Educational Psychology Review (2025) 37:18
https://doi.org/10.1007/s10648-025-09993-3

INTERVENTION STUDY

Does Instruction-First or Problem-Solving-First Depend on Learners' Prior Knowledge?

Cheng-Wen He[1] · **Logan Fiorella**[2] · **Paula P. Lemons**[3]

Accepted: 7 February 2025 / Published online: 26 February 2025

Abstract

This study tested competing theories about the effectiveness of different instructional sequences for learners with different levels of prior knowledge. Across two classroom experiments, undergraduates learned about noncovalent interactions in biochemistry by either receiving explicit instruction before problem-solving (I-PS group) or engaging in problem-solving before explicit instruction (PS-I group). Then all students completed near- and far-transfer tests on the material. In Experiment 1, participants were introductory biology students ($n = 367$), who had relatively low prior knowledge of the topic. Results indicated that the PS-I group significantly outperformed the I-PS group on the near-transfer test, providing support for productive failure. In Experiment 2, participants were biochemistry students ($n = 138$), who had relatively higher prior knowledge of the topic. In contrast to Experiment 1, results indicated that the I-PS group significantly outperformed the PS-I group, providing support for cognitive load theory. Neither experiment showed significant effects of instructional sequences on the far-transfer test. Overall, the findings suggest the effects of instructional sequences on students with different levels of topic-specific prior knowledge may not be as straightforward as existing theories suggest.

Keywords Instructional design · Cognitive load theory · Productive failure · Biology education

✉ Paula P. Lemons
plemons@uga.edu

[1] Department of Mathematics, Science and Social Studies, University of Georgia, Athens, USA

[2] Department of Educational Psychology, University of Georgia, Athens, USA

[3] Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, USA

Introduction

Researchers in cognitive science and educational psychology have provided substantial empirical evidence for the benefits of explicit instruction as well as opportunities for active engagement in problem-solving (Kapur & Bielaczyc, 2012; Sweller et al., 2011). Yet, an unresolved debate is whether and under what conditions students should receive explicit instruction *before* or *after* engaging in problem-solving (Chen & Kalyuga, 2020; Kalyuga & Singh, 2016; Kapur, 2014; Sinha & Kapur, 2021). Proponents of instruction-first (I-PS) approaches claim that explicit instruction should precede problem-solving to avoid cognitive overload, especially for students with lower prior knowledge (Sweller et al., 2011). Proponents of problem-solving-first (PS-I) approaches assert that problem-solving activities should precede explicit instruction to better prepare students for future learning, regardless of students' prior knowledge levels (Kapur, 2016; Schwartz & Martin, 2004).

One central component of the debate between I-PS and PS-I is the potential moderating role of students' prior knowledge (Ashman et al., 2020; Chen & Kalyuga, 2020; Zhang & Sweller, 2024). Existing theories yield competing predictions, yet there is currently very limited empirical evidence directly testing how instructional sequences differentially affect students with different levels of prior knowledge. The present study addresses this gap by comparing the effects of I-PS and PS-I for students with relatively lower (Experiment 1) or higher (Experiment 2) levels of prior knowledge using the same learning materials in the domain of biochemistry. Here we focus on the role of *topic-specific knowledge* (i.e., knowledge related to the content of the to-be-learned material; e.g., non-covalent interactions) rather than *general domain knowledge* (i.e., broader knowledge related to a particular field of study; e.g., biology; see McCarthy & McNamara, 2021).

Theoretical Background

Cognitive Load Theory

The I-PS approach is consistent with cognitive load theory, which posits that instruction should avoid overloading the capacity of students' limited working memory (Sweller et al., 2019). Specifically, instruction should reduce extraneous cognitive load (i.e., the load irrelevant to learning and caused by how the material is presented) so that cognitive resources can be devoted to intrinsic cognitive load (i.e., the load required for processing the inherent difficulty of the material) (Sweller et al., 2011). One common way to reduce extraneous load is for students to study worked examples before attempting problems on their own (Sweller et al., 2011). By showing students the full steps and solution to a problem, instructors can focus students' resources on building expertise (i.e., on developing schemas for solving specific types of problems) before asking them to engage in independent problem-solving (e.g., Sweller & Cooper, 1985). For example, Matlen and Klahr (2013) found that third graders demonstrated a better understanding of experimentation skills when instruction began with viewing explicit demonstrations compared to when instruction began with independent problem-solving.

Importantly, worked examples often become less effective (and eventually can even impair learning) as learners' expertise increases (Sweller et al., 2019), a pattern often referred to as the expertise reversal effect (Kalyuga, 2007). As expertise develops, worked examples become redundant with learners' previously-acquired schemas in long-term memory (Sweller et al., 2011). This requires learners to use their cognitive resources to relate and reconcile the information that already exists in their long-term memory, creating extraneous load and reducing the capacity for acquiring new knowledge. Thus, it may be more productive for advanced learners who have sufficient experience in a domain to generate a problem solution on their own instead of studying a worked example (Kalyuga et al., 2001). Overall, cognitive load theory suggests explicit instruction should precede problem-solving activities, particularly for novice learners. As students' knowledge develops, an I-PS approach may become less effective, and eventually—if students have acquired the appropriate problem-solving schemas—a PS-I approach may be more appropriate.

Preparation for Future Learning and Productive Failure

The PS-I approach is consistent with theories of preparation for future learning and productive failure (Kapur, 2016), which emphasize the value of challenging problem-solving experiences during learning (e.g., Bjork, 2017). Initial problem-solving experiences serve to reveal knowledge gaps, activate prior knowledge, and thereby facilitate a deeper understanding of the material (Kapur, 2016; Schwartz & Martin, 2004). Specifically, these theories suggest that problem-solving activities such as inventing methods, creating multiple representations and solutions, and analyzing contrasting cases can help learners identify and understand the critical features of problems (Kapur & Bielaczyc, 2012). This understanding, in turn, aids learners in recognizing these features during subsequent instruction and in future contexts (Schwartz & Martin, 2004).

Productive failure is a specific form of PS-I in which the experience of failure during an initial problem-solving phase prepares students to benefit more from subsequent instruction. The problem-solving phase prompts students to generate potential (and often incorrect) solutions on their own, thereby activating their relevant prior knowledge (DeCaro & Rittle-Johnson, 2012; Loibl et al., 2017). It also enhances their awareness of the problem context and their knowledge gaps (DeCaro & Rittle-Johnson, 2012; Loibl & Rummel, 2014; Loibl et al., 2017), focusing their attention on the deeper patterns over superficial aspects of the problem (Kapur & Bielaczyc, 2012). Furthermore, such preparatory problem-solving can increase students' interest in learning canonical solutions in the subsequent instruction phase (Lamnina & Chase, 2019). Taken together, these mechanisms suggest that a challenging initial problem-solving phase should precede explicit instruction, which supports the consolidation of knowledge (Kapur, 2016; Kapur & Bielaczyc, 2012).

Evidence and Limitations of Prior Studies Comparing Instructional Sequences

As Kapur (2016) argued, the evidence cited in favor of explicit forms of instruction like I-PS often comes from studies comparing explicit instruction with weak

controls, such as comparing worked examples to a condition that engages learners in independent problem-solving without guidance or feedback. Kapur argues that unguided problem-solving should be followed by explicit instruction that consolidates learning. Using this design, a growing body of evidence has shown that the PS-I approach can be more effective than the I-PS approach (Chowrira et al., 2019; Jacobson et al., 2017; Kapur, 2014; Kapur & Bielaczyc, 2012; Weaver et al., 2018). For instance, one study compared the effects of instructional sequences in a large undergraduate physics class, finding that the PS-I students exhibited better conceptual understanding and equal procedural knowledge than I-PS students (Weaver et al., 2018). This finding aligns with the notion that PS-I is particularly effective for conceptual understanding and transfer because it facilitates noticing and encoding deep structure and critical problem features (Kapur, 2014, 2016; Schwartz et al., 2011; Sinha & Kapur, 2021).

However, others have argued that comparative studies of I-PS and PS-I often have methodological limitations and that the effects depend on moderating factors (Chen and Kalyuga (2020). For instance, some studies implemented different learning materials for the I-PS and PS-I interventions (Dubovi, 2018), or students were taught by different instructors for the two interventions (Jacobson et al., 2017). Furthermore, the effectiveness of I-PS and PS-I may be moderated by factors such as the complexity of the learning materials and/or learners' prior knowledge (Chen & Kalyuga, 2020). A study by Ashman and colleagues (2020) compared the effectiveness of I-PS and PS-I on 5th-grade students' learning about light energy efficiency with learning materials that were either lower or higher complexity. For lower complexity materials, I-PS showed higher performance on problems similar to those used during instruction but no differences on transfer problems compared to PS-I. For higher complexity materials, I-PS was significantly more effective than PS-I for problems similar to those used during instruction and for transfer problems (Ashman et al., 2020).

Similarly, researchers have proposed students' level of prior knowledge as a potential moderator of the effects of I-PS and PS-I (Chen & Kalyuga, 2020; Sinha & Kapur, 2021). However, this moderating effect is based on comparisons across studies rather than controlled experiments directly comparing I-PS and PS-I with learners of different levels of prior knowledge using the same learning materials. Furthermore, prior research overwhelmingly includes participants classified as having low prior knowledge (Chen & Kalyuga, 2020). The authors of these reviews concede that their conclusions are limited and that further investigation into the issue of prior knowledge is needed (Chen & Kalyuga, 2020; Sinha & Kapur, 2021). Overall, there is ongoing debate about when PS-I and I-PS are most effective, particularly for students with different levels of prior knowledge.

Role of Prior Knowledge

The current literature includes very few studies directly testing the role of prior knowledge in learning from I-PS and PS-I (Chowrira et al., 2019; Zhang & Sweller, 2024). A recent study by Zhang and Sweller (2024) found that I-PS

was more effective for novice learners and less effective for advanced learners among students in a middle school physics class. However, these findings should be interpreted with caution because the study had a sample size of only 47 participants. On the other hand, Chowrira et al. (2019) found that the PS-I approach was more effective than the I-PS approach for low-, medium-, and high-achieving students in a large introductory biology course, with an especially strong effect for low-achieving students. However, the midterm score was used as a proxy for prior knowledge, which may reflect factors that are distinct from prior knowledge (e.g., motivation to study). Finally, Kapur also examined the relationship between PS-I efficacy and prior knowledge, showing that students had similar learning outcomes under PS-I conditions regardless of their level of prior knowledge (Kapur et al., 2023; Toh & Kapur, 2017). Yet, this study only tested the PS-I approach.

In related work, researchers have tested how the complexity of the learning materials interacts with I-PS and PS-I. According to cognitive load theory, the complexity of learning materials depends on learners' level of prior knowledge. That is, the same set of learning materials will be more complex for novice learners and less complex for advanced learners who have more extensive prior knowledge. Some studies have attempted to manipulate the complexity of materials to test whether it may moderate the effects of instructional sequences (Ashman et al., 2020; Chen et al., 2020, 2021). The study by Ashman and colleagues (2020) described above suggested that I-PS may be more effective when learning materials are high in element interactivity (i.e., more complex). However, other studies have found no effect of manipulating lesson complexity. For instance, Chen et al. (2021) compared the I-PS and PS-I approaches for learning materials with low versus high complexity in a college introductory chemistry setting. Their results showed no significant differences between I-PS and PS-I approaches for learning conceptual knowledge for either set of learning materials. Taken together, these findings suggest that the relationship between instructional sequences, lesson complexity, and prior knowledge may not be straightforward (Endres et al., 2023).

Overall, strong conclusions about the role of prior knowledge (and learning-material complexity) for different instructional sequences cannot be drawn based on existing empirical research. In the present study, we aimed to address the many limitations of the current literature by directly investigating the relationship between learners' prior knowledge and the effectiveness of the I-PS and PS-I approaches in a large student sample using the same set of learning materials. We focused on two student populations with different levels of topic-specific knowledge of a fundamental concept in biochemistry: the physical basis of noncovalent interactions.

Topic of Interest: Noncovalent Interactions

Noncovalent interaction is a pivotal concept that falls within the category of structure and function—a core concept across the undergraduate biology

curriculum (American Association for the Advancement of Science, 2011). This concept pertains to various biological topics and is also emphasized in important educational frameworks and studies (Brownell et al., 2014; Loertscher et al., 2014; Tansey et al., 2013). Solving problems about noncovalent interactions requires students to build schemas that include both conceptual knowledge, (i.e., general principle knowledge, symbolic knowledge, category knowledge, and knowledge of principles underlying procedures) (Chen et al., 2021) and procedural knowledge (i.e., knowledge of how to execute solution steps).

