Successful Problem Solving in Genetics Varies Based on Question Content

Jennifer S. Avena,†‡§ Betsy B. McIntosh,†‡ Oscar N. Whitney,† Ashton Wiens,‖ and Jennifer K. Knight†*

†Department of Molecular, Cellular, and Developmental Biology, ‡School of Education, and ‖Department of Applied Mathematics, University of Colorado Boulder, Boulder, CO 80309

ABSTRACT

Problem solving is a critical skill in many disciplines but is often a challenge for students to learn. To examine the processes both students and experts undertake to solve constructed-response problems in genetics, we collected the written step-by-step procedures individuals used to solve problems in four different content areas. We developed a set of codes to describe each cognitive and metacognitive process and then used these codes to describe more than 1800 student and 149 expert answers. We found that students used some processes differently depending on the content of the question, but reasoning was consistently predictive of successful problem solving across all content areas. We also confirmed previous findings that the metacognitive processes of planning and checking were more common in expert answers than student answers. We provide suggestions for instructors on how to highlight key procedures based on each specific genetics content area that can help students learn the skill of problem solving.

INTRODUCTION

The science skills of designing and interpreting experiments, constructing arguments, and solving complex problems have been repeatedly called out as critical for undergraduate biology students to master (American Association for the Advancement of Science, 2011). Yet each of these skills remains elusive for many students, particularly when the skill requires integrating and evaluating multiple pieces of information (Novick and Bassok, 2005; Bassok and Novick, 2012; National Research Council, 2012). In this paper, we focus on describing the steps students and experts take while solving genetics problems and determining whether the use of certain processes increases the likelihood of success.

The general process of solving a problem has been described as building a mental model in which prior knowledge can be used to represent ways of thinking through a problem state (Johnson-Laird, 2010). Processes used in problem solving have historically been broken down into two components: those that use domain-general knowledge and those that use domain-specific knowledge. Domain-general knowledge is defined as information that can be used to solve a problem in any field, including such strategies as rereading and identifying what a question is asking (Alexander and Judy, 1988; Prevost and Lemons, 2016). Although such steps are important, they are unlikely to be the primary determinants of success when specific content knowledge is required. Domain-specific problem solving, on the other hand, is a theoretical framework that considers one's discipline-specific knowledge and processes used to solve a problem (e.g., Prevost and Lemons, 2016). Domain-specific knowledge includes declarative (knowledge of content), procedural (how to utilize certain strategies), and conditional knowledge (when and why to utilize certain strategies) as they relate to a specific discipline (Alexander and Judy, 1988; Schraw and Dennison, 1994; Prevost and Lemons, 2016).

Previous studies on problem solving within a discipline have emphasized the importance of domain-specific declarative and conditional knowledge, as students need to understand and be able to apply relevant content knowledge to successfully solve problems (Alexander *et al.*, 1989; Alexander and Judy, 1988; Prevost and Lemons, 2016). Our prior work (Avena and Knight 2019) also supported this necessity. After students solved a genetics problem within a content area, they were offered a content hint on a subsequent content-matched question. We found that content hints improved performance overall for students who initially did not understand a concept. In characterizing the students' responses, we found that the students who benefited from the hint typically used the content language of the hint in their solution. However, we also found that some students who continued to struggle included the content language of the hint but did not use the information in their problem solutions. For example, in solving problems on predicted recombination frequency for linked genes, an incorrect solution might use the correct terms of map units and/or recombination frequency but not actually use map units to solve the problem. Thus, these findings suggest that declarative knowledge is necessary but not sufficient for complex problem solving and also emphasize the importance of procedural knowledge, which includes the "logic" of generating a solution (Avena and Knight, 2019). By definition, procedural knowledge uses both *cognitive* processes, such as providing reasoning for a claim or executing a task, and *metacognitive* processes, such as planning how to solve a problem and checking (i.e., evaluating) one's work (e.g., Kuhn and Udell, 2003; Meijer *et al.*, 2006; Tanner, 2012). We explore these processes in more detail below.

Cognitive Processing: Reasoning

Generating reasoning requires using one's knowledge to search for and explain an appropriate set of ideas to support or refute a given model (Johnson-Laird, 2010), so reasoning is likely to be a critical component of solving problems. Toulmin's original scheme for building a scientific argument (Toulmin, 1958) included generating a claim, identifying supporting evidence, and then using reasoning (warrant) to connect the evidence to the claim. Several studies have demonstrated a positive relationship between general reasoning "ability" (Lawson, 1978), defined as the ability to construct logical links between evidence and conclusions using conceptual principles, and performance (Cavallo, 1996; Cavallo *et al.*, 2004; Johnson and Lawson, 1998). As elaborated in more recent literature, there are many specific subcategories of reasoning. Students commonly use memorized patterns or formulas to solve problems: this approach is considered algorithmic and could be used to provide logic for a problem (Jonsson *et al.*, 2014; Nyachwaya *et al.*, 2014). Such algorithmic reasoning may be used with or without conveying an understanding of how an algorithm is used (Frey *et al.*, 2020). When an algorithm is not appropriate (or not used) in describing one's reasoning, but instead the solver provides a generalized explanation of underlying connections, this is sometimes referred to as "explanatory" or "causal" reasoning (Russ *et al.*, 2008). Distinct from causal reasoning is the domain-specific form of mechanistic reasoning, in which a mechanism of action of a biological principle is elaborated (Russ *et al.*, 2008; Southard *et al.*, 2016). Another common form of reasoning is quantitative reasoning, which can also be

described as statistical or, in other specialized situations, graph-construction reasoning (e.g., Deane *et al.*, 2016; Angra and Gardner, 2018). The detailed studies of these specific subcategories of reasoning have usually involved extensive interviews with students and/or very specific guidelines that prompt the use of a particular type of reasoning. Those who have explored students' unprompted general use of reasoning have found that few students naturally use reasoning to support their ideas (Zohar and Nemet, 2002; James and Willoughby, 2011; Schen, 2012; Knight *et al.*, 2015; Paine and Knight, 2020). However, with explicit training to integrate their knowledge into mental models (Kuhn and Udell, 2003; Osborne, 2010) or with repeated cueing from instructors (Russ *et al.*, 2008; Knight *et al.*, 2015), students can learn to generate more frequent, specific, and robust reasoning.

Metacognitive Processing

Successfully generating possible solutions to problems likely also involves *metacognitive thinking*. Metacognition is often separated into two components: metacognitive knowledge (knowledge about one's own understanding and learning) and metacognitive regulation (the ability to change one's approach to learning; Flavell, 1979; Jacobs and Paris, 1987; Schraw and Moshman, 1995). Metacognitive regulation is usually defined as including such processes as planning, monitoring one's progress, and evaluating or checking an answer (Flavell, 1979; Jacobs and Paris, 1987; Schraw and Moshman, 1995; Tanner, 2012). Several studies have shown that helping students use metacognitive strategies can benefit learning. For example, encouraging the planning of a possible solution beforehand and checking one's work afterward helps students generate correct answers during problem solving (e.g., Mevarech and Amrany, 2008; McDonnell and Mullally, 2016; Stanton *et al.*, 2015). However, especially compared with experts, students rarely use metacognitive processes, despite their value (Smith and Good, 1984; Smith, 1988). Experts spend more time orienting, planning, and gathering information before solving a problem than do students, suggesting that experts can link processes that facilitate generating a solution with their underlying content knowledge (Atman *et al.*, 2007; Peffer and Ramezani, 2019). Experts also check their problem-solving steps and solutions before committing to an answer, steps not always seen in student responses (Smith and Good, 1984; Smith, 1988). Ultimately, prior work suggests that, even when students understand content and employ appropriate cognitive processes, they may still struggle to solve problems that require reflective and regulative skills.

Theoretical Framework: Approaches to Learning

Developing domain-specific conceptual knowledge requires integrating prior knowledge and new disciplinary knowledge (Schraw and Dennison, 1994). In generating conceptual knowledge, students construct mental models in which they link concepts together to generate a deeper understanding (Johnson-Laird, 2001). These mental constructions involve imagining possible relationships and generating deductions and can be externalized into drawn or written models for communicating ideas (Chin and Brown, 2000; Bennett *et al.*, 2020). Mental models can also trigger students to explain their ideas to themselves (self-explanation), which can also help them solve problems (Chi *et al.*, 1989).

As our goal is to make visible how students grapple with their knowledge during problem solving, we fit this study into the approaches to learning framework (AtL: Chin and Brown, 2000). This framework, derived from detailed interviews of middle-school students solving chemistry problems, defines five elements of how students approach learning and suggests that these components promote deeper learning. Three of these elements are identifiable in the current study: engaging in explanations (employing reasoning through understanding and describing relationships and mechanisms), using generative thinking (application of prior knowledge and analogical transfer), and engaging in metacognitive activity (monitoring progress and modifying approaches). The remaining two elements: question asking (focusing on facts or on understanding) and depth of approaching tasks (taking a deep or a surface approach to learning: Biggs, 1987) could not be addressed in our study. However, previous studies showed that students who engage in a deep approach to learning also relate new information to prior knowledge and engage in reasoning (explanations), generate theories for how things work (generative thinking), and reflect on their understanding (metacognitive activity). In contrast, those who engage in surface approaches focus more on memorized, isolated facts than on constructing mental or actual models, demonstrating an absence of the three elements described by this framework. Biggs (1987) also previously provided evidence that intrinsically motivated learners tended to use a deep approach, while those who were extrinsically motivated (e.g., by grades), tended to use a surface approach. Because solving complex problems is, at its core, about how students engage in the learning process, these AtL components helped us frame how students' learning is revealed by their own descriptions of their thinking processes.

Characterizing Problem-Solving Processes

Thus far, a handful of studies have investigated the processes adult students use in solving biology problems, and how these processes might influence their ability to develop reasonable answers (Smith and Good, 1984; Smith, 1988; Nehm, 2010; Nehm and Ridgway, 2011; Novick and Catley, 2013; Prevost and Lemons, 2016; Sung et al., 2020). In one study, Prevost and Lemons (2016) collected and analyzed students' written documentation of their problem-solving procedures when answering multiple-choice questions. Students were taught to document their step-by-step thinking as they answered multiple-choice exam questions that ranged from Bloom's levels 2 to 4 (understand to analyze; Bloom et al., 1956), describing the steps they took to answer each question. The authors' qualitative analyses of students' documented problem solving showed that students frequently used domain-general test-taking skills, such as comparing the language of different multiple-choice distractors. However, students who correctly answered questions tended to use more domain-specific procedures that required knowledge of the discipline, such as analyzing visual representations and making predictions, than unsuccessful students. When students solved problems that required the higher-order cognitive skills of application and analysis, they also used more of these specific procedures than when solving lower-level questions. Another recent study explored how students solved exam questions on the genetic topics of recombination and nondisjunction through in-depth clinical interviews (Sung et al., 2020). These authors described two approaches that are not conceptual: using algorithms to bypass conceptual thinking and using non–biology specific test-taking strategies (e.g., length of answer, specificity of terminology). They also showed that students sometimes alternate between using an algorithm and a conceptual strategy, defaulting to the algorithm when they do not understand the underlying biological concept.

