecular Theory. *Measurement: Interdisciplinary Research and Perspectives*, *4*(1–2), 1–98.

Smith, J. P., III., & Girod, M. (2003). John Dewey & psychologizing the subject-matter: Big ideas, ambitious teaching, and teacher education. *Teaching and Teacher Education, 19*(3), 295–307.

Songer, C. J., & Mintzes, J. J. (1994). Understanding cellular respiration: An analysis of conceptual change in college biology. *Journal of Research in Science Teaching, 31*(6), 621–637.

Songer, N. B. (1989). *Promoting integration of instructed and natural world knowledge in thermodynamics*. University of California.

Stevens, S. Y., Delgado, C., & Krajcik, J. S. (2010). Developing a hypothetical multi-dimensional learning progression for the nature of matter. *Journal of Research in Science Teaching, 47*(6), 687–715.

Storey, R. D. (1992). Textbook errors & misconceptions in biology: Cell energetics. *The American Biology Teacher, 54*(3), 161–166.

Stroupe, D. (2014). Examining classroom science practice communities: How teachers and students negotiate epistemic agency and learn science-as-practice. *Science Education, 98*(3), 487–516.

Swackhamer, G. (2005). Cognitive resources for understanding energy. Retrieved January 10, 2022, from http://modeling.asu.edu/modeling/RESOURCES_10-03.pdf

Symeonidis, V., & Schwarz, J. F. (2016). Phenomenon-based teaching and learning through the pedagogical lenses of phenomenology: The recent curriculum reform in Finland. In *Forum Oświatowe* (Vol. 28, No. 2 (56), pp. 31–47). University of Lower Silesia.

Teichert, M. A., & Stacy, A. M. (2002). Promoting understanding of chemical bonding and spontaneity through student explanation and integration of ideas. *Journal of Research in Science Teaching, 39*(6), 464–496.

Todd, A., & Kenyon, L. (2016). Empirical refinements of a molecular genetics learning progression: The molecular constructs. *Journal of Research in Science Teaching, 53*(9), 1385–1418.

Wiggins, G., & McTighe, J. (2005).*Understanding by design*. (Expanded 2nd ed.). ASCD.

White, P. J. (2016). Molecular sculpting: Active learning of subcellular systems & processes. *The American Biology Teacher, 78*(6), 482–491.

Windschitl, M., Thompson, J., & Braaten, M. (2020). *Ambitious science teaching*. Harvard Education Press.

Windschitl, M., Thompson, J., Braaten, M., & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Science Education, 96*(5), 878–903.