Biology students encounter noncovalent interactions repeatedly from high school through the undergraduate curriculum, yet a vast body of literature documents the struggles students experience with noncovalent interaction problems (Becker et al., 2016; Cooper et al., 2015; Halmo et al., 2018; Loertscher et al., 2014, 2018). Students tend to rely on memorized definitions of the types of noncovalent interactions (Loertscher et al., 2018). They generally cannot explain the underlying causal mechanisms by which noncovalent interactions form (Becker et al., 2016; Cooper et al., 2015; Halmo et al., 2018), nor do they tend to consider the nuances of the biomolecular environment that influence these causal mechanisms (Loertscher et al., 2018). As a result, students show limited ability to analyze the noncovalent interactions present in a biological context and to predict how various changes in that context will change the noncovalent interactions (Halmo et al., 2018, 2020). Students' struggles with noncovalent interactions are likely due both to the shortcomings of standard instructional materials (Loertscher et al., 2018) and to the inherent difficulty of learning about phenomena that cannot be directly observed and require representations such as models, pictures, or equations (Cooper & Stowe, 2018; Gabel, 1999; Gilbert & Treagust, 2009; Johnstone, 1991; Taber, 2013). Thus, an important educational challenge for biochemistry educators is to determine how to support student learning of the underlying causal mechanisms of noncovalent interactions and to apply that knowledge to dynamic biological contexts.

The importance and challenges associated with the physical basis of noncovalent interactions serve as an ideal context to investigate the effectiveness of instructional sequences and its relationship with prior knowledge. Educators need targeted instructional materials and evidence-based pedagogies to promote students' application of knowledge for this persistently challenging science concept. This study compares the effect of pedagogical sequences on introductory biology (Experiment 1) and upper-level biochemistry (Experiment 2) students' learning about the physical basis of noncovalent interactions. Specifically, the following research questions guided this study:

1. Does the I-PS or PS-I approach lead to a better understanding of noncovalent interactions?
2. To what extent do the effects of I-PS and PS-I differ for students with lower or higher topic-specific prior knowledge of noncovalent interactions?

The Present Study

This study includes two experiments comparing the effects of I-PS and PS-I on students' understanding of noncovalent interactions. Experiment 1 focused on introductory biology students (who had relatively lower prior knowledge), whereas Experiment 2 targeted upper-level biochemistry students (who had relatively higher prior knowledge). The primary goal was to assess the impact of I-PS and PS-I on learning outcomes across student populations with varying levels of prior knowledge.

We tested two competing hypotheses about the effects of pedagogical sequences on learning. According to the cognitive load hypothesis, I-PS should be more effective than PS-I for students with lower prior knowledge (Experiment 1) and less effective for students with higher prior knowledge[1] (Experiment 2). According to the productive failure hypothesis, PS-I should be more effective than I-PS for both levels of prior knowledge (Experiment 1 and Experiment 2), especially for far-transfer problems that require deeper conceptual understanding.

Experiment 1: Introductory Biology Students

Method

Participants and Design

A priori power analysis using G*Power indicated 128 participants were needed to achieve power of 0.80, assuming a medium effect size (d = .50) and alpha of 0.05. Prior studies comparing the effects of instructional sequences have yielded medium to large effect sizes (Chen & Kalyuga, 2020). We recruited a much larger sample of 367 undergraduates from an introductory biology class at a large southeastern university in the USA. Students were randomly assigned to the I-PS group (n = 183) or PS-I group (n = 184). Participants were expected to have basic topic-specific prior knowledge about the concepts covered in the lesson from high school chemistry and high school biology but limited ability to explain or apply these concepts.

Materials and Measures

The materials consisted of a prior knowledge test, an instructional video (for the instruction phase), a set of two noncovalent-interaction problems (for the problem-solving phase), and a post-test. Each component focused on the causal

[1] As we discuss later in the manuscript, students in Experiment 2 had *relatively* higher topic-specific prior knowledge than those in Experiment 1. However, their performance on the prior knowledge test and post-tests suggests they had not previously acquired schemas for solving the types of problems presented during the learning phase. Thus, cognitive load theory would predict the benefits of I-PS to be stronger in Experiment 1, but I-PS may still be appropriate for students in Experiment 2. We discuss this possibility in greater detail in the "General Discussion" section.

mechanisms underlying noncovalent interactions. The instructional video explained how differences in the charge of various chemical groups lead to the formation of noncovalent interactions, emphasizing the emergence of negative and positive areas on chemical groups and the subsequent attraction between opposite charges. The problems asked students to analyze given noncovalent interactions and predict how changes in the biological context would impact the noncovalent interactions.

Prior Knowledge Test The prior knowledge test assessed students' topic-specific prior knowledge about noncovalent interactions. It included ten multiple-choice questions, seven true–false questions, and one matrix-table question with five items (Appendix 1). These questions assessed participants' knowledge about polarity, electronegativity, dipoles, charge characteristics of amino acids, covalent bonds, and noncovalent interactions, all of which contribute to understanding the physical basis of noncovalent interactions. Each question was worth one point, for a maximum possible score of 22. The Cronbach's α for the prior knowledge test was 0.67, which is in the acceptable range, particularly for measures that assess a range of knowledge (Tavakol & Dennick, 2011).

Instructional Video on Noncovalent Interactions (Instruction Phase) The instruction phase was administered via the introductory biology course's online learning management system. It included a four-part instructional video with four corresponding embedded quizzes to check engagement with the video. The first part of the video (11.5 min) introduced the types of intermolecular forces, electron movement, electronegativity, the types of charges on different chemical groups and what causes these differences, and the mechanisms by which six distinct types of noncovalent interactions arise from different types of charges. The second, third, and fourth parts of the video (6.5 min, 4 min, and 4 min, respectively) presented step-by-step canonical solutions for three different noncovalent-interaction problems and compared these canonical solutions to ideas commonly expressed by students who are learning this material. The canonical solutions consisted of a prediction and an explanation of the causal mechanistic reasons for the prediction. This design of building off students' responses and comparing them with canonical solutions was consistent with guidelines for implementing productive failure (Kapur, 2016). The content of the explicit instruction videos was developed by one of the authors with more than 20 years of experience teaching introductory biology and biochemistry. The content was also verified by one instructor with 7 years of experience teaching biochemistry and a second instructor with 7 years of experience teaching introductory biology. After each part of the video, students completed an embedded quiz, which together consisted of a total of six multiple-choice items.

Noncovalent-Interaction Problems (Problem-Solving Phase) During the problem-solving phase, students were asked to solve a noncovalent-interaction problem during their regular in-person class time. The problem presented students with a drawing representing a cytoplasmic protein and the chemical groups for several amino acids within the protein (Appendix 2). Students were asked to predict the effects of two amino acid mutations on one of the existing noncovalent interactions and

to explain their predictions scientifically. To solve the problem, students needed to identify the existing noncovalent interactions, characterize the original and substituted amino acids, identify what types of charges arise on those chemical groups, and explain how those charges lead to a particular type of noncovalent interaction. Students were prompted to predict any new noncovalent interactions that might occur with such mutations based on the evidence they identified. They also needed to provide a scientific explanation of how and why the new interactions form. The problems were designed to allow students to compare the two different amino acid mutations. The design of generating multiple representations and methods and comparing two situations is consistent with Kapur (2016)'s guidelines for implementing productive failure. Due to the persistent challenge students exhibit with the physical basis of noncovalent interactions, we expected students in the PS-I group to make errors while attempting to solve the problem.

We used an established codebook to analyze participants' problem-solving performance (Halmo et al., 2020). Each code captured a specific idea or piece of scientific reasoning, which we grouped into categories as evidence, claim, and reasoning to assess scientific explanations (McNeill & Krajcik, 2008). Each written response was independently coded by two raters. We calculated intercoder reliability using Cohen's kappa (Gisev et al., 2013). The intercoder reliability ranged from 0.59 to 0.64, reflecting moderate agreement (McHugh, 2012) for the two sub-items. Any disagreement in codes was resolved through discussion to reach a consensus.

Next, we assigned scores to problems based on the collection of codes present in the categories of evidence, claim, and reasoning (Halmo et al., 2020). The evidence category focused on students' categorization of the amino acids provided in the problem, as the explicit and implicit features of these chemical structures serve as the basis for students' claims and reasoning. If students categorized amino acids correctly and assigned the correct magnitude/permanency to the charges in those amino acids, we assigned three points. If they only categorized the amino acids correctly without indicating the magnitude/permanency of the charges, we assigned two points. If they displayed both correct and incorrect ideas about amino acid categorization, we assigned one point. If students provided no statements about amino acid categorization, we assigned zero points. The claim category focused on students' statements about the solution to the problem, i.e., their statements about the existing noncovalent interactions and the new ones that would arise in the given scenario. If students identified the existing interactions correctly and made high-quality predictions about the new interactions, we assigned three points. If students identified the existing interactions correctly and made low-quality predictions about the new interactions (or predicted the new interactions with high quality without identifying the existing interactions), we assigned two points. If students identified the existing interactions correctly without predicting any new interactions (or they made low-quality predictions about the new interactions without identifying the existing interactions), we assigned one point. If they identified the existing interactions correctly but showed a mixture of low and high-predictions on the new interactions, we assigned 1.5/2 points. If they had missing or incorrect ideas about predicting and identifying interactions, we assigned zero points for the claim category. The reasoning category focused on students' explanations about the mechanisms of the existing

or new interactions. If students correctly explained the mechanisms, we assigned two points. If they showed a mixture of correct and incorrect ideas about mechanisms, we assigned one point. If students showed no ideas about the mechanism or their ideas were incorrect, we assigned zero points.

The problem included two sub-problems, and each sub-problem was worth a total of 8 points: 3 points for evidence, 3 points for claim, and 2 points for reasoning. The maximum possible score for the two items was 16. The Cronbach's α for the two items was 0.71, which reflects an acceptable reliability (Tavakol & Dennick, 2011). We also used the codes for each item to assess the number of ideas participants generated during problem solving.

Post-Test The post-test included two sets of problems. The first set of problems (near-transfer items) asked students to solve three noncovalent-interaction problems that resemble the ones used during the instruction and problem-solving phases (Appendix 3). These problems require the same solution structure as the problems in the instruction and problem-solving phases. The solution requires students to predict and explain the impact of two amino acid mutations on the noncovalent interactions within a protein. This solution structure includes identifying the original interactions, characterizing the original and mutated amino acids' charges, predicting potential new interactions from the changes, and explaining the causal mechanism by which the interactions form. The second set of problems (far-transfer items) asked students to solve three noncovalent-interaction problems that do not closely resemble the ones used during the instruction or problem-solving phases but are based on the same underlying principles (Appendix 4). These problems require a different solution structure compared to the near-transfer problems. Students were asked to select and predict which amino acid (out of three) interacts noncovalently with a certain part of a drug. To solve the problem, students needed to characterize the amino acids and the drug, identify what types of charges arise on those chemical groups, and explain how those charges lead to particular types of noncovalent interaction. Students were prompted to choose one amino acid that interacts noncovalently with a drug based on the evidence they identified. They also needed to scientifically explain how and why the chosen amino acid's noncovalent interaction forms.

We used an established codebook to analyze participants' post-test performance (Halmo et al., 2020). As with the noncovalent-interaction problem used in the problem-solving phase, we scored this problem using specific codes for each idea and piece of scientific reasoning and grouped these codes into the categories of evidence, claim, and reasoning. Each written response was independently coded by two raters. We calculated intercoder reliability using Cohen's kappa (Gisev et al., 2013). The intercoder reliability ranged from 0.65 to 0.81, reflecting a moderate to strong agreement (McHugh, 2012) for the post-test items. Any disagreement in coding was resolved through discussion to reach a consensus.

Next, we assigned scores to problems based on the collection of codes present in the categories of evidence, claim, and reasoning (Halmo et al., 2020). The scoring system for near-transfer items is the same as that for problem-solving items. For scoring far-transfer items, we assigned points for the evidence and reasoning categories in the same way as we described for the problem-solving items. However, there was a slight difference for the claim category because students were required

to select one amino acid among three choices and predict the noncovalent interactions that would form given their selection. Each item was worth a total of 8 points (3 points for evidence, 3 points for claim, and 2 points for reasoning). The post-test items included three near-transfer and three far-transfer items, so the maximum possible score for the six transfer items was 48. The Cronbach's α for all the post-test items was 0.80, which reflects high reliability (Tavakol & Dennick, 2011).