From prior work specifically on students' understanding of genetics, we know that certain content areas are persistently challenging despite instructional focus (Smith et al., 2008). On these topics, students who enter a course with a particular misunderstanding are significantly more likely to retain this way of thinking at the end of the course than they are to switch to another answer (Smith and Knight, 2012). We have focused on these topic areas in previous studies (Prevost et al., 2016; Sieke et al., 2019) by designing questions to reveal what students are thinking and using the results as an instructional tool. In addition, we recently found that students performed better on answering constructed-response questions on chromosome separation and inheritance patterns than on calculating the probability of inheritance, although all three areas were challenging (Avena and Knight, 2019). To our knowledge, no prior work has compared the processes students use when solving different types of genetics problems or used a large sample of students to characterize all processes in which students engage. In the study described here, we ground our work in domain-specific problem solving (e.g., Prevost and Lemons, 2016), scientific argumentation and reasoning (Toulmin, 1958), and the AtL framework (Chin and Brown, 2000). We build upon these prior bodies of work to provide a complete picture of the cognitive and metacognitive processes described by both students and experts as they solve complex problems in four different content areas of genetics. Our research questions were as follows:

Research Question 1. How do experts and students differ in their description of problem-solving processes, using a much larger sample size than found in the previous literature (e.g., Chi et al., 1981; Smith and Good, 1984; Smith, 1988; Atman et al., 2007; Peffer and Ramezani, 2019).

Research Question 2. Are certain problem-solving processes more likely to be used in correct than in incorrect student answers?

Research Question 3. Do problem-solving processes differ based on content and are certain combinations of problem-solving processes associated with correct student answers for each content area?

METHODS

Mixed-Methods Approach

This study used a mixed-methods approach, combining both qualitative and quantitative research methods and analysis to understand a phenomenon more deeply (Johnson et al., 2007). Our goal was to make student thinking visible by collecting written documentation of student approaches to solving problems (qualitative data), in addition to capturing answer correctness (quantitative data), and integrating these together in our analyses. The student responses serve as a rich and detailed data set that can be interpreted using the qualitative process of assigning themes or codes to student writing (Hammer and Berland, 2014). In a qualitative study, the results of the coding

process are unpacked using examples and detailed descriptions to communicate the findings. In this study, we share such qualitative results but also convert the coded results into numerical representations to demonstrate patterns and trends captured in the data. This is particularly useful in a large-scale study, because the output can be analyzed statistically to allow comparisons between categories of student answers and different content areas.

Subjects

Students in this study were enrolled in an introductory-level undergraduate genetics course for biology majors at the University of Colorado in Spring 2017 ($n = 416$). This course is the second in a two-course introductory series, with the first course being Introduction to Cell and Molecular Biology. The students were majority white, 60% female, and 63% were in their first or second year. Ninety percent of the students were majoring in biology or a biology-related field (neuroscience, integrative physiology, biochemistry, biomedical engineering). Of the students enrolled in the course, 295 students consented to be included in the study; some of the student responses have been previously described in the prior study (Avena and Knight, 2019). We recruited experts from the Society for the Advancement of Biology Education Research Listserv by inviting graduate students, postdoctoral fellows, and faculty to complete an anonymous online survey consisting of the same questions that students answered. Of the responses received, we analyzed responses from 52 experts. Due to the anonymous nature of the survey, we did not collect descriptive data about the experts.

Problem Solving

As part of normal course work, students were offered two practice assignments covering four content areas related to each of two course exams (also described in Avena and Knight, 2019). Students could answer up to nine questions in blocks of three questions each, in randomized order, for three of the four content areas. Expert participants answered a series of four questions, one in each of the four content areas. All questions were offered online using the survey platform Qualtrics. All participants were asked to document their problem-solving processes as they completed the questions (as in Prevost and Lemons 2016), and they were provided with written instructions and an example in the online platform only (see Supplemental Material); no instructions were provided in class, and no explicit discussion of types of problem-solving processes to use were provided in class throughout the semester. Students could receive extra credit up to ~1% of the course point total, obtaining two-thirds credit for explaining their answer and an additional one-third if they answered correctly. All students who completed the assignment received credit regardless of their consent to participate in the research.

We used questions developed for a prior study (Avena and Knight, 2019) on four challenging genetics topics: calculation of the probability of inheritance across multiple generations (Probability), prediction of the cause of an incorrect chromosome number after meiosis (Nondisjunction), interpretation of a gel and pedigree to determine inheritance patterns (Gel/Pedigree), and prediction of the probability of an offspring's genotype using linked genes (Recombination; see example in Figure 1; all questions presented in Supplemental Material).

The Bustamante family, shown below, has a genetically inherited disease caused by a 117 bp deletion mutation in an exon of a single gene. DNA for this gene was amplified from each member of the family and run on the gel shown below. Using the pedigree and the gel, what is the most likely mode of inheritance for this disease?
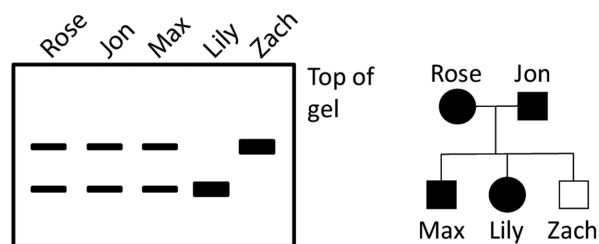

FIGURE 1. Sample problem for students from the Gel/Pedigree content area. Problems in each content area contain a written prompt and an illustrated image, as shown in this example.

These content areas have previously been shown to be challenging based on student performance (Smith *et al.*, 2008; Smith and Knight, 2012; Avena and Knight, 2019). Each content area contained three isomorphic questions that addressed the same underlying concept, targeted higher-order cognitive processes (Bloom *et al.*, 1956), and contained the same amount of information with a visual (Avena and Knight, 2019). Each question had a single correct answer and was coded as correct (1) or incorrect (0). For each problem-solving assignment, we randomized 1) the order of the three questions within each content area for each student and 2) the order in which each content area was presented. During each set of three isomorphic questions, while solving one of the isomorphic problems, students also had the option to receive a "content hint," a single most commonly misunderstood fact for each content area. We do not discuss the effects of the content hints in this paper (instead, see Avena and Knight, 2019).

Process Coding

Students may engage in processes that they do not document in writing, but we are limited to analyzing only what they provide in their written step-by-step descriptions. For simplicity, throughout this paper, a "process" is a thought documented by the participant that is coded as a particular process. When we refer to "failure" to use a process, we mean that a participant did not describe this thought process in the answer. Our initial analysis of student processes used a selection of codes from Prevost and Lemons (2016) and Toulmin's (1958) original codes of Claim and Reason. We note that all the problems we used can potentially be solved using algorithms, memorized patterns previously discussed and practiced in the class, which may have limited the reasoning students supplied. Because of the complexity of identifying different types of reasoning, we did not further subcategorize the reasoning category in the scheme we present, as this is beyond the scope of this paper. We used an emergent coding process (Saldana, 2015) to identify additional and different processes, including both cognitive and metacognitive actions. Thus, our problem-solving processes (PsP) coding

TABLE 1. Problem-solving process (PsP): Code categories, definitions, and examples[a]

Strategy category	Individual process codes	Description	Example
Orientation	Notice	Identifying components in the question stem.	"I'm underlining that this is autosomal recessive."
	Identify Similarity	Noticing similarity between problems.	"This is like the last problem."
	Identify Concept	Explicitly describing the type of problem.	"This is a meiosis problem."
	Recall	Remembering a fact or definition without direct application to the problem.	"Nondisjunction in meiosis 1 yields two different alleles in one gamete."
Metacognition	Plan	Outlining next steps.	"First, I need to figure out who the disease carriers are, determine the probabilities that they are affected, and then use the product law to determine the likelihood the child will be affected."
	Check	Checking solution steps and/or final answer.	"I had to go back and make sure I understood what I need to answer this question correctly."
	Assess Difficulty	Expressing perceived difficulty or unfamiliarity.	"This is a difficult problem and I'm not sure how to get started."
Execution	Use Information	Applying a single piece of information related to the problem.	"Individual A has a band that migrated slower than Individual B's band."
	Integrate	Linking visual representations with other information	"Looking at the gel and table together, I notice that Lily and Max are both affected, even though Lily has 2 copies of the disease gene and Max has only 1 copy."
	Draw	Drawing or visualizing problem components.	"I'm drawing a Punnett Square in my head."
	Calculate	Using any mathematical statement.	"Together, there is a ⅔ chance that both parents are heterozygous since $1 \times ⅔$ is ⅔."
Reasoning	Reason	Providing a logical explanation for a preliminary or final conclusion.	"I determined that II-6 has a 100% chance of being a carrier, because II-6 is not affected, but has a mother that is affected."
Conclusion	Eliminate	Ruling out a final answer.	"It couldn't be X-linked dominant."
	Claim	Providing an answer statement.	"There is a ⅔ chance that the child will be affected."
Error	Misinterpret	Misunderstanding the question stem.	"The problem was asking for the band that causes the disease." [Question was asking for the mode of inheritance]
General	Clarify	Clarifying the question stem and problem.	"I identified what the question was asking."
	State the Process	Stating an action abstractly (no details).	"I am determining the genotypes."
	Restate	Restating a previously stated process	

[a]Examples of student responses are to a variety of content areas and have been edited for clarity. Each individual process code captures the student's description, regardless of whether the statement is correct or incorrect.

scheme captures the thinking that students document while solving genetics problems (see individual process codes in Table 1). We used HyperRESEARCH software (ResearchWare, Inc.) to code each individual's documented step-by-step processes. A step was typically a sentence and sometimes contained multiple ideas. Each step was given one or more codes, with the exception of reasoning supporting a final conclusion (see Table 2 for examples of coded responses). Each individual process code captures when the student describes that process, regardless of whether the statement is correct or incorrect. Four raters (J.K.K., J.S.A., O.N.W., B.B.M.) coded a total of 24 student answers over three rounds of coding and discussion to reach consensus and identify a final coding scheme. Following

agreement on the codes, an additional 12 answers were coded by the four raters to determine interrater agreement. Specifically, in these 12 answers, there were 150 instances in which a code for a step was provided by one or more raters. For each of these 150 instances, we identified the number of raters who agreed. We then calculated a final interrater agreement of 83% by dividing the total number of raters who agreed for all 150 instances (i.e., 524) by the total number of possible raters to agree for four raters in 150 instances (i.e., 600). We excluded answers in which students did not describe their problem-solving steps and those in which students primarily or exclusively used domain-general processes (i.e., individual process codes within the General strategy category in Table 1) or made claims

