Problem Solving in Genetics: Content Hints Can Help

Jennifer S. Avena and Jennifer K. Knight*

Department of Molecular, Cellular, and Developmental Biology, University of Colorado–Boulder, Boulder, CO 80309

ABSTRACT

Problem solving is an integral part of doing science, yet it is challenging for students in many disciplines to learn. We explored student success in solving genetics problems in several genetics content areas using sets of three consecutive questions for each content area. To promote improvement, we provided students the choice to take a content-focused prompt, termed a "content hint," during either the second or third question within each content area. Overall, for students who answered the first question in a content area incorrectly, the content hints helped them solve additional content-matched problems. We also examined students' descriptions of their problem solving and found that students who improved following a hint typically used the hint content to accurately solve a problem. Students who did not improve upon receipt of the content hint demonstrated a variety of content-specific errors and omissions. Overall, ultimate success in the practice assignment (on the final question of each topic) predicted success on content-matched final exam questions, regardless of initial practice performance or initial genetics knowledge. Our findings suggest that some struggling students may have deficits in specific genetics content knowledge, which when addressed, allow the students to successfully solve challenging genetics problems.

INTRODUCTION

Problem solving has been defined in the literature as engaging in a decision-making process leading to a goal, in which the course of thought needed to solve the problem is not certain (Novick and Bassok, 2005; Bassok and Novick, 2012; National Research Council, 2012; Prevost and Lemons, 2016). Ample research shows that students have difficulty learning how to solve complex problems in many disciplines. For example, in biology and chemistry, students often omit critical information or recall information incorrectly and/or apply information incorrectly to a problem (Smith and Good, 1984; Smith, 1988; Prevost and Lemons, 2016). Furthermore, across many disciplines, researchers have found that experts use different procedural processes than nonexperts when solving problems (Chi *et al.*, 1981; Smith and Good, 1984; Smith *et al.*, 2013). While students often identify problems based on superficial features, such as the type of organism discussed in a problem, experts identify primary concepts and then link the concept with strategies on how to solve such a problem (Chi *et al.*, 1981; Smith and Good, 1984; Smith *et al.*, 2013). Experts also often check their work and problem solutions more frequently than nonexperts (Smith and Good, 1984; Smith, 1988). Given the difficulties students have in problem solving and the value of such skills to their future careers, there is clearly a need for undergraduate educators to assist students in developing problem-solving skills (American Association for the Advancement of Science, 2011; National Research Council, 2012).

Two kinds of knowledge have been described in the literature as important for solving problems: domain specific and domain general. Domain-specific knowledge is knowledge about a specific field, including the content (declarative knowledge), the procedural processes used to solve problems (procedural knowledge), and how to

apply content and process when solving problems (conditional knowledge; Alexander and Judy, 1988). Domain-general knowledge is knowledge that can be used across many contexts (Alexander and Judy, 1988; Prevost and Lemons, 2016). A third category, strategic knowledge, is defined as knowledge about problem-solving strategies that can be domain specific or domain general (Chi, 1981; Alexander and Judy, 1988). Research suggests that domain-specific knowledge is needed, but may not be sufficient, for applying strategic knowledge to solve problems (Alexander and Judy, 1988; Alexander *et al.*, 1989). Thus, helping students learn to solve problems likely requires teaching them how to activate their content knowledge, apply their knowledge to a problem, and logically think through the problem-solving procedure.

Previous research suggests that receiving help in a variety of forms, including procedure-based prompts (Mevarech and Amrany, 2008), a combination of multiple content- and procedure-based prompts (Pol *et al.*, 2008), and models (Stull *et al.*, 2012), can be beneficial to learning. Not surprisingly, accessing relevant prior knowledge has been shown to positively influence performance (Dooling and Lachman, 1971; Bransford and Johnson, 1972; Gick and Holyoak, 1980). For example, in genetics, successful problem solvers often identify similarities between problems, whereas unsuccessful problem solvers do not (Smith, 1988). Previous research also suggests that receiving procedural guidance can be beneficial to learning. In a study that asked students to examine different problems with related solutions, prompting students to consider previously reviewed problems helped most students subsequently solve a challenging problem (Gick and Holyoak, 1980). In another study, when students received guidance that included identifying similarities to other problems as well as other procedural skills, such as planning and checking their work, they were better able to solve subsequent problems than in the absence of such guidance (Mevarech and Amrany, 2008). However, although accessing prior knowledge is important, it is also important that students understand how to *apply* their prior knowledge to a given problem (Bransford and Johnson, 1972). Thus, while students may realize they need additional information to solve a problem, if they cannot make sense of this information in the context of a given problem, the information is unlikely to be useful.