Procedure

The experiment took place across six days, including three in-person class sessions (Fig. 1). Participants were randomly assigned to either the I-PS or the PS-I condition. On Day 1, all students were given 15 min to complete the prior knowledge test in class. After the prior knowledge test, the I-PS group was given Day 2 to complete the instruction phase by watching the instructional videos and completing the embedded quiz questions outside of class. Students could watch the video at their own pace (e.g., pausing or re-watching) but could not skip or adjust the play speed.

On Day 3, all students were given 25 min in class to complete the problem-solving phase by attempting to solve the noncovalent-interaction problem individually on their computers.

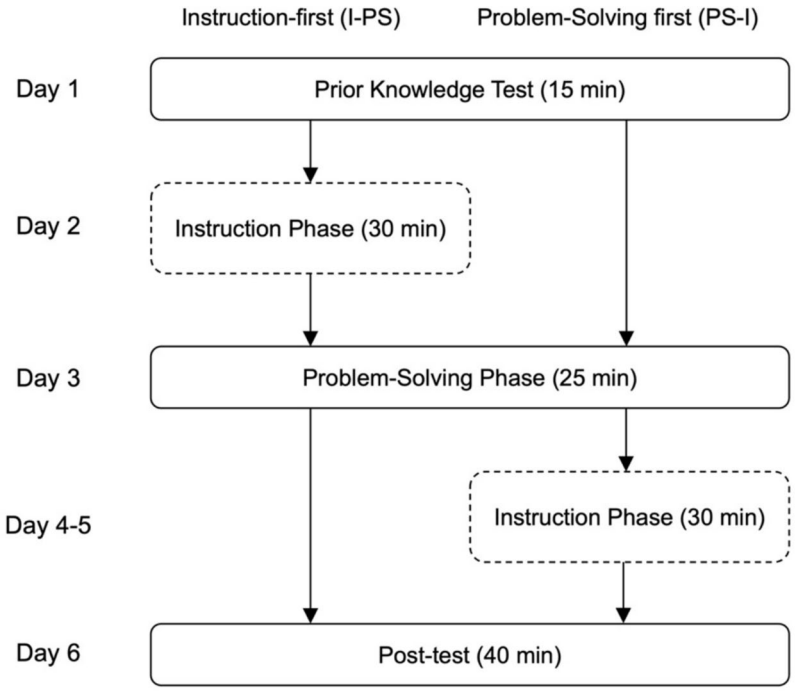


Fig. 1 Procedure for Experiment 1. Note. Boxes with dotted lines represent the instruction phase, which took place asynchronously outside of class either before or after the problem-solving phase. The solid-line boxes represent in-person classroom activities

A graduate assistant (one of the authors) first briefly introduced the problem to students and addressed solution-unrelated queries. Then, students read instructions for the problem on their computer, which varied slightly depending on whether they were assigned to the I-PS or PS-I. I-PS students were prompted to solve the problem based on the instructional video they watched, whereas PS-I students were prompted to think about what knowledge would be useful to solve the problem and do their best to generate possible solutions. Instructions for the PS-I students also emphasized that the goal was not to get a correct answer but to imagine possible solutions and to compare the two sub-problems. This implementation of the PS-I group is consistent with Kapur (2016)'s guidelines.

On Days 4 and 5, PS-I students were given two days to complete the instruction phase by watching the instructional videos and completing the embedded quiz questions outside of class.[2] Finally, on Day 6, all participants were given 40 min to complete the post-test in class.

Results

Preliminary Analyses

Data for Experiment 1 are publicly available via the Open Science Framework: https://osf.io/a7t4b/. First, we tested whether the I-PS and PS-I groups differed in performance on the topic-specific prior knowledge test. The prior knowledge test scores were not normally distributed, so we used the non-parametric Mann–Whitney *U* test. As expected, the results indicated that the I-PS (M = 12.30,SD = 3.52) and PS-I (M = 12.89,SD = 3.74) groups were not significantly different in prior knowledge test scores, U = 15423.5, $p = .163$. The median score for the I-PS group was 12, and for the PS-I group was 13 (out of 22 possible points).

Next, we computed the Pearson correlations among the prior knowledge test and post-test measures. Prior knowledge test performance was positively associated with near-transfer performance, $r(365) = .541$, $p < .001$, and far-transfer performance, $r(365) = .448$, $p < .001$. Therefore, we used the prior knowledge test score as a covariate in the analyses reported below. Near- and far-transfer performance was also strongly correlated, $r(365) = .671$, $p < .001$.

Performance During the Problem-Solving Phase and the Instruction Phase

Due to the non-normal distribution of the data, we conducted a non-parametric Quade ANCOVA to test whether the groups differed in problem-solving performance

[2] PS-I students were given two days to complete the instruction phase because Days 4 and 5 were during the weekend. Given this extra day for the PS-I group, we used a Chi-square test to determine whether the I-PS and PS-I groups differed in the number of times they watched the videos. The results showed the two groups did not significantly differ in the number of video plays across any of the four videos: $X^2(1) = 1.28$, $p = .25$; $X^2(1) = 2.81$, $p = .09$; $X^2 = 0.77$, $p = .29$; $X^2(1) = 3.01$, $p = .08$. Thus, the conditions did not significantly differ in time on task.

and the number of ideas generated during the problem-solving phase, controlling for prior knowledge test score. As expected, the I-PS students (M = 2.72,SD = 3.37) significantly outperformed the PS-I students (M = 0.93,SD = 1.58) during the problem-solving phase, $F(1, 365) = 44.54$, $p < .001$, $\eta_p^2 = .109$. The I-PS group (M = 7.78,SD = 3.73) also generated significantly more ideas than the PS-I group (M = 6.71,SD = 3.02), $F(1, 365) = 11.53$, $p = .001$, $\eta_p^2 = .031$.

Similarly, we conducted a non-parametric Quade ANCOVA to test whether the groups differed in performance on the embedded quizzes during the instruction phase, controlling for the prior knowledge test score. The results revealed that the PS-I group (M = 5.38,SD = 0.91) scored significantly higher on the quizzes than the I-PS group (M = 5.10,SD = 1.04), $F(1, 365) = 6.69$, $p = .010$, $\eta_p^2 = .018$. This suggests that experiencing the problem-solving phase first enhanced learning from the instructional video compared to the I-PS students who had not yet engaged in the problem-solving phase.

Post-Test Performance

Next, we tested whether the groups differed in post-test performance. Due to the non-normal distribution of near-transfer scores, we used non-parametric Quade ANCOVA to compare the two groups' near-transfer performance, controlling for prior knowledge test score. The results indicated that the PS-I students (M = 6.34,SD = 4.21) significantly outperformed the I-PS students (M = 5.07,SD = 3.76) on the near-transfer test, $F(1, 365) = 7.18$, $p = .008$, $\eta_p^2 = .019$. This finding supports the productive failure hypothesis.

Due to the non-normal distribution of far-transfer scores, a non-parametric Quade ANCOVA was conducted to compare the two groups' far-transfer performance, controlling for prior knowledge test score. The results indicated no significant differences between I-PS ($M = 6.40$,SD = 5.07) and PS-I students' (M = 7.95,SD = 6.28) on the far-transfer test, $F(1, 365) = 2.59$, $p = .108$, $\eta_p^2 = .007$. Figure 2 depicts the pattern of results across groups for the near- and far-transfer test.

It is important to note that although student's near-transfer and far-transfer scores are numerically somewhat low, our analytic coding of student responses provides evidence that both groups of students demonstrated an understanding of key scientific ideas from the lesson, including their ability to formulate scientific claims about the nature of noncovalent interactions and provide evidence to support the claims (see Appendix 5).

Discussion

Findings from Experiment 1 generally support the productive failure hypothesis: the PS-I group significantly outperformed the I-PS group on the near-transfer test. This was despite worse performance and fewer ideas generated during the problem-solving phase. According to theories of PS-I, the problem-solving phase may have served to activate relevant formal or informal knowledge, reveal knowledge gaps, and motivate students to

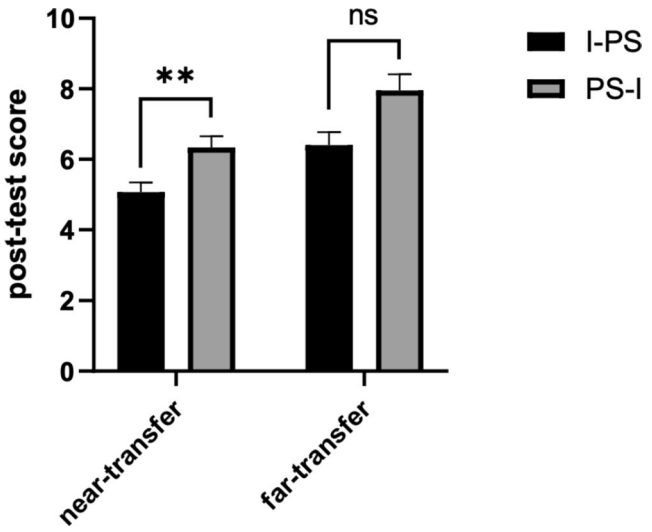


Fig. 2 Comparison of post-test performance between the I-PS and PS-I groups in Experiment 1. Note. Error bars represent the standard error of the mean. $^{**}p < .01$

engage more productively with subsequent explicit instruction, resulting in a better ability to apply their knowledge to similar problems (Kapur, 2016; Schwartz & Martin, 2004). However, we did not find a significant effect on far-transfer problems, which may be in part due to the difficulty level of the problems. We discuss this further in the "General Discussion" section. Importantly, the benefits of PS-I for near transfer were found in students who had low topic-specific prior knowledge of noncovalent interactions. While consistent with productive failure, this is inconsistent with predictions from cognitive load theory, which posits that novice learners should receive explicit instruction prior to problem-solving. In Experiment 2, we tested whether a similar pattern emerged when we compared PS-I and I-PS using the same learning materials but with a sample of biochemistry students who had relatively higher prior knowledge of noncovalent interactions.

Experiment 2: Biochemistry Students

Method

Participants and Design

For Experiment 2, we recruited 138 students from a biochemistry class at a large southeastern university in the USA. Students were randomly assigned to the I-PS group ($n = 64$) or PS-I group ($n = 74$). Participants in this course had previously completed several prerequisite courses, including two semesters of general chemistry, organic chemistry 1, and one semester of introductory biology. Thus, we expected students in Experiment 2 to have more topic-specific prior knowledge of

noncovalent interactions than students in Experiment 1. To confirm, we conducted a nonparametric Mann–Whitney *U* test of prior knowledge test scores between the two student populations. As expected, the biochemistry students in Experiment 2 (Mdn = 16) scored significantly higher on the prior knowledge test than the introductory biology students in Experiment 1 (Mdn = 12), U = 11,904.5, $p < .001$. Thus, students in Experiment 2 had *relatively* higher topic-specific prior knowledge than those in Experiment 1.

Materials and Measures

The materials and measures used in Experiment 2 were identical to those used in Experiment 1. We checked the reliability of the measures and coding for Experiment 2. The Cronbach's α for the prior knowledge test items was 0.69. The intercoder reliability for the problems in the problem-solving phase ranged from 0.67 to 0.73. The Cronbach's α for the two sub-problems in the problem-solving phase was 0.71. The intercoder reliability for the post-test items ranged from 0.70 to 0.83. The Cronbach's α for all the post-test items was 0.84. All the measures and scoring procedures were identical to Experiment 1.

Procedure

The procedure in Experiment 2 was identical to Experiment 1 (see Fig. 1). The only exception was the fourth phase (Days 4–5). In Experiment 2, I-PS and PS-I students were both given one day to complete the instruction phase.

Results

Preliminary Analyses

Data for Experiment 2 are publicly available via the Open Science Framework: https://osf.io/a7t4b/. First, we tested whether the two groups differed in performance on the topic-specific prior knowledge test. The prior knowledge test scores were not normally distributed, so we used a non-parametric Mann–Whitney *U* test. As expected, the results indicated that the I-PS (M = 15.98,SD = 3.06) and PS-I (M = 16.23,SD = 3.37) groups were not significantly different, $U = 2200$, $p = .471$. The median scores for the I-PS and PS-I groups were 16 and 17, respectively.