TABLE 2. Examples of expert and student documented problem solving on a Gel/Pedigree problem with associated process codes[a]

Expert answer: Eliot	Process
The question states that the mutation is a deletion in a single gene.	Clarify
We don't know yet if one copy of the mutation can cause the disease.	Use Information
We have a gel to look at and a pedigree, so there's lots of information, and I can use both to make sure I have the inheritance right.	Plan
I look at the pedigree and think it looks like a dominant disease because of the inheritance pattern.	Claim and Reason
Actually, it has to be dominant, just from the pedigree, because otherwise Zach could not be unaffected.	Claim and Reason
I need the gel to decide if it's X linked or autosomal.	Plan
The gel shows two alleles for just about everyone, so that almost answers the question right off the bat—the smaller allele is the mutant allele, and Rose, Jon and Max all have one copy of this allele and one copy of the normal, larger allele, and they have the disease.	Integrate
Sounds like autosomal dominant to me.	Claim
To be sure I check out just the males in the pedigree—	Check
Zach has only normal allele copies and all the other males have two different alleles.	Use Information
Thus, the disease cannot be caused by a gene on the X chromosome; since males only have one copy of the X, they would only have one allele.	Eliminate and Reason
It must be autosomal dominant.	Claim

Correct student answer: Cassie	Process
The gel shows that Lily and Zach do not have two separate alleles for the gene, they both only have one.	Use Information
I read that the question mentions a missing exon.	Notice
This means that the missing exon probably codes for something that, when missing, causes a disease phenotype.	Use Information
Lily has the disease, and Zach doesn't, as seen in the pedigree.	Use Information
Rose and Jon, the parents, have both alleles.	Integrate
This means they are heterozygous for the disease	Use Information
If both parents are heterozygous for the disease, then it is probably inherited in a dominant manner.	Claim and Reason
Everyone with the smaller segment of the gene (Rose, Jon, Max, and Lily) has the disease. This must be the dominant allele that causes the phenotype.	Integrate and Reason
Since Zach doesn't have that gene, he is normal.	Integrate
The disease doesn't have any seeming tie to the X chromosome, so it is autosomal. It is inherited autosomal dominant.	Eliminate and Claim

Incorrect student answer: Ian	Process
Read question	Clarify
Look at gel and pedigree	Clarify
Notice both son and daughter are affected.	Use Information
One son is not.	Use Information
Do a X-linked dominant cross.	Draw
Outcome was 1/2 daughters homozygous dominant and 1/2 heterozygous.	Integrate
1/2 sons affected and 1/2 not.	Integrate
[Final answer indicated in a section following this documentation: X-linked dominant]	

[a]The responses above are all solutions to the question in Figure 1.

without any other supporting codes. The latter two exclusion criteria were used because such responses lacked sufficient description to identify the thought processes. The final data set included a total of 1853 answers from 295 students and 149 answers from 52 experts. We used only correct answers from experts to serve as a comparison to student answers, excluding an additional 29 expert answers that were incorrect.

After initial coding and analyses, we identified that student use of drawing was differentially associated with correctness based on content area. Thus, to further characterize drawing use, two raters (J.S.A. and J.K.K.) explored incorrect student answers from Probability and Recombination. One rater examined 33 student answers to identify an initial characterization, and then two raters reviewed a subset of answers to agree upon a final scheme. Each rater then individually categorized a portion of the student answers, and the final interrater agreement

on 10 student answers was 90%. Interrater agreement was calculated as described earlier, with each answer serving as one instance, so we divided the total number of raters agreeing for each answer (i.e., 18) by the total possible number of raters agreeing (i.e., 20).

Statistical Analyses

The unit of analysis for all models considered is an individual answer to a problem. We investigate three variations of linear models, specified below. The response variable in all cases is binary (presence/absence of process or correct/incorrect answer). Thus, the models are generalized linear models, and, more specifically, logistic regression models. Because our data contain repeated measures in the form of multiple answers per student, we specifically use generalized linear mixed models (GLMM) to include a random effect on the intercept term in all

models, grouped by participant identifier (Gelman and Hill, 2006; Theobald, 2018). This component of the model accounts for variability in the baseline outcome between participants. In our case, we can model each student's baseline probability of answering a problem correctly or each participant's baseline probability of using a given process (e.g., one student may use Reason more frequently than another student). Accounting for this variation yields better estimates of the fixed effects in the models.

In these logistic regression models, the predicted value of the response variable is the log odds of the presence of a process or correctness. The log odds can take on any real number and can be transformed into a more interpretable probability using the logistic function $f(x) = \dfrac{e^x}{1 + e^x}$, as they are in some tables presented in this study (Gelman and Hill, 2006). For additional ease of interpretation, we report cumulative predicted probabilities for each predictor group by adding the relevant combination of coefficient estimates to the intercept, which represents the baseline case.

The fitted models give some, but not all, pairwise comparisons among predictor groups. We conducted pairwise post hoc comparisons (e.g., expert vs. correct student, expert vs. incorrect student, correct student vs. incorrect student, or among the four content areas) to draw inferences about the differences among all groups. In particular, we performed Tukey pairwise honestly significant difference (HSD) tests for all pairs of groups, comparing estimated marginal means (estimated using the fitted model) on the logit scale. Using estimated marginal means corrects for unbalanced group sample sizes, and using the Tukey HSD test provides adjusted p values, facilitating comparison to a significance level of $\alpha = 0.05$.

To ease reproducibility, we use "formula" notation conventionally used in R to specify the models we employ in this paper, which has the following general form: outcome = fixed effect + (1 | group). The random effects component is specified within parentheses, with the random effect on the left of the vertical bar and the grouping variable on the right.

To address whether the likelihood of process use differs among experts, correct students, and incorrect students (research questions 1 and 2), we used the following model (model 1):

Process present = Expert/Student answer status + (1| ID)

where "Process present" is the binary factor response variable: absent (0)/present (1); "Expert/Student answer status" is the fixed effect: Factor-level grouping: incorrect student (0)/correct student (1)/correct expert (2); and "(1|ID)" is the random effect intercept based on participant ID. This random effect intercept accounts for variability in the baseline outcome between participants. To address whether the likelihood of process use differs by content area (research question 3), we used the following model (model 2):

Process present = Content area + (1| ID)

where "Process present" is the response variable as described for model 1; "Content area" is the fixed effect: Factor-level grouping: Probability (1)/Nondisjunction (2)/Gel-Pedigree (3)/Recombination (4); and "(1|ID)" is the random effect as described for model 1.

To examine what combination of factors are associated with correctness within each content area (research question 3), we used a GLMM model with a lasso penalty (Groll and Tutz, 2014; Groll, 2017). The lasso model was used for variable selection and to prevent overfitting when many predictors are available (the model is not used specifically to identify significance of predictors). We used a lasso model in this case to find out which of the process variables, in combination, were most predictive of student answer correctness in the GLMM, using the following model (model 3):

Student answer correctness = Process 1 + Process 2 + …
+ Process X + (1| ID)

where "Student answer correctness" is the response variable: incorrect (0)/correct (1); "Process 1 + Process 2 + … + Process X" is the list of process factors entered into the model as the fixed effect: absent (0)/present (1); and "(1|ID)" is the random effect as described for models 1 and 2. We identified which components were associated with correctness by seeing which predictor coefficients remained non-zero in a representative lasso model. We identified a representative model for each content area by first identifying the lasso penalty with the lowest Akaike information criterion (AIC) to reduce variance and then identifying a lasso penalty with a similar AIC that could be used across all content areas. Because a penalty parameter of 25 and the penalty parameter with the lowest AIC for each content area had similar AIC values, we consistently used a penalty parameter of 25. Note that when the penalty parameter is set to zero, the GLMM model is recovered. On the other hand, when the penalty parameter is very large, no predictors are included in the model. Thus, the selected penalty parameter forced many, but not all, coefficients to 0, giving a single representative model for each content area.

All models and tests were performed in R (v. 3.5.1). We used the lme4 package in R (Bates *et al.*, 2015) for models 1 and 2, and estimation of parameters was performed using residual maximum likelihood. For model 3, we used the glmmLasso package, and the model was fit using the default EM-type estimate. Post hoc pairwise comparisons were performed using the emmeans package.

Human Subjects Approval

Human research was approved by the University of Colorado Institutional Review Board (protocols 16-0511 and 15-0380).

RESULTS

The PsP Coding Scheme Helps Describe Written Cognitive and Metacognitive Processes

We developed a detailed set of codes, which we call the PsP scheme to characterize how individuals describe their solutions to complex genetics problems. Table 1 shows the 18 unique processes along with descriptions and examples for each. With the support of previous literature, we grouped the individual processes into seven strategies, also shown in Table 1. All strategies characterized in this study were domain specific except the General category, which is domain general. We categorized a set of processes as Orientation based on a previously published taxonomy for think-aloud interviews (Meijer *et al.*, 2006) and on information management processes from the Metacognitive Awareness Inventory (Schraw and Dennison, 1994). Orienting

processes include: Notice (identifying important information in the problem), Recall (activating prior knowledge without applying it), Identify Similarity (among question types), and Identify Concept (the "type" of problem). Orientation processes are relatively surface level, in that information is observed and noted, but not acted on. The Metacognition category includes the three common elements of planning (Plan), monitoring (Assess Difficulty), and evaluating (Check) cited in the metacognitive literature (e.g., Schraw and Moshman, 1995; Tanner, 2012). The Execution strategy includes actions taken to explicitly solve the problem, including Use Information (apply information related to the problem), Integrate (i.e., linking together two visual representations provided to solve the problem or linking a student's own drawing to information in the problem), Draw, and Calculate. The Use Information category is distinguished from Recall by a student applying a piece of information (Use Information) rather than just remembering a fact without directly using it in the problem solution (Recall). Students may Recall and then Use Information, just Recall, or just Use Information. If a student used the Integrate process, Use Information was not also coded (i.e., Integrate supersedes Use Information). The Reasoning strategy includes just one general process of Reason, which we define as providing an explanation or rationale for a claim, as previously described in Knight *et al.* (2013), Lawson (2010), and Toulmin (1958). The Conclusion strategy includes Eliminate and Claim, processes that provide types of responses to address the final answer. The single process within the Error strategy category, Misinterpret, characterizes steps in which students misunderstand the question stem. Finally, the General category includes the codes Clarify, State the Process, and Restate, all of which are generic statements of execution, representing processes that are domain general (Alexander and Judy, 1988; Prevost and Lemons, 2016).