Next, we computed Pearson correlations among the prior knowledge test and post-test measures. Prior knowledge test performance was positively associated with the near-transfer performance, $r(136) = .591$, $p < .001$. Prior knowledge test performance was also positively associated with the far-transfer performance, $r(136) = .563$, $p < .001$. Therefore, we used the prior knowledge test score as a covariate in the analyses reported below. Near- and far-transfer performance were also strongly correlated, $r(136) = .715$, $p < .001$.

Performance During the Problem-Solving Phase and the Instruction Phase

Due to the non-normal distribution of the data, we conducted a non-parametric Quade ANCOVA to examine whether the groups differed in problem-solving performance and the number of ideas generated during the problem-solving phase, controlling for prior knowledge test score. As expected, the I-PS students (M = 6.64,SD = 4.07) outperformed the PS-I students (M = 2.66,SD = 2.34) in the problem-solving phase, $F(1, 136) = 60.51$, $p < .001$, $\eta_p^2 = .308$. The I-PS group (M = 9.75,SD = 4.17) also generated significantly more ideas than the PS-I group (M = 7.26,SD = 3.13), $F(1, 136) = 20.56$, $p < .001$, $\eta_p^2 = .131$.

Similarly, due to the non-normal distribution of the quiz score, we conducted a Quade ANCOVA to examine whether the groups differed in performance on the embedded quizzes during the instruction phase, controlling for prior knowledge test score. Unlike the results from Experiment 1, the I-PS group (M = 5.77,SD = 0.50) scored significantly higher on the quizzes than the PS-I group (M = 5.40,SD = 0.83), $F(1, 136) = 11.41$, $p = .001$, $\eta_p^2 = .077$. Interestingly, this suggests that the initial problem-solving phase may have interfered with students' learning from the instructional video.

Post-Test Performance

Next, we tested whether the groups differed in post-test performance. Near-transfer performance was normally distributed, so we used ANCOVA to compare the I-PS and PS-I students' near-transfer performance, controlling for prior knowledge test score. The results indicated that unlike Experiment 1, the I-PS students (M = 10.72,SD = 5.34) significantly outperformed the PS-I students (M = 9.47,SD = 4.88) on the near-transfer test, $F(1, 136) = 4.51$, $p = .036$, $\eta_p^2 = .033$. This finding supports the cognitive load hypothesis, which posits that I-PS will be more effective than PS-I for students with limited topic-specific knowledge. Students in Experiment 2—while they had higher topic-specific prior knowledge than students in Experiment 1—still did not have sufficient knowledge of these types of problems, as reflected by their scores on the prior knowledge test and the post-test.

Due to the non-normal distribution of far-transfer scores, we used a non-parametric Quade ANCOVA to compare the I-PS and PS-I groups on far-transfer performance, controlling for prior knowledge test score. The results indicated that there were no significant differences between the I-PS (M = 12.67,SD = 6.37) and PS-I (M = 11.25,SD = 6.63) groups on the far-transfer test, $F(1, 136) = 4.14$, $p = .44$, $\eta_p^2 = .030$. Figure 3 depicts the pattern of results across groups on the near- and far-transfer test.

Like Experiment 1, students' near-transfer and far-transfer scores are numerically somewhat low, but our analytic coding of student responses indicated that both groups of students demonstrated an understanding of key scientific ideas from the lesson (Appendix 6).

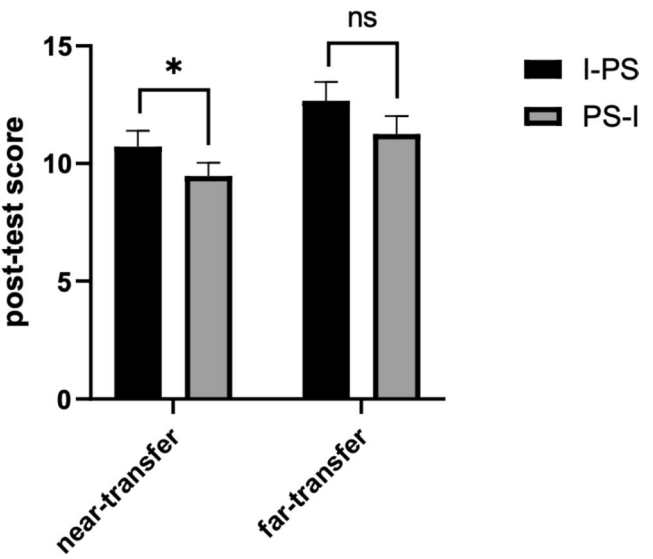


Fig. 3 Comparison of post-test performance between the I-PS and PS-I groups in Experiment 2. Note. Error bars represent the standard error of the mean. $*p < .05$

Discussion

Experiment 2 compared I-PS and PS-I with a sample of biochemistry students who had significantly higher topic-specific prior knowledge of noncovalent interactions than the introductory biology students in Experiment 1. Results indicated that the I-PS group significantly outperformed the PS-I group on the near-transfer test. The I-PS group also generated more ideas during the problem-solving phase than the PS-I group and scored higher on the embedded quiz during the instruction phase. Interestingly, unlike Experiment 1, this pattern of results does not support the productive failure hypothesis and instead supports the cognitive load hypothesis. The productive failure hypothesis posits that PS-I should be more effective than I-PS regardless of students' level of prior knowledge. The cognitive load hypothesis posits that I-PS is more effective than PS-I, particularly when students have low prior knowledge. Although students in Experiment 2 had significantly higher topic-specific prior knowledge than students in Experiment 1, their post-test performance suggests they were still at a novice level with respect to these learning materials. We provide potential explanations for the different patterns of findings across the two experiments in the General Discussion.

General Discussion

The present study advances prior work by comparing I-PS and PS-I for the same topic across two student populations with varying levels of topic-specific prior knowledge using the same learning materials. This study took place within an authentic classroom

setting, included a large sample size (Experiment 1: n= 367; Experiment 2: n = 138), and involved a complex topic in biology. The results showed that for introductory biology students (Experiment 1), the PS-I group significantly outperformed the I-PS on near-transfer problems, whereas for biochemistry students (Experiment 2), the I-PS group significantly outperformed the PS-I group on near-transfer problems. It suggests prior knowledge might play some role in the effects of instructional sequences (at least for near transfer), though not in line with predictions derived from productive failure (Experiment 1) and cognitive load theory (Experiment 2). Importantly, this study cannot directly attribute the differential effects of instructional sequences across the two experiments to differences in topic-specific prior knowledge,[3] as other factors may have also contributed to the divergent findings, which we discuss below.

Findings from Experiment 1 primarily support the productive failure hypothesis: students who engaged in problem-solving prior to instruction outperformed students who received explicit instruction first on near-transfer problems. As predicted by productive failure, this was despite the PS-I group performing worse during problem-solving than the I-PS group. Prior research suggests this effect may result from several factors, including activation of prior knowledge during the initial problem-solving activity (DeCaro & Rittle-Johnson, 2012; Loibl et al., 2017), raised awareness of the knowledge gaps and the problem situation (DeCaro & Rittle-Johnson, 2012; Loibl & Rummel, 2014; Loibl et al., 2017), facilitation of focused attention on searching deeper patterns of the problems (Kapur & Bielaczyc, 2012), an increased desire to learn more about the canonical solutions (Lamnina & Chase, 2019), and increased learner agency and engagement to learn the target concept (Clifford, 1984; Hiebert & Grouws, 2007). This finding does not support the cognitive load hypothesis, which posits that PS-I would deter learning for low-knowledge learners by creating excessive cognitive load.

In contrast, Experiment 2 indicated that biochemistry students benefited more on near-transfer problems from receiving explicit instruction first (I-PS) rather than problem-solving first (PS-I). This finding does not support the productive failure hypothesis, which posits that PS-I should be effective for students with higher prior knowledge as well as lower prior knowledge. At first glance, this finding may also seem to go against the cognitive load hypothesis, which posits that I-PS should be less effective for students with high prior knowledge. However, performance in Experiment 2 suggests that while students did have significantly *higher* topic-specific prior knowledge than students in Experiment 1, they did not show evidence of having previously acquired schemas for solving the types of problems presented in the learning materials (Sweller et al., 2011). This suggests that students in Experiment 2 still experienced the materials as highly complex rather than redundant to what they already knew. Therefore, the findings of Experiment 2 can be interpreted as supporting cognitive load theory. Yet the contradictory findings from Experiment

[3] As supplementary analyses, we also tested whether prior knowledge score moderated the effects of instructional sequences for each experiment (see Supplementary Information for full details). The results indicated that prior knowledge was not a significant moderator in either experiment. It is important to note that this analysis is limited because the students in each experiment had a restricted range of prior knowledge scores.

1 (which involved even lower-knowledge students) suggest the relationship between prior knowledge and instructional sequences may not be straightforward. As we discuss below, other factors may be involved, including possible differences in students' expectations, beliefs, or motivations toward the PS-I approach.

A possible explanation is that the different backgrounds of the two student populations influenced how students approached the learning material, particularly for the PS-I sequence. For example, the PS-I students in Experiment 2 may have assumed they knew how to solve the problems, given their relatively higher familiarity with the content. Biochemistry students had taken pre-requisite courses covering the concepts and supporting ideas in the lesson used for this study. However, research shows that even students who complete advanced chemistry and biochemistry courses struggle to use causal mechanistic reasoning to solve complex problems about non-covalent interactions (Becker et al., 2016; Halmo et al., 2018), relying instead on heuristics and superficial ideas that sometimes but not always lead to correct solutions (Becker et al., 2016; Halmo et al., 2018). Consequently, it is plausible that biochemistry students approached the problem-solving phase in a somewhat superficial manner, using heuristics without fully recognizing the causal, mechanistic features of the problems or identifying gaps in their own knowledge. Thus, the PS-I approach may have hindered their ability to benefit from subsequent explicit instruction. Our findings from the embedded quizzes students took during the instruction phase align with this hypothesis: In Experiment 2, PS-I students performed significantly worse on the embedded quizzes than the I-PS group despite having an opportunity to engage in problem solving before watching the video (whereas I-PS students had only watched the video before taking the quiz). This suggests the initial problem-solving phase may have interfered with biochemistry students' learning from the instructional video.

On the other hand, the PS-I students in Experiment 1, who possessed relatively lower prior knowledge and had more limited prior exposure to the core concepts and supporting ideas, may have approached the initial problem-solving phase with a greater sense of openness and willingness to explore potential problem-solving solutions. This openness could have facilitated a broader activation of their prior knowledge and awareness of their knowledge gaps. As a result, the PS-I students in Experiment 1 may have been better prepared to leverage subsequent explicit instruction in the PS-I approach, as reflected by their superior near-transfer test performance. Along these same lines, Toh and Kapur (2017) found that providing prerequisite knowledge to students in a PS-I approach did not yield higher learning gains and reported lower lesson engagement and greater mental effort during the subsequent instruction compared to those without the provision of prerequisite knowledge. Taken together, students' levels of prior knowledge and their corresponding approaches to problem-solving activities may help explain the unexpected pattern of results in this study. However, we need further empirical work to directly assess how prior knowledge interacts with students' expectations or beliefs about problem-solving activities during PS-I.

In summary, our findings suggest there may not be a straightforward relationship between students' level of prior knowledge and the optimal instructional sequence. Students with very low topic-specific prior knowledge benefited from PS-I, perhaps

due to the opportunity to activate intuitive knowledge and become aware of knowledge gaps without expecting to solve the problem successfully. Students with relatively higher topic-specific prior knowledge, but still no expertise in the domain, benefited more from I-PS, perhaps because they had not yet acquired schemas for solving the types of problems presented during the instruction phase and/or because they may have approached PS-I with stronger expectations of solving the problem successfully. Of course, these explanations are speculative, but they provide promising directions for future research.

As for far-transfer performance across the two experiments, the results had the same trend as near-transfer performance across the two experiments, but the difference was not statistically significant between the I-PS and PS-I groups. This lack of effect on far-transfer measures in both experiments does not support predictions from productive failure theory. Far-transfer problems involve the application of knowledge and skills to contexts that are substantially different from those presented in the instructional materials. Although the productive failure approach predicts that PS-I should yield a deeper understanding and prepare students for further transfer of learning, we did not observe this in our studies. Far transfer is notoriously difficult to achieve, particularly in highly complex domains like biochemistry, for which students of all backgrounds tend to enter with surface-level or heuristic-based knowledge. The problem-solving activities and instructional video may have been insufficient in duration or capacity to foster deep knowledge changes that support transfer to substantially different problems. Future research within complex domains like biochemistry is needed to determine how I-PS and PS-I approaches need to be adapted or expanded to support far transfer.

Practical Contribution

This study provides practical implications for implementing I-PS and PS-I sequences within an undergraduate biology context. Our results show that instructional sequences may have different effects for students with different backgrounds, including (but not limited to) differences in topic-specific knowledge. Our findings from Experiment 1 suggest that biology instructors should be open to implementing PS-I approaches in introductory courses, such as incorporating preparatory activities prior to explicit instruction to prepare students for subsequent explicit instruction. Our findings in Experiment 2 suggest that instructors might emphasize I-PS approaches in advanced-level courses for which students possess more prior knowledge but have not yet mastered the content. Having students receive explicit instruction and a worked example before solving a complex problem on their own may benefit their near-transfer learning.

Of course, instructors should interpret these findings and their implications within the appropriate context. First, the effect sizes in these two experiments were relatively small and limited to near-transfer performance. Thus, instructors might choose to focus more on *how* they implement explicit instruction and problem-solving (e.g.,

the quality of their examples, explanations, and feedback) rather than the sequence of these activities per se. Additionally, the present study implemented the instruction phase via asynchronous out-of-class instructional videos rather than typical live, in-class instruction. This study also focused on a single topic at the undergraduate level. As such, only cautious and conservative recommendations can be made regarding when instructors should choose I-PS or PS-I. One possibility is that a mix of PS-I and I-PS approaches may be most appropriate to accommodate a wide range of learner backgrounds by leveraging the strengths of both instructional sequences. For example, in a PS-I-PS sequence, the initial problem-solving phase engages learners by activating their existing knowledge, the instruction phase provides structured support and introduces problem-solving schemas, and the subsequent problem-solving phase allows students to apply and practice the learned schemas. Alternating these phases allows all students to have opportunities for both active problem-solving and guided knowledge integration. More empirical research is needed to provide practical contributions regarding using instructional sequences in varied educational contexts.

Limitations and Future Research Directions

One limitation of this study was the method used to categorize students into low- and high-prior knowledge groups. We relied on the students' course levels (introductory vs. advanced) and a prior knowledge test to gauge their existing understanding of noncovalent interactions. While the prior knowledge test scores confirmed that the biochemistry students had relatively higher topic-specific prior knowledge than the introductory biology students, there is still overlap in the distributions of prior knowledge test scores across courses. Thus, course level is a useful but imperfect indicator of students' prior knowledge. Accordingly, being enrolled in an upper-level biochemistry course or having completed prerequisite biology courses does not necessarily equate to a deep or adequate understanding of noncovalent interactions, a difficult concept for many undergraduates (Cooper et al., 2015; Loertscher et al., 2014). Biochemistry students had relatively higher topic-specific prior knowledge, but their understanding was still limited in the context of the causal mechanistic reasoning required by our learning materials. Moreover, this study only focused on topic-specific knowledge without measuring general domain knowledge. Future research should systematically examine the impact of a fuller range of levels of prior knowledge (e.g., general domain knowledge and topic-specific knowledge) in moderating the impact of different instructional sequences, particularly in highly complex domains in which even relatively advanced students may still have only a superficial understanding.

Another limitation of our study is that students watched the instructional videos outside of class. With this design, we could not control the precise time and conditions under which students completed the instruction

phase. Furthermore, in Experiment 1, the PS-I group was afforded an extra day to watch the video because the timing of the study included a weekend. We implemented the intervention this way to accommodate the needs of the instructors teaching the courses and because it is common for students to be assigned lecture videos at home (e.g., flipped classrooms). Importantly, our data indicate that the two groups in Experiment 1 did not significantly differ in the number of times they watched the instructional videos. Nonetheless, it is still possible that having an extra day for the instruction phase might have provided some advantages to the PS-I group, such as having more control over when they chose to watch the instructional video. Conversely, completing academic tasks over the weekend can also create unique challenges for students compared to the typical school week. Future research should replicate these findings by implementing a more controlled instruction phase.

On a related note, conducting the instructional phase outside of class may have resulted in lower overall student engagement. While student engagement during the instructional phase was assessed using an embedded quiz, it remains unclear whether students were as engaged with the instructional videos as they might have been with live, in-class instruction. This may have influenced the results. For instance, students who received instruction first had worse quiz scores in Experiment 1 but better in Experiment 2. This raises the question of whether introductory-level students are less engaged with initial instructional videos assigned for outside-class viewing than their more advanced peers. Thus, the observed results may reflect the combined effects of instructional sequencing and the flipped classroom design (instruction delivered outside of class) rather than instructional sequencing alone. Future studies should systematically investigate whether the setting of the instruction and/or problem-solving phase moderates any effect of instructional sequences.

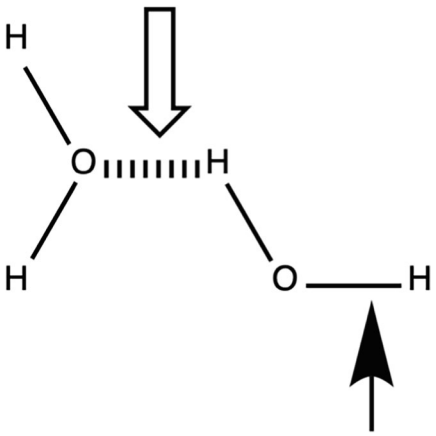
Moreover, a possible criticism of our problem-solving phase is that students worked individually rather than collaboratively. We implemented it this way to maintain greater experimental control over the problem-solving phase, particularly given that we had less control over the instruction phase, as described above. Some have argued that collaboration is important for the effectiveness of PS-I, such as by increasing engagement, providing students the opportunity to cue each other's prior knowledge, and building on the complementary knowledge of others (Nokes-Malach et al., 2015; Sinha & Kapur, 2021). However, prior research involving similar contexts and methodologies has not shown added benefits of collaboration (e.g., Brand et al., 2023; Weaver et al., 2018). Further work is needed to specify how instructors can leverage the unique benefits of collaboration in PS-I and I-PS while managing the potential costs.

Finally, future research on instructional sequences should broaden the scope beyond a single concept to a broader range of fundamental topics in biochemistry. While our study provides insights into the concept of noncovalent

interactions, it represents just one among several foundational ideas essential to biochemistry education (Loertscher et al., 2014). It is crucial to explore instructional sequences across various concepts due to the differences in complexity and student familiarity, which may influence learning outcomes. For instance, metabolic pathway dynamics and regulation is another foundational concept in biochemistry (American Association for the Advancement of Science, 2011; Brownell et al., 2014; Loertscher et al., 2014). Solving problems about metabolic pathway dynamics and regulation requires the interpretation of complex visual representations (Bhatia et al., 2022; Offerdahl et al., 2017; Wright et al., 2017) and the integration of challenging concepts that students initially encounter in introductory chemistry and biology courses (Bhatia et al., 2022; Villafañe et al., 2021). Moreover, it would be constructive for future research to consider longer intervention periods (e.g., Chowrira et al., 2019). The current study was limited to a brief timeframe that may not capture the full effect a particular instructional sequence could have over a longer duration. Further, given the low performance of both introductory biology and biochemistry students on near and far-transfer tests, students may have needed additional explicit instruction and/or problem-solving activities to achieve mastery of the knowledge required. Longitudinal studies extending over a full course would provide a richer perspective on the sustained influence of the PS-I versus I-PS approaches on student learning and retention.

Appendix 1. Prior Knowledge Test

Q1 The picture below shows two water molecules. Respond to each true/false statement about this picture.



	Select True or False	
	True	False
The open arrow points to a noncovalent interaction.	○	○
The open arrow represents a pair of shared electrons.	○	○
The solid arrow points to covalent bond.	○	○
The solid arrow represents an attractive force between areas of high and low electron density.	○	○

Q2 Characterize each of the following statements as true or false:

	True	False
Electronegativity of an atom is the ability of the atom to repel electrons.	○	○
The greater the difference in electronegativity between two bonded atoms, the stronger the bond will be.	○	○
Electronegativity of an atom depends on the size of the atom.	○	○

Q3 Which of the following atoms has the highest electronegativity? (with a Periodic Table of Element shown).

○ Nitrogen

○ Oxygen

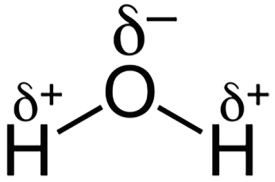
○ Hydrogen

○ Carbon

Q4 Which of the following best describes the term dipole?

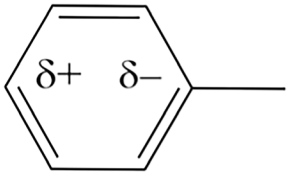
- A dipole occurs when electrons in a covalent bond share their electrons unequally.
- A dipole occurs when electrons in a covalent bond share electrons equally.
- A dipole occurs when both atoms in a covalent bond have partial positive charges.
- A dipole occurs when both atoms in a covalent bond have partial negative charges.

Q5 The molecule below shows a snapshot of charges on atoms at a given moment in time. What is the most important cause of the charges indicated on the molecule?



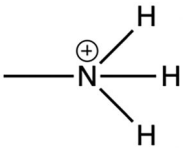
- O and H have different electronegativity.
- O and H have gained or lost an electron.
- Electrons are moving continuously around a chemical group.

Q6 The molecule below shows a snapshot of charges on atoms at a given moment in time. What is the most important cause of the charges indicated on the molecule?



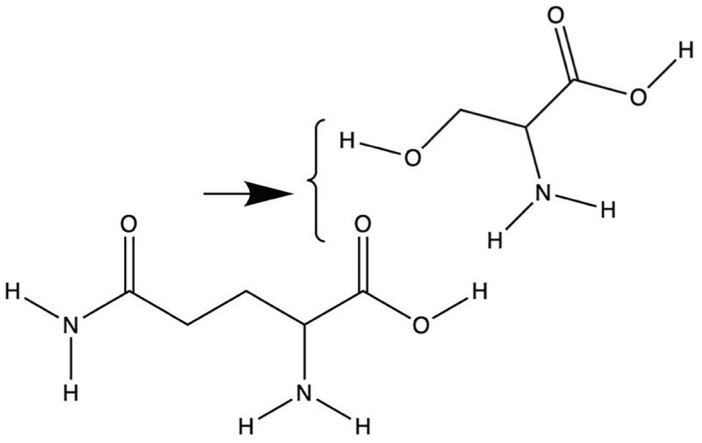
- C and H have different electronegativity.
- C and H have gained or lost an electron.
- Electrons are moving continuously around a chemical group.

Q7 The molecule below shows a snapshot of charges on atoms at a given moment in time. What is the most important cause of the charges indicated on the molecule?



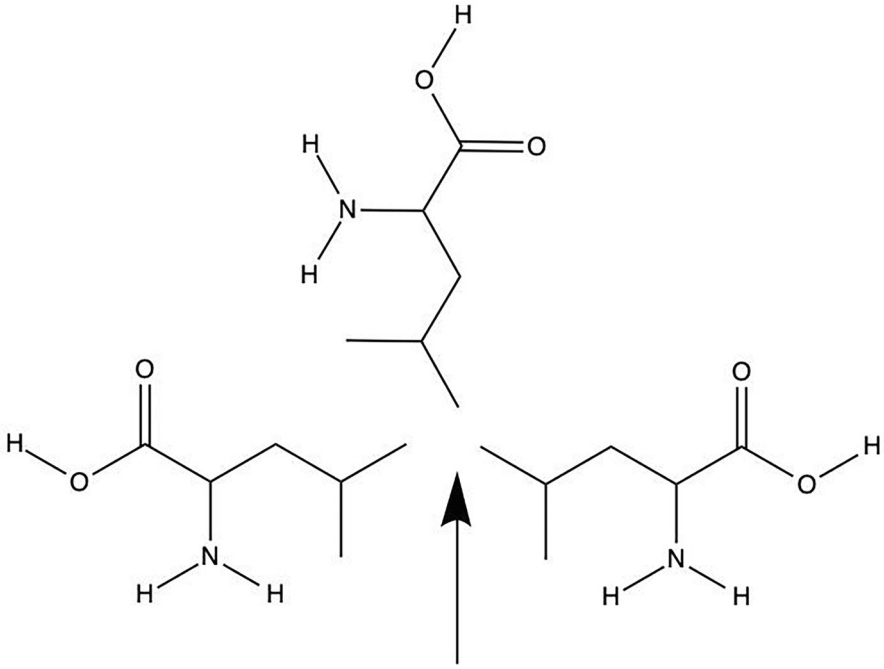
- N and H have different electronegativity.
- N and/or H has gained or lost an electron.
- Electrons are moving continuously around a chemical group.