To help visualize the series of steps students took and how these steps differed across answers and content areas, we provide detailed examples in Tables 2 and 3. In Table 2, we provide three examples of similar-length documented processes to the same Gel/Pedigree problem (Figure 1) from a correct expert, a correct student, and an incorrect student. Note the multiple uses of planning and reasoning in the expert answer, multiple uses of reasoning in the correct student answer, and the absence of both such processes in the incorrect student answer. The reasoning used in each case provides a logical explanation for the claim, which either immediately precedes or follows the reasoning statement. For example, in the second incident of Claim and Reason for Eliot, "because otherwise Zach could not be unaffected" is a logical explanation for the claim "it has to be dominant." Similarly, for Cassie's Claim and Reason code, "If both parents are heterozygous for the disease" is a logical explanation for the claim "it is probably inherited in a dominant manner." Table 3 provides additional examples of correct student answers to the remaining three content areas. Note that for Probability and Recombination questions, the Reason process often explains why a certain genotype or probability is assigned (e.g., "otherwise all or none of the children would have the disease" explains why "Both parents of H and J must be Dd" in Li's Probability answer) or how a probability is calculated, for example, "using the multiplication rule" (Li's Probability explanation) or "multiply that by the 100% chance of getting 'af' from parent 2" (Preston's Recombination explanation). In Non-

disjunction problems, a student may claim that a nondisjunction occurred in a certain stage of meiosis (the Claim) because it produces certain gamete genotypes consistent with such an error (the Reason), as seen in Gabrielle's answer.

Across All Content Areas, Expert Answers Are More Likely Than Student Answers to Contain Orientation, Metacognition, and Execution Processes

For each category of answers (expert, correct student, and incorrect student), we calculated the overall percent of answers that contained each process and compared these frequencies. Note that, in all cases, frequency represents the presence of a process in an answer, not a count of all uses of that process in an answer. The raw frequency of each process is provided in Table 4, columns 2–4. To determine statistical significance, we used GLMM to account for individual variability in process use. The predicted likelihood of each process per group and pairwise comparisons between groups from this analysis is provided in Table 4, columns 5–10. These comparisons show that expert answers were significantly more likely than student answers to contain the processes of Identify concept, Recall, Plan, Check, and Use Information (Table 4 and Supplemental Table S1). The answers in Table 2 represent some of the typical trends identified for each group. For example, expert Eliot uses both Plan and Check, but these metacognitive processes are not used by either student, Cassie (correct answer) or Ian (incorrect answer).

Across All Content Areas, Correct Student Answers Are More Likely Than Incorrect Answers to Contain the Processes of Reason and Eliminate

Students most commonly used the processes Use Information, Reason, and Claim, each present in at least 50% of both correct and incorrect student answers (Table 4). The processes Notice, Recall, Calculate, and Clarify were present in 20–50% of both correct and incorrect student answers (Table 4). In comparing correct and incorrect student answers across all content areas, we found that Integrate, Reason, Eliminate, and Clarify were more likely used in correct compared with incorrect answers (Table 4). As illustrated in Table 2, the problem-solving processes in Cassie's correct answer include: reasoning for a claim of dominant inheritance and eliminating when ruling out the possibility of an X-linked mode of inheritance. However, in describing the incorrect answer, Ian fails to document use of either of these processes.

Process Use Varies by Question Content

To determine whether student answers contain different processes depending on the content of the problem, we separated answers, regardless of correctness, by content area. We then excluded some processes: we did not analyze the Error and General codes, as well as Claim, which was seen in virtually every answer across content areas. We also excluded the very rarely seen processes of Identify Similarity and Identify Concept, which were present in 5% or fewer of both incorrect and correct student answers. For the remaining 11 processes, we found that each content area elicited different frequencies of use, as shown in Table 5 and Supplemental Table S2. Some processes were nearly absent in a content area: Calculate was rarely seen in answers to Nondisjunction and Gel/Pedigree

TABLE 3. Examples of correct student documented problem solving with associated process codes for each content area[a]

Correct student answer: Li for Probability (Wilson's disease question)	Process
We must initially conclude that Hillary and Justin's parents are carriers for the disease, because they both have children who are affected.	Use Information and Reason
However, neither Hillary nor Justin has the disease, so they must not be recessive for it (dd).	Use Information and Reason
Both parents of H and J must be Dd, otherwise all or none of the children would have the disease.	Use Information and Reason
Chance of Hillary/Justin being Dd is 2/3. Chance being DD is 1/3.	Use Information
If Hillary and Justin are Dd, then their child has a 1/4 chance of being diseased.	Use Information
So, using the multiplication rule, the child has a 2/3*2/3*1/4 chance of being diseased, which is 1/9.	Reason and Calculate and Claim

Correct student answer: Preston for Recombination (Aldose gene question)	Process
Write down the intended offspring which would be Af/af	Use Information
See if you have to use recombination (Are the goal alleles already paired in the parents?)	Plan
Yes they are we need one of the alleles to be af which one parents has two of and the other needs to be Af which the other parent has one of	Use Information
What is the probability the offspring will inherit af from one parent?	Plan
100% because that is the only allele he has to give	Use Information
What is the probability the offspring will inherit Af from the other parent?	Plan
The offspring should be able to inherit Af 50% because its either one or the other	Use Information and Reason
BUT we have to consider the possibility of recombination	Use Information
Since the two genes are 10 map units apart it means that there is a 10% chance they will recombine	Recall
So now instead of 50% of inheriting Af from the first parent, there is a 10% of recombination meaning there is a 90% chance of getting either of the non-recomb alleles	Eliminate and Claim
So now split 90% in half	Calculate
There is now a 45% chance of either Af or aF from parent 1	Use Information
Multiply that by the 100% chance of getting af from parent 2	Reason and Calculate
The answer is there is a 45% chance the offspring will be able to only produce Aldose	Claim

Correct student answer: Gabrielle for Nondisjunction (chromosome 7 Q/q gene question)	Process
Read the question	Clarify
Observe the chromosomes present in Daryl's normal diploid cell	Clarify
Observe the chromosomes present in Daryl's genotype and understand a nondisjunction occurred	Claim
Draw a Daryl's diploid cell going through meiosis with a nondisjunction in meiosis 1 and normal meiosis 2	Draw
See that it produces gametes with genotypes BBQ and q or QqB and B. Recognize these do not match the genotype of the gamete	Integrate and Eliminate
Draw Daryl's diploid cell going through meiosis with nondisjunction in meiosis 2 and regular meiosis 1	Draw
See that it produces gametes with genotypes QQB and B, qqB and B, qqB and B or BBq and q	Reason
Recognize that the nondisjunction in meiosis 2 creates the genotype seen in the gamete. Conclude that meiosis 2 was affected	Claim

[a]Responses edited slightly for clarity. See Table 2 for a correct student documented solution to the Gel/Pedigree problem.

questions and Eliminate was rarely seen in answers to Probability and Recombination questions. Furthermore, in answering Probability questions, students were more likely to use the processes Plan and Use Information than in any other content area. Recall was most likely in Recombination and least likely in Gel/Pedigree. Examples of student answers showing some of these trends are shown in Table 3.

The Combination of Processes Linked to Correctness Differs by Content Area

Performance varied by content area. Students performed best on Nondisjunction problems (75% correct), followed by Gel/Pedigree (73%), Probability (54%), and then Recombination (45%). Table 6 shows the raw data of process prevalence for correct and incorrect student answers in each of the four content areas. To examine the combination of problem-solving processes associated with correct student answers for each content area, we used a representative GLMM model with a lasso penalty. This type of analysis measures the predictive value of a process on answer correctness, returning a coefficient value. The presence of a factor with a higher positive coefficient increases the probability of answering correctly more than a factor with a lower positive coefficient. With each additional positive factor in the model, the likelihood of answering correctly increases in an additive manner (Table 7 and Supplemental Table S3). To interpret these values, we show the probability estimates (%) for each process, which represent the probability that an answer will be correct in the presence of one or more processes (Table 7). The strength of association of the process with correctness, measured by positive coefficient size, is listed in descending order. Thus, for each content area, the process with the strongest positive association to a correct answer is listed first. A process with a negative coefficient (a negative association with correctness) is listed last, and models with negative associations are highlighted in gray in Table 7. An example of how to interpret the GLMM model is as follows. For the content area of Probability, Calculate

TABLE 4. Comparison of students' and experts' process use across all four content areas[a]

Process	Prevalence of process (% of answers)			Predicted probability (%) from GLMM			GLMM *p* value (pairwise comparisons)		
	Incorrect student	Correct student	Correct expert	Incorrect student (i)	Correct student (c)	Correct expert (e)	i–c	i–e	c–e
Orientation									
Notice	29.69	32.28	32.21	26.29	22.08	23.52	ns	ns	ns
Identify Similarity	3.64	5.31	0.67	0.52	0.83	0.11	ns	ns	ns
Identify Concept	3.64	4.59	20.13	0.08	0.09	1.39	ns	**	**
Recall	24.16	28.42	45.64	21.40	22.19	44.77	ns	***	***
Metacognition									
Plan	18.62	25.81	51.68	15.43	17.28	53.49	ns	***	***
Check	6.34	10.43	47.65	2.50	4.06	46.23	ns	***	***
Assess Difficulty	7.29	4.41	14.09	1.55	0.83	4.29	*	ns	**
Execution									
Use Information	70.85	70.41	91.28	75.53	71.39	93.28	ns	***	***
Integrate	16.46	23.47	30.20	13.30	19.40	27.33	**	**	ns
Draw	20.38	16.28	24.83	10.83	7.61	16.49	ns	ns	ns
Calculate	44.13	40.65	41.61	45.93	38.66	40.40	*	ns	ns
Reasoning									
Reason	82.05	92.36	93.29	91.80	96.68	97.40	***	*	ns
Conclusion									
Eliminate	4.45	12.86	16.78	2.85	8.56	12.72	***	***	ns
Claim	97.44	98.38	96.64	99.96	99.97	99.94	ns	ns	ns
Error									
Misinterpret	2.29	0.18	0.67	0.02	0.00	0.01	NA	NA	NA
General									
Clarify	37.79	50.99	73.83	17.80	39.74	99.06	***	***	ns
State the Process	5.80	3.42	14.09	0.55	0.35	2.22	ns	ns	**
Restate	3.91	4.95	4.03	2.81	3.56	2.91	ns	ns	ns
n answers	741	1112	149						

[a]Pairwise comparison: incorrect students to correct students (i–c), incorrect students to correct experts (i–e), correct students to correct experts (c–e). NA, no comparison made due to predicted probability of 0 in at least one group. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; ns: $p > 0.05$. See Supplemental Table S1 for standard error of coefficient estimates. Interpretation example: 82.05% and 92.36% of incorrect and correct student answers, respectively, contained Reason. The GLMM, after accounting for individual variability, predicts the probability of an incorrect student using Reason to be 91.80%, while the probability of a correct student using Reason is 96.68%.