Q8 The item below corresponds to the most prominent noncovalent interaction present in the space indicated by the arrow. What is the name of this noncovalent interaction? Select one option.



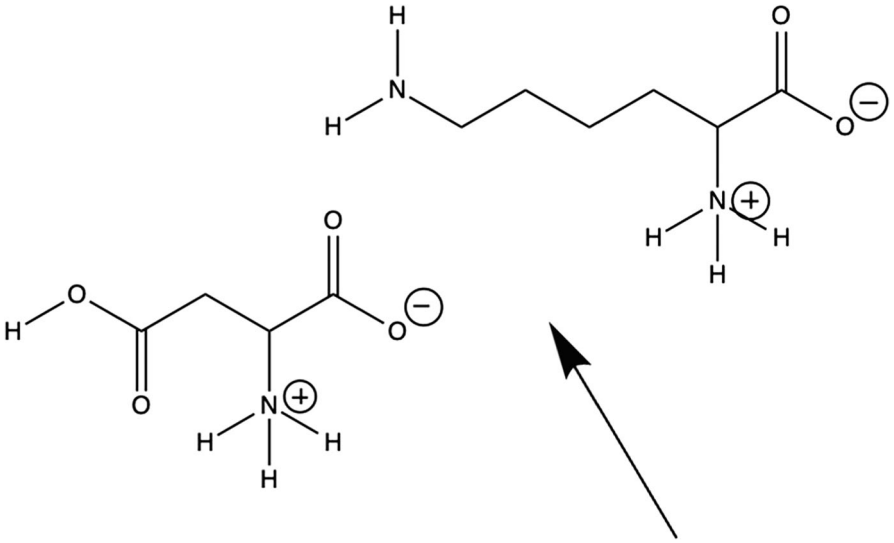
- Hydrogen bond (a special type of dipole-dipole interaction)
- Ion pairing (also called an ion-ion interaction, charge-charge interactions, or salt bridges)
- Van der Waals Interaction (also called London Dispersion Forces or an induced dipole-induced dipole interaction)

Q9 The item below corresponds to the most prominent noncovalent interaction present in the space indicated by the arrow. What is the name of this noncovalent interaction? Select one option.



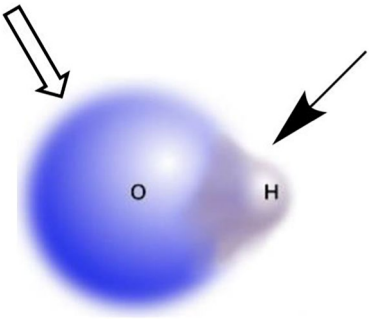
- Hydrogen bond (a special type of dipole-dipole interaction)
- Ion pairing (also called an ion-ion interaction, charge-charge interactions, or salt bridges)
- Van der Waals Interaction (also called London Dispersion Forces or an induced dipole-induced dipole interaction)

Q10 The item below corresponds to the most prominent noncovalent interaction present in the space indicated by the arrow. What is the name of this noncovalent interaction? Select one option.



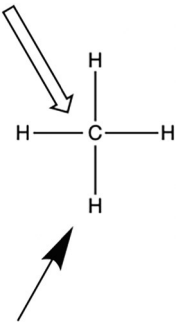
- Hydrogen bond (a special type of dipole-dipole interaction)
- Ion pairing (also called an ion-ion interaction, charge-charge interactions, or salt bridges)
- Van der Waals Interaction (also called London Dispersion Forces or an induced dipole-induced dipole interaction)

Q11 The figure below represents an OH group. Which region has the highest electron density?



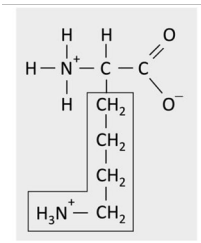
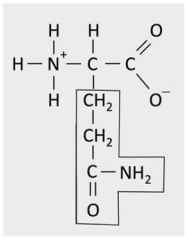
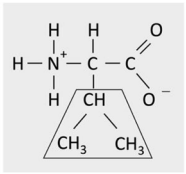
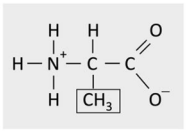
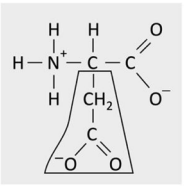
- ○ The region indicated by the open arrow.
- ○ The region indicated by the solid arrow.
- ○ The two regions have equal electron density.

Q12 The figure below represents a CH group. Which region has the highest electron density?



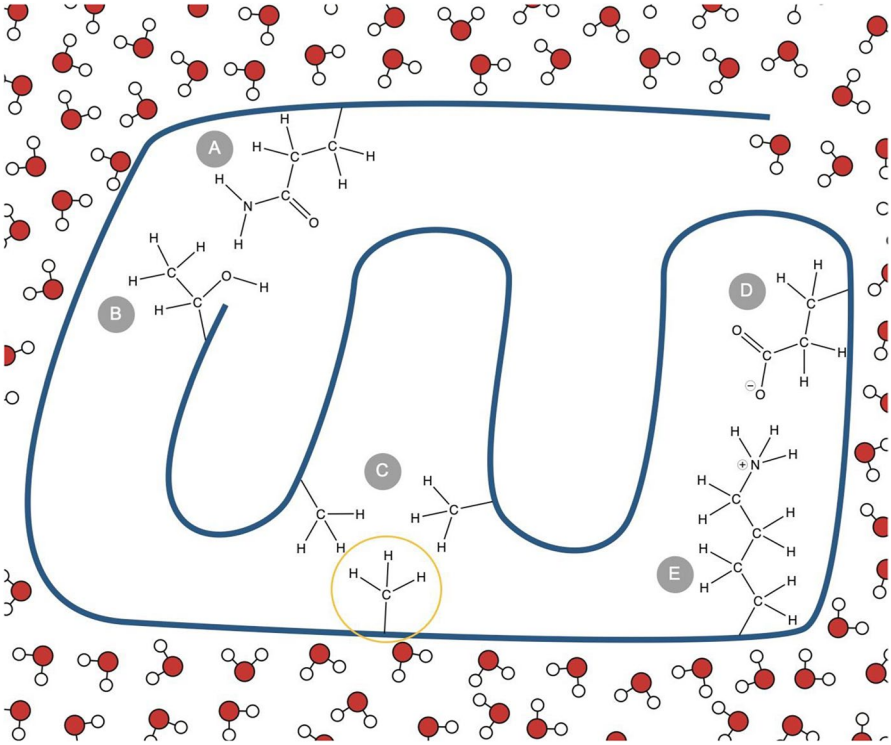
- ○ The region indicated by the open arrow
- ○ The region indicated by the solid arrow
- ○ The two regions have equal electron density

Q13 The molecules below are amino acids. The chemical groups with a box around them are unique for every amino acid. The chemical groups without a box are common to all amino acids. Classify the chemical group in each box as polar, nonpolar, or polar charged.

	Polar	Nonpolar	Polar Charged
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○

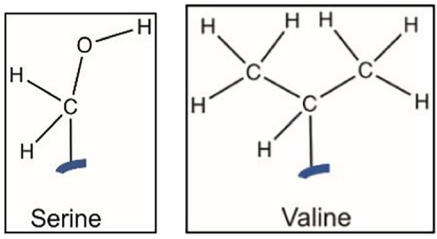
Appendix 2. Noncovalent-Interaction Problems in Problem-Solving Phase (Based on Problems Published by Halmo et al. (2020))

Protein W, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and white). The environment has a pH of 7.4. The blue line represents the Protein W backbone; some but not all of the amino acid side chains (R groups) are shown.



The amino acids shown are: (A) glutamine, (B) threonine, (C) alanine, (D) glutamate, and (E) lysine. Sometimes, a mutation occurs that substitutes alanine (yellow circle) with serine or valine (below).

Sometimes, a mutation occurs that substitutes alanine (yellow circle) with serine or valine (below).

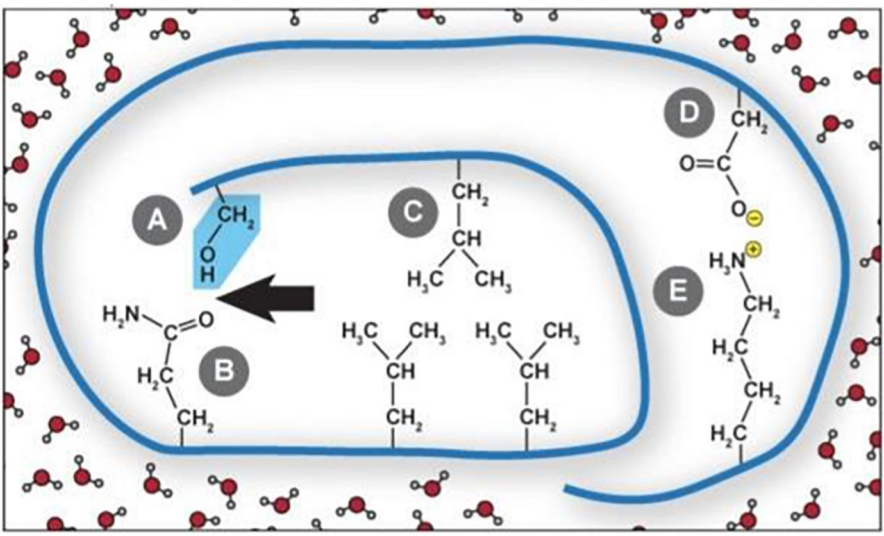


Predict the effect of **Serine substitution** on the existing noncovalent interactions. Compare and contrast the effects of the two amino acid substitutions. Explain your reasoning.

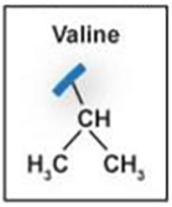
Predict the effect of **Valine substitution** on the existing noncovalent interactions. Compare and contrast the effects of the two amino acid substitutions. Explain your reasoning.

Appendix 3. Noncovalent-Interaction Near-Transfer Problems in Posttest (Based on Problems Published by Halmo et al. (2020))

Protein X, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and gray). This environment has a pH of 7.4. The blue line represents the protein X backbone. Some, but not all, of the amino acid side chains are shown in the chemical notation.

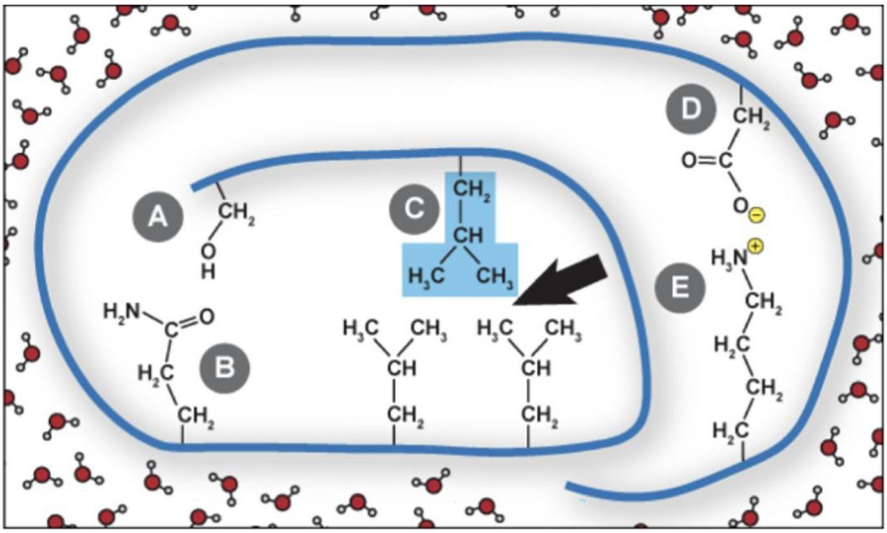


The amino acids shown are: (A) serine, (B) glutamine, (C) leucine, (D) aspartate, and (E) lysine. Sometimes, a mutation occurs that substitutes serine (blue highlight) with valine (below).

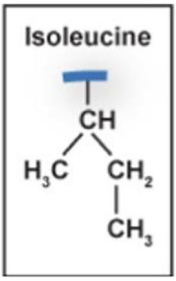


Do you predict that such a mutation would affect the non-covalent interaction pointed to by the arrow? Explain your reasoning, including a description of the original interaction and any new interactions.

Protein X, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and gray). This environment has a pH of 7.4. The blue line represents the protein X backbone. Some, but not all, of the amino acid side chains are shown in the chemical notation.