(strongest association with correctness), Use Information, and Reason (weakest association with correctness) in combination are positively associated with correctness; Draw is the only negative predictor of correctness. For this content area, the intercept indicates a 7.31% likelihood of answering correctly in the absence of any of the processes tested. If an answer contains Calculate only, there is a 40.19% chance the answer will be correct. If an answer contains both Calculate and Use Information, there is a 58.60% chance the answer will be correct, and if the answer contains the three processes of Calculate, Use Information, and Reason combined, there is a 67.56% chance the answer will be correct. If Draw is present in addition to these three processes, the chance the answer will be correct slightly decreases to 66.40%. For Recombination, the processes of Calculate, Recall, Use Information, Reason, and Plan in combination are associated with correctness, and Draw and Assess Difficulty are negatively associated with correctness. For Nondisjunction, the processes of Eliminate, Draw, and Reason in combination are associated with correctness. For Gel/Pedigree, only the process of Reason was associated with correctness. The examples of cor-

rect student answers for each content area, as shown in Tables 2 and 3, were selected to include each of the positively associated processes described.

To identify why drawing may be detrimental for Probability and Recombination problems, we further characterized how students described their process of Draw in incorrect answers from these two content areas. We identified two categories: Inaccurate drawing and Inappropriate drawing application. Table 8 provides descriptions and student examples for each category. For Probability problems, 49% of the incorrect student answers that used Draw were Inaccurate, as they identified incorrect genotypes or probabilities while drawing a Punnett square. Thirty-one percent of the answers contained Inappropriate drawing applications such as drawing a Punnett square for each generation of a multiple-generation pedigree rather than multiplying probabilities. Five percent of the answers displayed both Inaccurate and Inappropriate drawing (Figure 2). For Recombination, 83% of incorrect student answers using Draw used an Inappropriate drawing application, typically treating linked genes as if they were unlinked by drawing a

TABLE 5. Likelihood of processes in student answers varies by content area[a]

Process	Prevalence of process (% of correct and incorrect student answers)				Predicted probability (%) from GLMM				GLMM p value (pairwise comparisons)					
	Probability	Recombination	Nondisjunction	Gel/Pedigree	Probability (P)	Recombination (R)	Nondisjunction (N)	Gel/ Pedigree (G)	P–R	P–N	P–G	R–N	R–G	N–G
Orientation														
Notice	36.32	32.89	26.22	29.02	33.21	25.87	17.77	18.46	ns	***	***	*	ns	ns
Recall	26.15	43.41	20.65	9.27	18.35	40.29	14.15	5.07	***	ns	***	***	***	***
Metacognition														
Plan	32.20	17.70	20.42	23.90	28.24	11.28	11.44	16.71	***	***	**	ns	ns	ns
Check	5.81	5.51	11.83	13.41	2.12	1.83	4.38	6.5	ns	ns	**	**	***	ns
Assess Difficulty	4.36	5.01	7.66	5.37	1.06	1.04	1.71	1.29	ns	ns	ns	ns	ns	ns
Execution														
Use Information	93.70	82.30	36.66	65.85	96.25	86.75	31.46	69.53	***	***	***	***	***	***
Integrate	20.82	8.35	14.39	45.12	17.82	5.67	9.47	43.53	***	**	***	ns	***	***
Draw	22.76	13.19	28.77	8.54	13.65	4.08	15.36	2.64	***	ns	***	***	ns	***
Calculate	77.97	76.13	0.00	0.24	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Reasoning														
Reason	94.43	84.98	84.69	90.49	97.52	94.10	92.88	96.93	*	**	ns	ns	*	**
Conclusion														
Eliminate	0.48	0.00	10.44	31.46	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
n answers	413	599	431	410										

[a]All student answers (correct and incorrect) are reported. Processes excluded from analyses include Claim, those within the Error and General strategies, processes that were present in 5% or fewer of both incorrect and correct student answers. Pairwise comparisons between: Probability (P), Recombination (R), Nondisjunction (N), and Gel/Pedigree (G). NA: no comparison made due to prevalence of 0% in at least one group. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; ns: $p > 0.05$. See Supplemental Table S2 for standard errors of coefficient estimates. Interpretation example: In Probability questions, 94.43% of answers contain Reason, while in Nondisjunction, 84.69% of answers contain Reason. Based on GLMM estimates to account for individual variability in process use, a question in the Probability content area had a 97.52% chance of using Reason, and a question in the Nondisjunction content area had a 92.88% chance of using this process.

TABLE 6. Prevalence of processes in correct and incorrect student answers by content area[a]

| | Prevalence of process (% of correct and incorrect student answers) | | | | | | | |
| | Probability | | Recombination | | Nondisjunction | | Gel/Pedigree | |
Process	Incorrect student	Correct student	Incorrect student	Correct student	Incorrect student	Correct student	Incorrect student	Correct student
Orientation								
Notice	32.81	39.37	31.21	34.94	24.07	26.93	25.22	30.43
Recall	21.88	29.86	29.70	60.22	24.07	19.50	11.71	8.36
Metacognition								
Plan	26.04	37.56	13.94	22.30	18.52	21.05	19.82	25.42
Check	4.17	7.24	5.15	5.95	10.19	12.38	9.91	14.72
Assess Difficulty	5.21	3.62	6.97	2.60	11.11	6.50	8.11	4.35
Execution								
Use Information	88.54	98.19	73.94	92.57	37.96	36.22	63.06	66.89
Integrate	18.75	22.62	10.30	5.95	5.56	17.34	41.44	46.49
Draw	28.65	17.65	21.52	2.97	12.96	34.06	9.91	8.03
Calculate	58.33	95.02	65.15	89.59	0.00	0.00	0.00	0.33
Reasoning								
Reason	90.10	98.19	79.70	91.45	75.93	87.62	81.08	93.98
Conclusion								
Eliminate	0.52	0.45	0.00	0.00	2.78	13.00	26.13	33.44
n answers	192	221	330	269	108	323	111	299

[a]All student answers (correct and incorrect) are reported. Processes excluded from analyses include Claim, those within the Error and General strategies, processes that were present in 5% or fewer of both correct and incorrect student answers.

TABLE 7. The combination of processes associated with the probability of a correct student answer varies by content area[a]

Content area	Predicted probability of correct answer (%)	Combination of processes present
Probability	7.31	Intercept
	40.19	Calculate
	58.60	Calculate + Use Information
	67.56	Calculate + Use Information + Reason
	66.40	Calculate + Use Information + Reason + Draw [negative predictor]
Recombination	6.31	Intercept
	17.57	Calculate
	38.83	Calculate + Recall
	63.02	Calculate + Recall + Use Information
	74.91	Calculate + Recall + Use Information + Reason
	76.17	Calculate + Recall + Use Information + Reason + Plan
	44.29	Calculate + Recall + Use Information + Reason + Plan + Draw [negative predictor]
	40.81	Calculate + Recall + Use Information + Reason + Plan + Draw [negative predictor] + Assess Difficulty [negative predictor]
Nondisjunction	70.48	Intercept
	86.34	Eliminate
	89.96	Eliminate + Draw
	90.22	Eliminate + Draw + Reason
Gel/Pedigree	56.95	Intercept
	71.56	Reason

[a]Based on a representative GLMM model with a lasso penalty predicting answer correctness with a moderate penalty parameter (lambda = 25). The intercept represents the likelihood of a correct answer in the absence of all processes initially entered into the model: Notice, Plan, Recall, Check, Assess Difficulty, Use Information, Integrate, Draw, Calculate, Reasoning, Eliminate. Shaded rows indicate the inclusion of negative predictors in combination with positive predictors. Probabilities were calculated
[AQ 5] using the inverse logit of the sum of the combination of log odds coefficient estimates and the intercept from Supplemental Table S3.

TABLE 8. Drawing use categorization

Categories	Description	Example
Inaccurate drawing	Drawing contains incorrect components or is incorrectly interpreted.	Student draws a Punnett square/cross and identifies the incorrect offspring probability. In the Probability Giraffe question, we know that II-6 is heterozygous, but the student answer here indicates II-6 is 2/3 likely to be heterozygous: "I look at the pedigree and try to decide what the genotypes of the parents are. I do tests crosses for Rrxrr and RR x rr for II-6. I determine that II-7s parents are both Rr since one of their kids is rr, or short necked. So both II-6 and II-7 have a 2/3 chance of being a carrier. Doing a cross for their kid, if both parents are Rr, she will have a 1/4 chance of being short necked. The total probability of the kid being short necked is (2/3)x(2/3)x(1/4), 1/9." —Incorrect student answer: Ingrid
Inappropriate drawing application	The type of drawing used was not appropriate for the concept,	Student draws multiple Punnett squares instead of using the Product Rule to take into account the probability of parent genotypes. In the Probability Dimples problem, we know that Pritya has a 2/3 likelihood of being heterozygous, but the student answer here creates two separate accurate Punnett squares to account for the uncertainty in Pritya's genotype instead of using genotype probabilities multiplied over multiple generations: "1. read the question and look at the pedigree 2. try to give genotypes to people 2a. Narayan has to be heterozygous because she has to have one recessive allele from her mother but she is not affected with the dimple phenotype, so she has one dominant allele. 2b. Pritya can be either homozygous dominant or heterozygous because her parents were heterozygotes and she has dimples. 3. drew a punnett square of the possible crosses between Pritya and Narayan (Dd x Dd and DD x Dd). 3a. if Pritya is homozygous dominant, the child will have dimples. 3b. if Pritya is heterozygous, the child has a 1/4 chance of not having dimples. 4. final answer: we have to know the genotype of Pritya before making a conclusion on the child, so 1/4 or 0."—Incorrect student answer: Isabella Student draws a Punnett square of unlinked genes for a scenario in which genes are linked. In the Recombination Aldose gene question, the distance between genes (e.g., map units) must be considered, but the student answer here does not account for linked genes: "1. Read through the question 2. Examined the cross 3. Made a punnet square crossing AaFf with aaff 4. Determined that 1/4 or 25% of the offspring are Aaff and that phenotype would only produce aldose, not fructose." —Incorrect student answer: Ivan

Punnett square to calculate probability. Ten percent of answers used both Inappropriate and Inaccurate drawing (Figure 2).