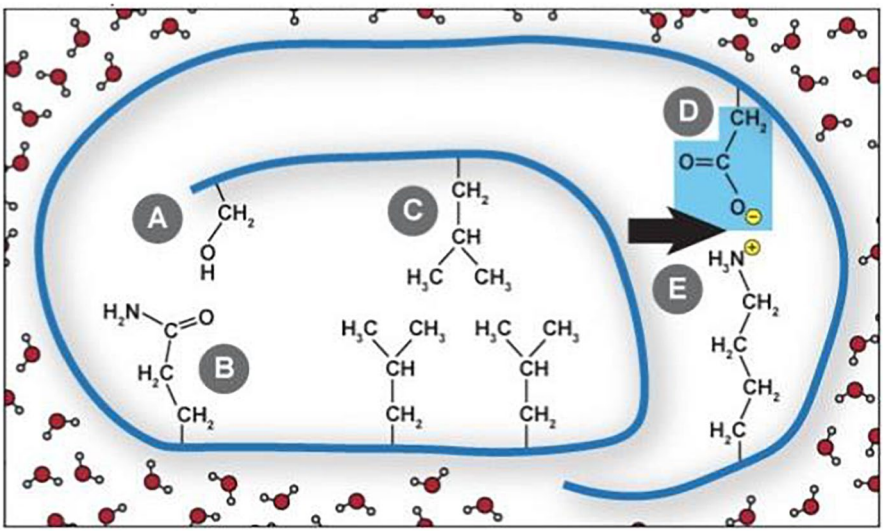


The amino acids shown are: (A) serine, (B) glutamine, (C) leucine, (D) aspartate, and (E) lysine. Sometimes, a mutation occurs that substitutes leucine (blue highlight) with isoleucine (below).



Do you predict that such a mutation would affect the non-covalent interaction pointed to by the arrow? Explain your reasoning, including a description of the original interaction and any new interaction.

Protein X, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and gray). This environment has a pH of 7.4. The blue line represents the protein X backbone. Some, but not all, of the amino acid side chains are shown in the chemical notation.



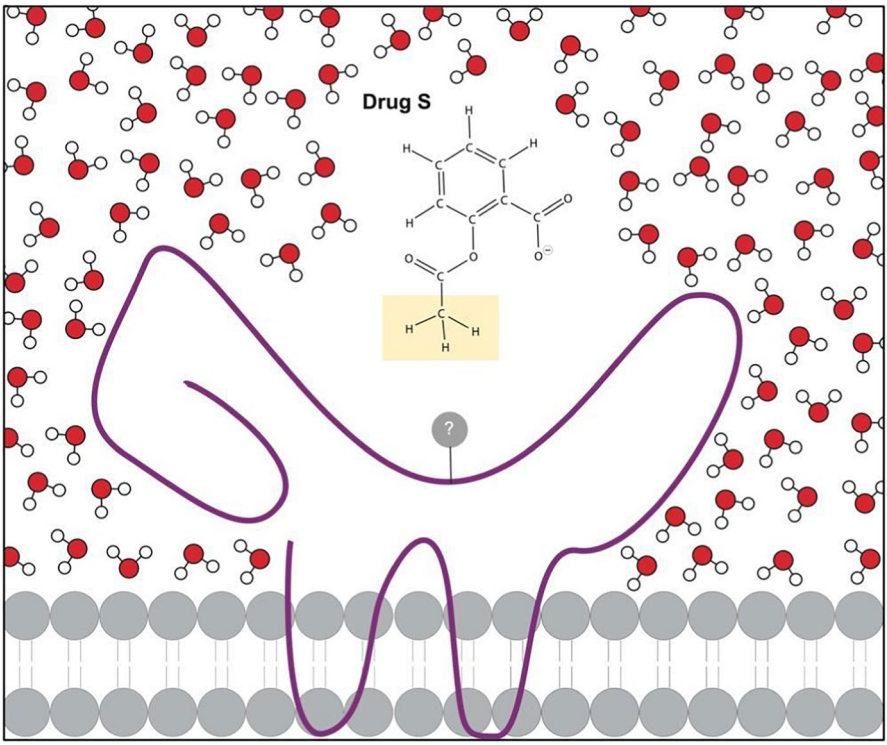
The amino acids shown are: (A) serine, (B) glutamine, (C) leucine, (D) aspartate, and (E) lysine. Sometimes, a mutation occurs that substitutes aspartate (blue highlight) with alanine (below).

Do you predict that such a mutation would affect the non-covalent interaction pointed to by the arrow? Explain your reasoning, including a description of the original interaction and any new interaction.

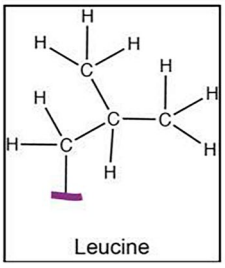
Appendix 4. Noncovalent-Interaction Far-transfer Problems in Posttest (Based on Problems Published by Halmo et al. (2020))

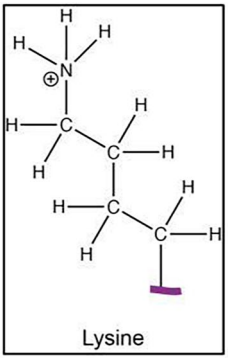
Below is a model of Drug S and a protein with which it may interact. The protein is located on the cell surface situated within the cell membrane and surrounded by water molecules (red and white). The environment has a pH of 7.4. The purple line represents the protein backbone, and the section labeled with a question mark is a site for an amino acid side chain (R group).

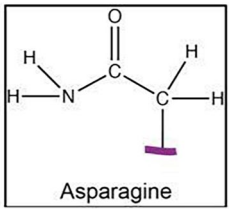
Model of Drug S and Protein:



Amino Acid Side Chains:



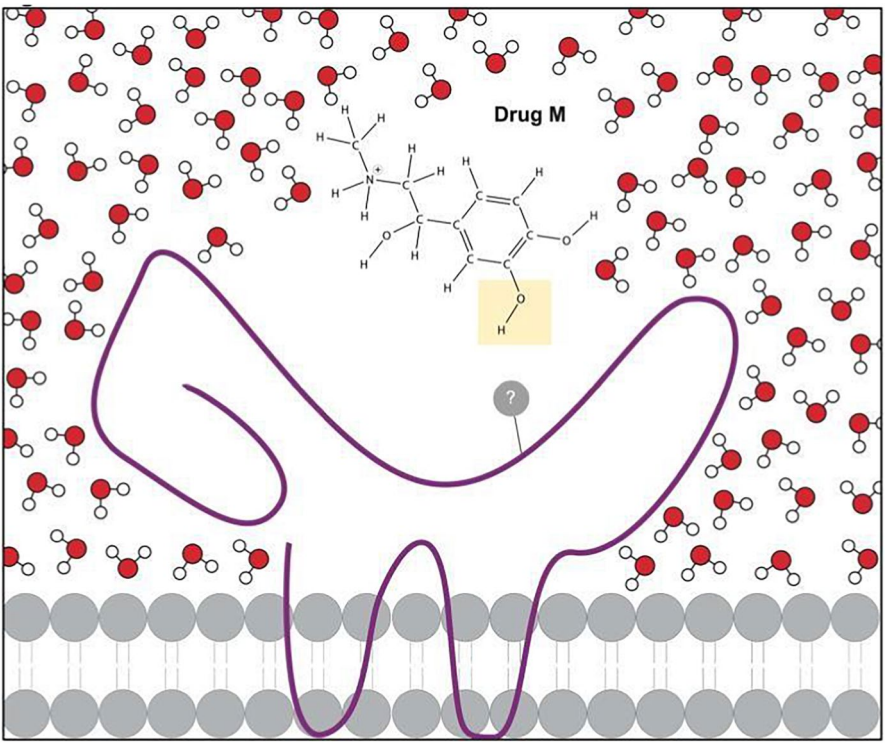




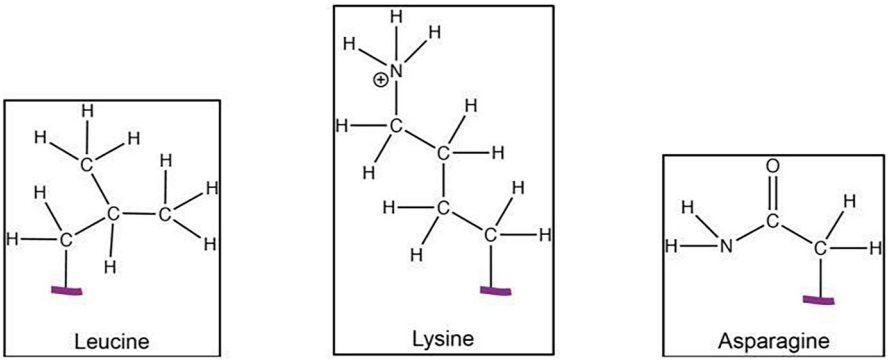
Which amino acid would interact noncovalently with the yellow highlighted section of Drug S? Provide a scientific explanation describing how Drug S interacts noncovalently with the amino acid you selected. Be sure to describe how this interaction forms.

Below is a model of Drug M and a protein with which it may interact. The protein is located on the cell surface situated within the cell membrane and surrounded by water molecules (red and white). The environment has a pH of 7.4. The purple line represents the protein backbone, and the section labeled with a question mark is a site for an amino acid side chain (R group).

Model of Drug M and Protein:



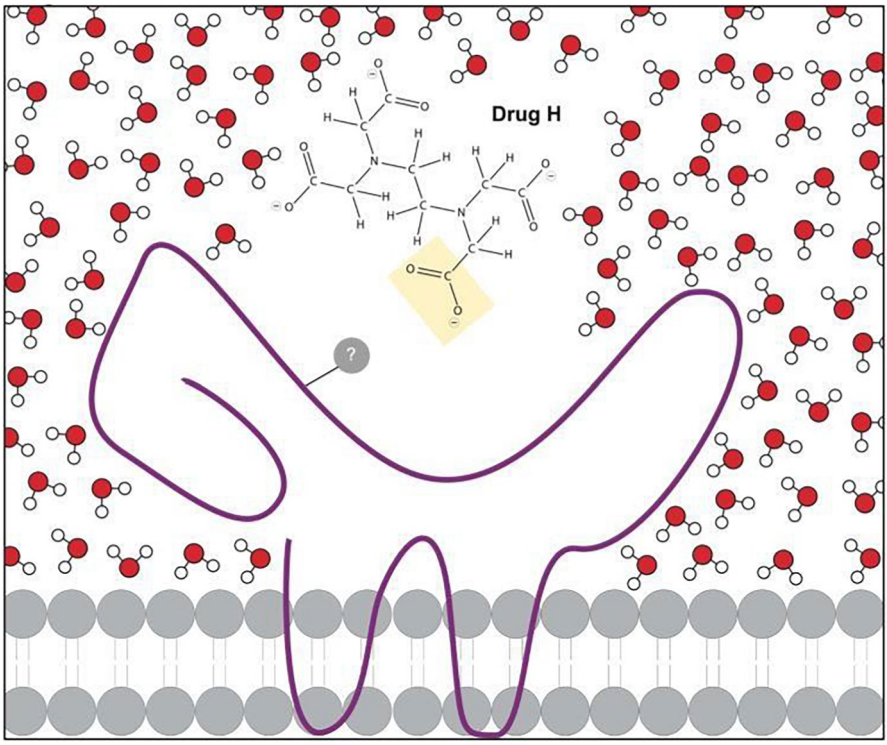
Amino Acid Side Chains:



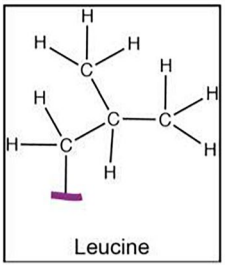
Which amino acid would interact noncovalently with the yellow highlighted section of Drug M? Provide a scientific explanation describing how Drug M interacts noncovalently with the amino acid you selected. Be sure to describe how this interaction forms.

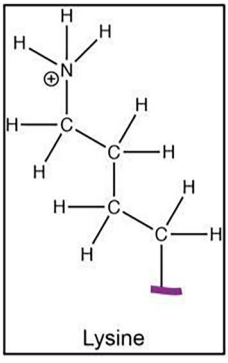
Below is a model of Drug H and a protein with which it may interact. The protein is located on the cell surface situated within the cell membrane and surrounded by water molecules (red and white). The environment has a pH of 7.4. The purple line represents the protein backbone, and the labeled Sects. (1) and (2) are two sites for amino acid side chains (R groups).