DISCUSSION

In this study, we identified and characterized the various processes that a large sample of students and experts used to document their answers to complex genetics problems. Overall, although their frequency of use differed, experts and students used the same set of problem-solving strategies. Experts were more likely to use orienting and metacognitive strategies than students, confirming prior findings on expert–novice differences (e.g., Chi *et al.*, 1981; Smith and Good, 1984; Smith, 1988; Atman *et al.*, 2007; Smith *et al.*, 2013; McDonnell and Mullally, 2016; Peffer and Ramezani, 2019). For students, we also identified which strategies were most associated with correct answers. The use of reasoning was consistently associated with correct answers across all content areas combined as well as for each individual content area. Students used other processes more or less frequently depending on the content of the question, and the combination of processes associated with correct answers also varied by content area.

Domain-Specific Problem Solving

We found that most processes students used (i.e., all but those in the General category) were domain specific, relating directly to genetics content. Prevost and Lemons (2016), who examined students' process of solving multiple-choice biology problems, found that domain-general processes were more common in answers to lower-order than higher-order questions. They also found that using more domain-specific processes was associ-

ated with correctness. In our study, students solved only higher-order problems that asked them to apply or analyze information. Students also had to construct their responses to each problem, rather than selecting from multiple predetermined answer options. These two factors may explain the prevalence of domain-specific processes in the current study, which allowed us to investigate further the types of domain-specific processes that lead to correct answers.

Metacognitive Activity: Orienting and Metacognitive Processes Are Described by Experts but Not Consistently by Students

Our results support several previous findings from the literature comparing the problem-solving tactics of experts and students: experts are more likely to describe orienting and metacognitive problem-solving strategies than students, including planning solutions, checking work, and identifying the concept of the problem.

Planning. While some students used planning in their correct answers, experts solving the same problems were more likely to do so. Prior studies of solutions to complex problems in both engineering and science contexts found that experts more often used the orienting/planning behavior of gathering appropriate information compared with novices (Atman *et al.*, 2007; Peffer and Ramezani, 2019). Experts likely have engaged in authentic scientific investigations of their own, and planning is more likely when the problem to be solved is more complex (e.g., Atman *et al.*, 2007), so experts are likely more familiar with and see value in planning ahead before pursuing a certain problem-solving approach.
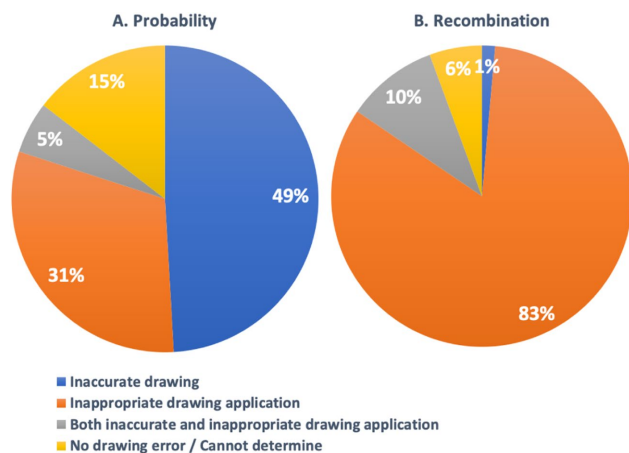
FIGURE 2. Drawing is commonly inaccurate or inappropriate in incorrect student answers for Probability and Recombination. Drawing categorization from student answers that used Draw and answered incorrectly for content areas of (A) Probability (*n* = 55) and (B) Recombination (*n* = 71). Each category is mutually exclusive, so those that have both Inaccurate drawing/Inappropriate drawing are not in the individual use categories. "No drawing error" indicates neither inaccurate nor inappropriate drawings were described. "Cannot determine" indicates not enough information was provided in the students' written answer to assign a drawing use category.

Checking. Experts were much more likely than students to describe their use of checking work, as also shown in previous work (Smith and Good, 1984; Smith, 1988; McDonnell and Mullally, 2016). McDonnell and Mullally (2016) found greater levels of unprompted checking after students experienced modeling of explicitly checking prompts and were given points for demonstrating checking. These researchers also noted that when students reviewed their work, they usually only checked some problem components, not all. Incomplete checking was associated with incorrect answers, while complete checking was associated with correct answers. In the current study, we did not assess the completeness of checking, and therefore may have missed an opportunity to correlate checking with correctness. However, if most students were generally checking their answers in a superficial way (i.e., only checking one step in the problem-solving process versus checking all steps), this could explain why there were no differences in the presence of checking between incorrect and correct student answers. In contrast to our study, Prevost and Lemons (2016) found checking was the most common domain-specific procedure used by students when answering both lower- and higher-order multiple-choice biology questions. The multiple-choice format may prompt checking, as the answers have already been provided in the scenario. In addition, while that study assessed answers to graded exam questions, we examined answers to extra-credit assignments. Thus, a lack of motivation may have influenced whether the students in the current study reported checking their answers.

Identifying the Concept of a Problem. Although this strategy was relatively uncommon even among experts, they were more likely than students to describe identifying the concept of a

problem in their solutions. This is consistent with previous research showing that nonexperts use superficial features to solve problems (Chi *et al.*, 1981; Smith and Good, 1984; Smith *et al.*, 2013), a tactic also associated with incorrect solutions (Smith and Good, 1984). The process of identifying relevant core concepts in a problem allows experts to identify the appropriate strategies and knowledge needed for any given problem (Chi *et al.*, 1981). Thus, we suggest that providing students with opportunities to recognize the core concepts of different problems, and thus the similarity of their solutions, could be beneficial for learning successful problem solving.

Engaging in Explanations: Using Reasoning Is Consistently Associated with Correct Answers

Our findings suggest that, although reasoning is frequently used by both correct and incorrect students, it is strongly associated with correct student answers across all content areas. Correct answers were more likely than incorrect answers to use reasoning; furthermore, reasoning was associated with a correct answer for each of the four content areas we explored. This supports previous work showing that reasoning ability in general is associated with overall biology performance (Cavallo, 1996; Johnson and Lawson, 1998). Students who use reasoning may be demonstrating their ability to think logically and sequentially connect ideas, essentially building an argument for why their answers make sense. In fact, teaching the skill of argumentation helps students learn to use evidence to provide a reason for a claim, as well as to rebut others' claims (Toulmin, 1958; Osborne, 2010), and can improve their performance on genetics concepts (Zohar and Nemet, 2002). Thus, the genetics students in the current study who were able to explain the rationale behind each of their problem-solving steps are likely to have built a conceptual understanding of the topic that allowed them to construct logical rationales for their answers.

In the future, think-aloud interviews should be used to more closely examine the types of reasoning students use. Students may be more motivated and better able to explain their rationales verbally, or with a combination of drawn and verbal descriptions, than they are inclined to do when typing their answers in a writing-only situation. Interviewers can also ask follow-up questions, confirming student explanations and ideas, something that cannot be obtained from written explanations. In addition, the problems used in this study were near-transfer problems, similar to those that students previously solved during class. Such problems can often be solved using an algorithmic approach, as also recently described by Frey *et al.* (2020) in chemistry. Future studies could identify whether and when students use more complex approaches such as causal reasoning (providing connections between ideas) or mechanistic reasoning (explaining the biological mechanism as part of making causal connections (Russ *et al.*, 2008; Southard *et al.*, 2016) in addition to or instead of algorithmic reasoning.

Students Use Different Processes to Answer Questions in Different Content Areas

Overall, students answered 60% of the questions correctly. Some content areas were more challenging than others: Recombination was the most difficult, followed by Probability, then Gel/Pedigree and Nondisjunction (see also Avena and Knight, 2019). While our results do not indicate that a certain combination of

processes are both necessary and sufficient to solve a problem correctly, they can be useful to instructors wishing to guide students in their strategy use when considering their solutions to certain types of problems. In the following section, we discuss the processes that were specifically associated with correctness in student answers for each content area.

Probability. Solving a Probability question requires calculation, while many other types of problems do not. To solve the questions in this study, students needed to consider multiple generations from two families to calculate the likelihood of independent events occurring by using the product rule. Smith (1988) found that both successful and unsuccessful students often find this challenging. Our previous work also found that failing to use the product rule, or using it incorrectly, was the second most common error in incorrect student answers (Avena and Knight, 2019). Correctly solving probability problems likely also requires a conceptual understanding of the reasoning behind each calculation (e.g., Deane *et al.*, 2016). This type of reasoning, specific to the mathematical components of a problem, is referred to as statistical reasoning, a suggested competency for biology students (American Association for the Advancement of Science, 2011). The code of Reason includes reasoning about other aspects of the problem (e.g., determining genotypes; see Table 3) in addition to reasoning related to calculations. While reasoning was prevalent in both incorrect and correct answers to Probability problems, using reasoning still provided an additional 9% likelihood of answering correctly for students who had also used calculating and applying information in their answers.

Generally, calculation alone was not sufficient to answer a Probability question correctly. When students applied information to solving the specific problem (captured with the Use Information code), such as determining genotypes within the pedigree or assigning a probability, their likelihood of generating a correct answer was 40%. This only increased to 59% if they also used Calculate (see Table 7). We previously found that the most common content error in these types of probability problems was mis-assigning a genotype or probability due to incorrectly using information in the pedigree; this error was commonly seen in combination with a product rule error (Avena and Knight, 2019). This correlates with our current findings on the importance of applying procedural knowledge: both Use Information and Calculate, under the AtL element of generating knowledge, contribute to correct problem-solving.

Recombination. Both the Probability and Recombination questions are fundamentally about calculating probabilities; thus, not surprisingly, Calculate is also associated with correct answers to Recombination questions. Determining map units and determining the frequency of one possible genotype among possible gametes both require calculation. Use of Recall in addition to Calculate increases the likelihood of answering correctly from 18 to 39%. This may be due to the complexity of some of the terms in these problems. As shown previously, incorrect answers to Recombination questions often fail to use map units in their solution (Avena and Knight, 2019). Appropriately using map units thus likely requires remembering that the map unit designation is representative of the probability of recombination and then applying this definition to the problem. When

students Used Information, along with Calculate and Recall, their likelihood of answering correctly increased to 63%.

Reasoning and planning also contribute to correct answers in this content area. In their solutions, students needed to consider the genotypes of the offspring and both parents to solve the problem. The multistep nature of the problem may give students the opportunity to plan their possible approaches, either at the very beginning of the problem and/or as they walk through these steps. This was seen in Preston's solution (Table 3), in which the student sequentially made a short-term plan and then immediately used information in the problem to carry out that plan.