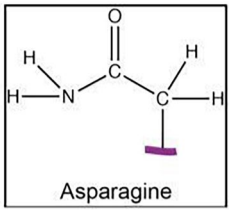
Model of Drug H and protein:



Amino Acid Side Chain:







Which amino acid would interact noncovalently with the yellow highlighted section of Drug H? Provide a scientific explanation describing how Drug H interacts noncovalently with the amino acid you selected. Be sure to describe how this interaction forms.

Appendix 5. Example Responses in Experiment 1

A. Example responses to the first near-transfer problem

I predict that this mutation would affect the non-covalent interaction. This would cause a change because there is a significant amount more hydrogen in the valine than the serine. The serine and glutamine form dipole-dipole bonding, and the non-polar valine bonding with glutamine would decrease the strength of the bond. (from an *I-PS student, 1 pt evidence, 1 pt claim*)

A hydrogen bond intermolecular force (IMF) occurs with Serine and Glutamine. However, Valine has nonpolar covalent bonds. Because Valine has a weaker intermolecular force than Serine, this means that the IMF won't be as tightly held. Therefore, the shape changes, which changes the function. (from a *PS-I student, 1 pt evidence, 1 pt claim*)

B. Example responses to the first far-transfer problem.

The amino acid that would interact non-covalently is leucine. The amino acids would interact between a CH3 on both sides. This would create a nonpolar bond meaning that the only intermolecular force present is London dispersion forces thus non-covalent. (from an *I-PS student, 3 pt claim*)

Leucine would provide the best non covalent interaction as there are no specific charges associated with the molecule. Leucine also provides three additional hydrogen atoms whereas the two molecules can experience a van der Waals force (from a *PS-I student, 3 pt claim*).

Appendix 6. Example Responses in Experiment 2

A. Example responses to the first near-transfer problem.

The original interaction involves the intermolecular attraction of hydrogen bonding between the H attached to the electronegative O of serine and the O of the carbonyl group of glutamine. However, if a mutation replaced serine with nonpolar valine, there would be a dipole-induced-dipole interaction between groups A and B instead. This is because group A is now nonpolar and does not have any F,O, or N to use as a hydrogen acceptor for hydrogen bonding. (from an *I-PS student, 1 pt evidence, 2 pt claim*)

Yes, as of now, serine is able to interact with glutamine by hydrogen bonding due to the electronegativity difference between hydrogen and oxygen. However, when serine is replace with valine, a non-polar amino acid, the non-covalent interaction will be affected as hydrogen bonding would no longer be able to take place. The new non-covalent interaction will most likely be a permanent dipole to induce dipole interaction. (from a *PS-I student, 1 pt evidence, 2 pt claim*)

B. Example responses to the first far-transfer problem

Leucine would interact non-covalently with the drug because it is the only option with a non polar side chain. This means London dispersion interactions could most effectively occur between the drug and the side chain as opposed to a polar or charged side chain. (from an *I-PS student, 1 pt evidence, 3 pt claim*)

The methyl side chain of Drug S (highlighted) can only interact via London Dispersion forces. There is no permanent or partial charge associated with this side chain because carbon and hydrogen have very similar electronegativities. Therefore, leucine, lysine, and asparagine will all interact with this side chain, but leucine will probably interact best because it can also only interact via London Dispersion forces. Lysine and asparagine are more likely to interact with a different side chain of the drug or the surrounding water molecules rather than the highlighted portion. (from a *PS-I student, 1 pt evidence, 3 pt claim*)

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10648-025-09993-3.

Acknowledgements We would like to acknowledge the Biology Education Research Group at the University of Georgia for their feedback on the manuscript and Brooke Coates, Kyle Nuckles, and Seeta Patel, undergraduates who assisted with this work.

Funding This work was supported through a grant from the National Science Foundation, Grant No. DUE-2055117 awarded to Logan Fiorella and Paula P. Lemons. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation.

Data Availability Data for this study are publicly available via the Open Science Framework: https://osf.io/a7t4b.

Declarations

Ethics Approval The approval for this human subject research was obtained from the Institutional Review Board of the author's institution.

Conflict of Interest The authors declare no competing interests.

References

American Association for the Advancement of Science. (2011). Vision and change in undergraduate biology education: A call to action. Washington, DC.

Ashman, G., Kalyuga, S., & Sweller, J. (2020). Problem-solving or explicit instruction: Which should go first when element interactivity is high? *Educational Psychology Review, 32*(1), 229–247.

Becker, N., Noyes, K., & Cooper, M. (2016). Characterizing students' mechanistic reasoning about London dispersion forces. *Journal of Chemical Education, 93*(10), 1713–1724.

Bhatia, K. S., Stack, A., Sensibaugh, C. A., & Lemons, P. P. (2022). Putting the pieces together: Student thinking about transformations of energy and matter. *CBE – Life Sciences Education, 21*(4), 60.

Bjork, R. (2017). *Creating desirable difficulties to enhance learning*. Crown House Publishing.

Brand, C., Hartmann, C., Loibl, K., & Rummel, N. (2023). Do students learn more from failing alone or in groups? Insights into the effects of collaborative versus individual problem solving in productive failure. *Instructional Science, 51*(6), 953–976.

Brownell, S. E., Freeman, S., Wenderoth, M. P., & Crowe, A. J. (2014). BioCore guide: A tool for interpreting the core concepts of vision and change for biology majors. *CBE – Life Sciences Education, 13*(2), 200–211.

Chen, O., & Kalyuga, S. (2020). Exploring factors influencing the effectiveness of explicit instruction first and problem-solving first approaches. *European Journal of Psychology of Education, 35*(3), 607–624.

Chen, O., Retnowati, E., & Kalyuga, S. (2020). Element interactivity as a factor influencing the effectiveness of worked example–problem solving and problem solving–worked example sequences. *British Journal of Educational Psychology, 90*, 210–223.

Chen, O., Woolcott, G., & Kalyuga, S. (2021). Comparing alternative sequences of examples and problem-solving tasks: The case of conceptual knowledge. *The Educational and Developmental Psychologist, 38*(1), 158–170.

Chowrira, S. G., Smith, K. M., Dubois, P. J., & Roll, I. (2019). DIY productive failure: Boosting performance in a large undergraduate biology course. *npj Science of Learning, 4*(1), 1.

Clifford, M. M. (1984). Thoughts on a theory of constructive failure. *Educational Psychologist, 19*(2), 108–120.

Cooper, M. M., & Stowe, R. L. (2018). Chemistry education research—From personal empiricism to evidence, theory, and informed practice. *Chemical Reviews, 118*(12), 6053–6087.

Cooper, M. M., Williams, L. C., & Underwood, S. M. (2015). Student understanding of intermolecular forces: A multimodal study. *Journal of Chemical Education, 92*(8), 1288–1298.

DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology, 113*(4), 552–568.

Dubovi, I. (2018). Designing for online computer-based clinical simulations: Evaluation of instructional approaches. *Nurse Education Today, 69*, 67–73.

Endres, T., Lovell, O., Morkunas, D., Rieß, W., & Renkl, A. (2023). Can prior knowledge increase task complexity?–Cases in which higher prior knowledge leads to higher intrinsic cognitive load. *British Journal of Educational Psychology, 93*, 305–317.

Gabel, D. (1999). Improving teaching and learning through chemistry education research: A look to the future. *Journal of Chemical Education, 76*(4), 548.

Gilbert, J. K., & Treagust, D. F. (2009). *Multiple representations in chemical education* (Vol. 4). Springer.

Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater agreement and interrater reliability: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy, 9*(3), 330–338.

Halmo, S. M., Sensibaugh, C. A., Bhatia, K. S., Howell, A., Ferryanto, E. P., Choe, B., Kehoe, K., Watson, M., & Lemons, P. P. (2018). Student difficulties during structure–function problem solving. *Biochemistry and Molecular Biology Education, 46*(5), 453–463.

Halmo, S. M., Sensibaugh, C. A., Reinhart, P., Stogniy, O., Fiorella, L., & Lemons, P. P. (2020). Advancing the guidance debate: Lessons from educational psychology and implications for biochemistry learning. *CBE – Life Sciences Education, 19*(3), ar41.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. *Second Handbook of Research on Mathematics Teaching and Learning, 1*(1), 371–404.

Jacobson, M. J., Markauskaite, L., Portolese, A., Kapur, M., Lai, P. K., & Roberts, G. (2017). Designs for learning about climate change as a complex system. *Learning and Instruction, 52*, 1–14.

Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning, 7*(2), 75–83.
Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509–539.
Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*(3), 579.
Kalyuga, S., & Singh, A.-M. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review, 28*(4), 831–852.
Kapur, M. (2014). Productive failure in learning math. *Cognitive Science, 38*(5), 1008–1022.
Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist, 51*(2), 289–299.
Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21*(1), 45–83.
Kapur, M., Saba, J., & Roll, I. (2023). Prior math achievement and inventive production predict learning from productive failure. *npj Science of Learning, 8*(1), 15.
Lamnina, M., & Chase, C. C. (2019). Developing a thirst for knowledge: How uncertainty in the classroom influences curiosity, affect, learning, and transfer. *Contemporary Educational Psychology, 59*, 101785.
Loertscher, J., Green, D., Lewis, J. E., Lin, S., & Minderhout, V. (2014). Identification of threshold concepts for biochemistry. *CBE – Life Sciences Education, 13*(3), 516–528.
Loertscher, J., Lewis, J. E., Mercer, A. M., & Minderhout, V. (2018). Development and use of a construct map framework to support teaching and assessment of noncovalent interactions in a biochemical context. *Chemistry Education Research and Practice, 19*(4), 1151–1165.
Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review, 29*(4), 693–715.
Loibl, K., & Rummel, N. (2014). Knowing what you don't know makes failure productive. *Learning and Instruction, 34*, 74–85.
Matlen, B. J., & Klahr, D. (2013). Sequential effects of high and low instructional guidance on children's acquisition of experimentation skills: Is it all in the timing? *Instructional Science, 41*, 621–634.
McCarthy, K. S., & McNamara, D. S. (2021). The multidimensional knowledge in text comprehension framework. *Educational Psychologist, 56*(3), 196–214.
McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282.
McNeill, K. L., & Krajcik, J. (2008). Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 45*(1), 53–78.
Nokes-Malach, T. J., Richey, J. E., & Gadgil, S. (2015). When is it better to learn together? Insights from research on collaborative learning. *Educational Psychology Review, 27*, 645–656.
Offerdahl, E. G., Arneson, J. B., & Byrne, N. (2017). Lighten the load: Scaffolding visual literacy in biochemistry and molecular biology. *CBE – Life Sciences Education, 16*(1), es1.
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.
Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction, 22*(2), 129–184.
Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research, 91*(5), 761–798.
Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59–89.
Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.
Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*, 261–292.
Taber, K. S. (2013). Three levels of chemistry educational research. *Chemistry Education Research and Practice, 14*(2), 151–155.
Tansey, J. T., Baird, T., Jr., Cox, M. M., Fox, K. M., Knight, J., Sears, D., & Bell, E. (2013). Foundational concepts and underlying theories for majors in "biochemistry and molecular biology." *Biochemistry and Molecular Biology Education, 41*(5), 289–296.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53.

Toh, P. L. L., & Kapur, M. (2017). Is having more prerequisite knowledge better for learning from productive failure? *Instructional Science, 45*, 377–394.

Villafañe, S. M., Minderhout, V., Heyen, B. J., Lewis, J. E., Manley, A., Murray, T. A., Tienson-Tseng, H., & Loertscher, J. (2021). Design and implementation of a tool to assess students' understanding of metabolic pathways dynamics and regulation. *CBE – Life Sciences Education, 20*(3), ar35.

Weaver, J. P., Chastain, R. J., DeCaro, D. A., & DeCaro, M. S. (2018). Reverse the routine: Problem solving before instruction improves conceptual knowledge in undergraduate physics. *Contemporary Educational Psychology, 52*, 36–47.

Wright, L. K., Cardenas, J. J., Liang, P., & Newman, D. L. (2017). Arrows in biology: Lack of clarity and consistency points to confusion for learners. *CBE – Life Sciences Education, 17*(1), ar6.

Zhang, L., & Sweller, J. (2024). Instructional sequences in science teaching: Considering element interactivity when sequencing inquiry-based investigation activities and explicit instruction. *European Journal of Psychology of Education, 39*(4), 3791–3801.