Drawing: A Potentially Misused Strategy in Probability and Recombination Solutions. Only one process, Drawing, was negatively associated with correct answers in solutions to both Probability and Recombination questions. Drawing is generally considered beneficial in problem solving across disciplines, as it allows students to generate a representation of the problem space and/or of their thinking (e.g., Mason and Singh, 2010; Quillin and Thomas, 2015; Heideman *et al.*, 2017). However, when students generate inaccurate drawings or use a drawing methodology inappropriately, they are unlikely to reach a correct answer. In a study examining complex meiosis questions, Kindfield (1993) found that students with more incorrect responses provided drawings with features not necessary to solving the problem. In our current study, we found that the helpfulness of a drawing depends on its quality and the appropriateness or context of its use.

When answering Recombination problems, many students described drawing a Punnett square and then calculating the inheritance as if the linked genes were actually genes on separate chromosomes. In doing so, students revealed a misunderstanding of when and why to appropriately use a Punnett square as well as their lack of understanding that the frequency of recombination is connected to the frequency of gametes. Because we have also shown that planning is beneficial in solving Recombination problems, we suggest that instructors emphasize that students first plan to look for certain characteristics in a problem, such as linked versus unlinked genes, to identify how to proceed. For example, noting that genes are linked would suggest *not* using a Punnett square when solving the problem. Similarly, in Probability questions, students must realize that uncertainty in genotypes over multiple generations of a family can be resolved by multiplying probabilities together rather than by making multiple possible Punnett squares for the outcome of a single individual. These findings connect to the AtL elements of generative thinking and taking a deep approach: drawing can be a generative behavior, but students must also be thinking about the underlying context of the problem rather than a memorized fact.

Nondisjunction. In Nondisjunction problems, students were asked to predict the cause of an error in chromosome number. Our model for processes associated with correctness in nondisjunction problems (Table 7) suggested that the likelihood of answering correctly in the absence of several processes was 70%. This may explain the higher percent of correct answers in this content area (75%) compared with other content areas. Nonetheless, three processes were shown to help students

answer correctly. The process Eliminate, even though used relatively infrequently (10%), provides a benefit. Using elimination when there are a finite number of obvious solutions is a reasonable strategy, and one previously shown to be successful (Smith and Good, 1984). Ideally, this strategy would be coupled with drawing the steps of meiosis and then reasoning about which separation errors could not explain the answer. Drawing was associated with correct answers in this content area, though it was neither required nor sufficient. Instead of drawing, some students may have used a memorized series of steps in their solutions. This is referred to as an "algorithmic" explanation, in which a memorized pattern is used to solve the problem. For example, such a line of explanation may go as follows: "beginning from a diploid cell heterozygous for a certain gene, two of the same alleles being present in one gamete indicates a nondisjunction in meiosis II." Such algorithms can be applied without a conceptual understanding (Jonsson *et al.*, 2014; Nyachwaya *et al.*, 2014), and thus students may inaccurately apply them without fully understanding or being able to visualize what is occurring during a nondisjunction event (Smith and Good, 1984; Nyachwaya *et al.*, 2014). Using a drawing may help provide a basis for analytic reasoning, providing logical links between ideas and claims that are thoughtful and deliberate (Alter *et al.*, 2007). Indeed, in Kindfield's study (1993), in which participants (experts and students) were asked to complete complex meiosis questions, they found that those with more accurate models of meiosis used their drawings to assist in their reasoning process. Kindfield (1993) suggested that these drawings allowed for additional working memory space, thus supporting an accurate problem-solving process.

Gel/Pedigree. Unlike other content areas, the only process associated with correctness in the Gel/Pedigree model was Reasoning, which provided a greater contribution to correct solutions than in any other content area. In these problems, students are asked to find the most likely mode of inheritance given both a pedigree of a family and a DNA gel that shows representations of alleles for each family member. The two visuals, along with the text of the problem, provide students an opportunity to provide logical explanations at many points in the problem. Students use reasoning to support intermediate claims as they think through possible solutions, and again for their final claims, or for why they eliminate an option. Almost half of both correct and incorrect student answers to these questions integrated features from both the gel and pedigree to answer the problem. Even though many correct and incorrect answers integrate, correct answers also reason. We suggest that the presence of two visual aids prompts students to integrate information from both, thus potentially increasing the likelihood of using reasoning.

Limitations

In this study, we captured the problem-solving processes of a large sample of students by asking them to write their step-by-step processes as part of an online assignment. In so doing, we may not have captured the entirety of a student's thought process. For example, students may have felt time pressure to complete an assignment, may have experienced fatigue after answering multiple questions on the same topic, or simply may not have documented everything they were thinking. Students may also have been less likely to indicate they were engaging in

drawing, as they were answering questions using an online text platform; exploring drawing in more detail in the future would require interviews or the collection of drawings as a component of the problem-solving assignment. Additionally, students may not have felt that all the steps they engaged in were worth explaining in words; this may be particularly true for metacognitive processes. Students are not likely accustomed to expressing their metacognitive processes or admitting uncertainty or confusion during assessment situations. However, even given these limitations, we have captured some of the primary components of student thinking during problem solving.

In addition, our expert–student comparison may be biased, as experts had different reasons than students for participating in the study. The experts likely did so because they wanted to be helpful and found it interesting. Students, on the other hand, had very different motivations, such as using the problems for practice in order to perform well on the next exam and/or to get extra credit. Although it is likely not possible to put experts and students in the same affective state while they are solving problems, it is worth realizing that the frequencies of processes they use could reflect their different states while answering the questions.

Finally, the questions in the assignments provided to students were similar to those seen previously during in-class work. The low prevalence of metacognitive processes in their solutions could be due to students' perception that they have already solved similar questions. This may prevent them from articulating their plans or from checking their work. More complex, far-transfer problems would likely elicit different patterns of processes for successful problem solving.

SUGGESTIONS FOR INSTRUCTION

We have shown that successful problem solving in genetics varies depending upon the concepts presented in the problem. However, for all content areas, the general skill of explaining one's answer (Reasoning) supports students' use of declarative knowledge, increasing their likelihood of constructing correct solutions. Instructors could make a practice of suggesting certain processes to students to highlight strategies correlated with successful problem solving, along with pointing out the processes that may be detrimental. Here we provide a generalized summary of recommendations.

- Calculating: In questions regarding probability, students will need to be familiar with mathematical representations and calculations. Practicing probabilistic thinking is critical.
- Drawing: Capturing thought processes with a drawing can help visualize the problem space and can be used to generate supportive reasoning for one's thinking (e.g., a drawing of the stages of meiosis). However, a cautionary note: drawing can lead to unsuccessful problem solving when used in an inappropriate context, such as a Punnett square when considering linked genes or using multiple Punnett squares when other rules should be used, such as multiplication of probabilities from multiple generations.
- Eliminating: In questions with clear alternate final answers, eliminating answers, preferably while explaining one's reasons, is particularly useful.
- Practicing metacognition: Although there were few significant differences in metacognitive processes between correct

and incorrect student answers, we still suggest that planning and checking are valuable across content areas, as demonstrated by the more frequent use of these processes by experts.

In summary, we suggest that instructors not only emphasize key pieces of challenging content for each given topic, but also consistently demonstrate possible problem-solving strategies, provide many opportunities for students to practice thinking about how to solve problems, and encourage students to explain to themselves and others *why* each of their steps makes sense.

ACKNOWLEDGMENTS

REFERENCES

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research*, *58*(4), 375–404. https://doi.org/10.3102/00346543058004375

Alexander, P. A., Pate, P. E., Kulikowich, J. M., Farrell, D. M., & Wright, N. L. (1989). Domain-specific and strategic knowledge: Effects of training on students of differing ages or competence levels. *Learning and Individual Differences*, *1*(3), 283–325. https://doi.org/10.1016/1041-6080(89)90014-9

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*(4), 569–576. https://doi.org/10.1037/0096-3445.136.4.569

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

Angra, A., & Gardner, S. M. (2018). The graph rubric: Development of a teaching, learning, and research tool. *CBE—Life Sciences Education*, *17*(4), ar65. https://doi.org/10.1187/cbe.18-01-0007

Atman, C. J., Adams, R. S., Cardella, M. E., Turns, J., Mosborg, S., & Saleem, J. (2007). Engineering design processes: a comparison of students and expert practitioners. *Journal of Engineering Education*, *96*(4), 359–379. https://doi.org/10.1002/j.2168-9830.2007.tb00945.x

Avena, J. S., & Knight, J. K. (2019). Problem solving in genetics: Content hints can help. *CBE—Life Sciences Education*, *18*(2), ar23. https://doi.org/10.1187/cbe.18-06-0093

Bassok, M., & Novick, L. R. (2012). Problem solving. In Holyoak, K. J., & Morrison, R. G. (Eds.), *Oxford handbook of thinking and reasoning* (pp. 413–432). New York, NY: Oxford University Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bennett, S., Gotwals, A. W., & Long, T. M. (2020). Assessing students' approaches to modelling in undergraduate biology. *International Journal of Science Education*, *42*(10), 1697–1714. https://doi.org/10.1080/09500693.2020.1777343

Biggs, J. B. (1987). *Student approaches to learning and studying. Research monograph*. Hawthorn, Australia: Australian Council for Educational Research.

Bloom, B. S., Krathwohl, D. R., & Masia, B. B. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. New York, NY: David McKay.

Cavallo, A. M. L. (1996). Meaningful learning, reasoning ability, and students' understanding and problem solving of topics in genetics. *Journal of Research in Science Teaching*, *33*(6), 625–656. https://doi.org/10.1002/(SICI)1098-2736(199608)33:6<625::AID-TEA3>3.0.CO;2-Q

Cavallo, A. M. L., Potter, W. H., & Rozman, M. (2004). Gender differences in learning constructs, shifts in learning constructs, and their relationship to

course achievement in a structured inquiry, yearlong college physics course for life science majors. *School Science and Mathematics*, *104*(6), 288–300. https://doi.org/10.1111/j.1949-8594.2004.tb18000.x

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*(2), 145–182. https://doi.org/10.1207/s15516709cog1302_1

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Chin, C., & Brown, D. E. (2000). Learning in science: A comparison of deep and surface approaches. *Journal of Research in Science Teaching*, *37*(2), 109–138. https://doi.org/10.1002/(SICI)1098-2736(200002)37:2<109::AID-TEA3>3.0.CO;2-7

Deane, T., Nomme, K., Jeffery, E., Pollock, C., & Birol, G. (2016). Development of the Statistical Reasoning in Biology Concept Inventory (SRBCI). *CBE—Life Sciences Education*, *15*(1), ar5. https://doi.org/10.1187/cbe.15-06-0131

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Frey, R. F., McDaniel, M. A., Bunce, D. M., Cahill, M. J., & Perry, M. D. (2020). Using students' concept-building tendencies to better characterize average-performing student learning and problem-solving approaches in general chemistry. *CBE—Life Sciences Education*, *19*(3), ar42. https://doi.org/10.1187/cbe.19-11-0240

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, England: Cambridge University Press.

Groll, A. (2017). *glmmLasso: Variable selection for generalized linear mixed models by L1-penalized estimation*. *R package version*, *1*(1), 25.

Groll, A., & Tutz, G. (2014). Variable selection for generalized linear mixed models by L1-penalized estimation. *Statistics and Computing*, *24*(2), 137–154. https://doi.org/10.1007/s11222-012-9359-z

Hammer, D., & Berland, L. K. (2014). Confusing claims for data: A critique of common practices for presenting qualitative research on learning. *Journal of the Learning Sciences*, *23*(1), 37–46. https://doi.org/10.1080/10508406.2013.802652

Heideman, P. D., Flores, K. A., Sevier, L. M., & Trouton, K. E. (2017). Effectiveness and adoption of a drawing-to-learn study tool for recall and problem solving: Minute sketches with folded lists. *CBE—Life Sciences Education*, *16*(2), ar28. https://doi.org/10.1187/cbe.16-03-0116

Jacobs, J. E., & Paris, S. G. (1987). Children's metacognition about reading: Issues in definition, measurement, and instruction. *Educational Psychologist*, *22*(3–4), 255–278. https://doi.org/10.1080/00461520.1987.9653052

James, M. C., & Willoughby, S. (2011). Listening to student conversations during clicker questions: What you have not heard might surprise you! *American Journal of Physics*, *79*(1), 123–132. https://doi.org/10.1119/1.3488097

Johnson, M. A., & Lawson, A. E. (1998). What are the relative effects of reasoning ability and prior knowledge on biology achievement in expository and inquiry classes? *Journal of Research in Science Teaching*, *35*(1), 89–103. https://doi.org/10.1002/(SICI)1098-2736(199801)35:1<89::AID-TEA6>3.0.CO;2-J

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, *1*(2), 112–133. https://doi.org/10.1177/1558689806298224

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences*, *5*(10), 434–442. https://doi.org/10.1016/S1364-6613(00)01751-4

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences USA*, *107*(43), 18243–18250. https://doi.org/10.1073/pnas.1012933107

Jonsson, B., Norqvist, M., Liljekvist, Y., & Lithner, J. (2014). Learning mathematics through algorithmic and creative reasoning. *Journal of Mathematical Behavior*, *36*, 20–32. https://doi.org/10.1016/j.jmathb.2014.08.003

Kindfield, A. C. H. (1993). Biology diagrams: Tools to think with. *Journal of the Learning Sciences*, *3*(1), 1–36.

Knight, J. K., Wise, S. B., Rentsch, J., & Furtak, E. M. (2015). Cues matter: Learning assistants influence introductory biology student interactions during clicker-question discussions. *CBE—Life Sciences Education*, *14*(4), ar41. https://doi.org/10.1187/cbe.15-04-0093

Knight, J. K., Wise, S. B., & Southard, K. M. (2013). Understanding clicker discussions: Student reasoning and the impact of instructional cues. *CBE—Life Sciences Education*, *12*(4), 645–654. https://doi.org/10.1187/cbe.13-05-0090

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development*, *74*(5), 1245–1260. https://doi.org/10.1111/1467-8624.00605

Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, *15*(1), 11–24. https://doi.org/10.1002/tea.3660150103

Lawson, A. E. (2010). Basic inferences of scientific reasoning, argumentation, and discovery. *Science Education*, *94*(2), 336–364. https://doi.org/10.1002/sce.20357

Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Norwood, NJ: Ablex Publishing. Retrieved July 30, 2020, from http://eric.ed.gov/?id=ED362379

Mason, A., & Singh, C. (2010). Helping students learn effective problem solving strategies by reflecting with peers. *American Journal of Physics*, *78*(7), 748–754. https://doi.org/10.1119/1.3319652

McDonnell, L., & Mullally, M. (2016). Teaching students how to check their work while solving problems in genetics. *Journal of College Science Teaching*, *46*(1), 68.

Meijer, J., Veenman, M. V. J., & van Hout-Wolters, B. H. A. M. (2006). Metacognitive activities in text-studying and problem-solving: Development of a taxonomy. *Educational Research and Evaluation*, *12*(3), 209–237. https://doi.org/10.1080/13803610500479991

Mevarech, Z. R., & Amrany, C. (2008). Immediate and delayed effects of meta-cognitive instruction on regulation of cognition and mathematics achievement. *Metacognition and Learning*, *3*(2), 147–157. https://doi.org/10.1007/s11409-008-9023-3

National Research Council. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, DC: National Academies Press. https://doi.org/10.17226/13362

Nehm, R. H. (2010). Understanding undergraduates' problem-solving processes. *Journal of Microbiology & Biology Education*, *11*(2), 119–122. https://doi.org/10.1128/jmbe.v11i2.203

Nehm, R. H., & Ridgway, J. (2011). What do experts and novices "see" in evolutionary problems? *Evolution: Education and Outreach*, *4*(4), 666–679. https://doi.org/10.1007/s12052-011-0369-7

Novick, Laura R., & Catley, K. M. (2013). Reasoning about evolution's grand patterns college students' understanding of the Tree of Life. *American Educational Research Journal*, *50*(1), 138–177. https://doi.org/10.3102/0002831212448209

Novick, L. R., & Bassok, M. (2005). Problem Solving. In Holyoak, K. J., & Morrison, R. G. (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 321–349). New York. NY: Cambridge University Press.

Nyachwaya, J. M., Warfa, A.-R. M., Roehrig, G. H., & Schneider, J. L. (2014). College chemistry students' use of memorized algorithms in chemical reactions. *Chemistry Education Research and Practice*, *15*(1), 81–93. https://doi.org/10.1039/C3RP00114H

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, *328*, 463–466.

Paine, A. R., & Knight, J. K. (2020). Student behaviors and interactions influence group discussions in an introductory biology lab setting. *CBE—Life Sciences Education*, *19*(4), ar58. https://doi.org/10.1187/cbe.20-03-0054

Peffer, M. E., & Ramezani, N. (2019). Assessing epistemological beliefs of experts and novices via practices in authentic science inquiry. *International Journal of STEM Education*, *6*(1), 3. https://doi.org/10.1186/s40594-018-0157-9

Prevost, L. B., & Lemons, P. P. (2016). Step by step: Biology undergraduates' problem-solving procedures during multiple-choice assessment. *CBE—Life Sciences Education*, *15*(4), ar71. https://doi.org/10.1187/cbe.15-12-0255

Prevost, L. B., Smith, M. K., & Knight, J. K. (2016). Using student writing and lexical analysis to reveal student thinking about the role of stop codons in the central dogma. *CBE—Life Sciences Education*, *15*(4), ar65. https://doi.org/10.1187/cbe.15-12-0267

Quillin, K., & Thomas, S. (2015). Drawing-to-Learn: A framework for using drawings to promote model-based reasoning in biology. *CBE—Life Sciences Education*, *14*(1), es2. https://doi.org/10.1187/cbe.14-08-0128

Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, *92*(3), 499–525. https://doi.org/10.1002/sce.20264

Saldana, J. (2015). *The coding manual for qualitative researchers*. Los Angeles, CA: Sage.

Schen, M. (2012, March 25). Assessment of argumentation skills through individual written instruments and lab reports in introductory biology. Paper presented at: Annual Meeting of the National Association for Research in Science Teaching (Indianapolis, IN).

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, *19*(4), 460–475. https://doi.org/10.1006/ceps.1994.1033

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, *7*(4), 351–371. https://doi.org/10.1007/BF02212307

Sieke, S. A., McIntosh, B. B., Steele, M. M., & Knight, J. K. (2019). Characterizing students' ideas about the effects of a mutation in a noncoding region of DNA. *CBE—Life Sciences Education*, *18*(2), ar18. https://doi.org/10.1187/cbe.18-09-0173

Smith, J. I., Combs, E. D., Nagami, P. H., Alto, V. M., Goh, H. G., Gourdet, M. A. A., … & Tanner, K. D. (2013). Development of the biology card sorting task to measure conceptual expertise in biology. *CBE—Life Sciences Education*, *12*(4), 628–644. https://doi.org/10.1187/cbe.13-05-0096

Smith, M. K., & Knight, J. K. (2012). Using the Genetics Concept Assessment to document persistent conceptual difficulties in undergraduate genetics courses. *Genetics*, *191*, 21–32.

Smith, M. K., Wood, W. B., & Knight, J. K. (2008). The Genetics Concept Assessment: A new concept inventory for gauging student understanding of genetics. *CBE—Life Sciences Education*, *7*(4), 422–430.

Smith, M. U. (1988). Successful and unsuccessful problem solving in classical genetic pedigrees. *Journal of Research in Science Teaching*, *25*(6), 411–433. https://doi.org/10.1002/tea.3660250602

Smith, M. U., & Good, R. (1984). Problem solving and classical genetics: Successful versus unsuccessful performance. *Journal of Research in Science Teaching*, *21*(9), 895–912. https://doi.org/10.1002/tea.3660210905

Southard, K., Wince, T., Meddleton, S., & Bolger, M. S. (2016). Features of knowledge building in biology: Understanding undergraduate students' ideas about molecular mechanisms. *CBE—Life Sciences Education*, *15*(1), ar7. https://doi.org/10.1187/cbe.15-05-0114

Stanton, J. D., Neider, X. N., Gallegos, I. J., & Clark, N. C. (2015). Differences in metacognitive regulation in introductory biology students: When prompts are not enough. *CBE—Life Sciences Education*, *14*(2), ar15. https://doi.org/10.1187/cbe.14-08-0135

Sung, R.-J., Swarat, S. L., & Lo, S. M. (2020). Doing coursework without doing biology: Undergraduate students' non-conceptual strategies to problem solving. *Journal of Biological Education*, 1–13. https://doi.org/10.1080/00219266.2020.1785925

Tanner, K. D. (2012). Promoting student metacognition. *CBE—Life Sciences Education*, *11*(2), 113–120. https://doi.org/10.1187/cbe.12-03-0033

Theobald, E. (2018). Students are rarely independent: When, why, and how to use random effects in discipline-based education research. *CBE—Life Sciences Education*, *17*(3), rm2. https://doi.org/10.1187/cbe.17-12-0280

Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, *39*(1), 35–62. https://doi.org/10.1002/tea.10008