In addition to knowledge, students need practice. Within the field of psychology, many studies have examined the association between practice and performance. Completing a practice test leads to better performance on a subsequent final test compared with other conditions in which students do not test themselves, such as studying or completing an unrelated or no activity (e.g., Roediger and Karpicke, 2006; Adesope *et al.*, 2017). In a meta-analysis, this effect, termed the "testing effect," was found to occur regardless of whether feedback was given and regardless of the time between the practice test and the final test (Adesope *et al.*, 2017). The benefits of practice testing on later performance can occur not only when using the same questions (retention) but also when students are asked to transfer information to nonidentical questions, including questions that require application of concepts. In one of the few studies on the testing effect using transfer questions, students who took practice tests performed better on transfer questions on a final test for both factual (i.e., a single fact in a sentence) and con-ceptual (i.e., a cohesive idea across multiple sentences) questions than those who studied but did not take practice tests (Butler, 2010). This study also found that those who performed well on their practice tests were more likely to do well than those who performed poorly on their practice tests 1 week after practice on a subsequent final test, which included conceptual questions that required application (Butler, 2010).

In the current study, we focused on whether students who are incorrectly solving a problem can apply content knowledge given to them as a prompt to correctly solve subsequent genetics problems. We address the following questions: 1) Does providing a single content-focused prompt help students answer similar questions during subsequent practice, and does this practice help on later exams? 2) When unable to apply content prompts, what content errors and omissions do students make that lead them to continue to answer incorrectly?

METHODS

Participants

We invited students enrolled in an introductory-level undergraduate genetics course for biology majors (total of 416 students in the course) at a 4-year institution during Spring 2017 to complete each of two practice assignments containing content related to course exams. The first practice assignment was taken immediately before a unit exam, and the second assignment was taken either immediately before the next unit exam or after this exam in preparation for the cumulative final exam (see Supplemental Figure S1 for timeline). Each assignment was offered online (using the survey platform Qualtrics) for up to 6 points of extra credit (650 total course points). Students received 4 points for answering the question with an explanation of their problem-solving process and an additional 2 points if they answered correctly. The practice assignments were announced in class and by email, with encouragement to complete the assignment as preparation for an upcoming exam. Students had the option to consent to have their answers used for research purposes, and all students who completed the assignment received credit regardless of their consent.

Course Performance Metrics

Students in the course were given the option to complete the Genetics Concept Assessment (GCA; Smith *et al.*, 2008) online at the beginning of the semester (within the first week of classes) for participation extra credit. The 25 GCA questions address eight of the 11 learning objectives taught in this course. Initial performance on the GCA is reported as the pretest. Students answered the same GCA questions again on the cumulative final exam, for credit, along with instructor-generated questions that also addressed the content from practice assignments along with other course content. The instructor-generated questions on the final exam comprised 15% of the student's final course grade, and the GCA questions comprised just under 8% of the student's final course grade.

Practice Assignment Content

We selected content areas known to be challenging for genetics students (Smith *et al.*, 2008; Smith and Knight, 2012) and developed sets of questions on the following five topics: calculation of the probability of inheritance across multiple generations ("probability"), prediction of the cause of an

Jill is a carrier for the autosomal recessive disease cystic fibrosis (Ff on chromosome 7). She is also a carrier for the X-linked trait muscular dystrophy (X^DX^d). These chromosomes (unreplicated) are shown below in one of Jill's normal cells.
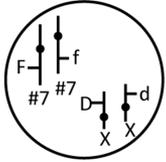
For this problem, assume no crossing over occurs during meiosis. Jill produces an egg with the genotype $F X^DX^d$, as shown below.
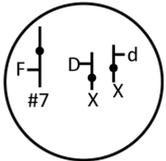
Just considering the steps of meiosis, what has occurred to produce this egg? Specify the phase of meiosis that was affected, if any.

FIGURE 1. Example of a practice question used for problem solving on the content of nondisjunction. Each question in the study had a visual aid, was constructed response, and had a single correct answer.

incorrect chromosome number after meiosis ("nondisjunction"), interpretation of a gel and pedigree to determine inheritance patterns ("gel/pedigree"), prediction of the probability of an offspring's genotype using linked genes ("recombination"), and determination of the parental germ line from which a gene is imprinted ("imprinting").

For each content area, we wrote three questions intended to be isomorphic that had the following characteristics: they addressed the same underlying concept but used different superficial characteristics, targeted higher-order cognitive processes as assessed by Bloom's level (Bloom *et al.*, 1956), contained the same amount of information, and required students to perform similar processes to solve the problem. The questions were in constructed-response format but had a single correct answer, and each question also had a coinciding visual aid (example in Figure 1; see all questions in the Supplemental Material). The questions were initially based on previously used exam questions in the course and were tested and modified through individual think-aloud interviews (16 students and seven genetics faculty) and/or a focus group (three students).

The three questions within a given content area (referred to as a "trio") were given sequentially in the practice assignments, with the first, second, and third questions referred to as "Q1," "Q2," and "Q3," respectively. For each problem-solving assignment, we randomized for each student the order of the three questions within each content area and the order in which each content area was presented. In the first problem-solving assignment, to prevent fatigue, students answered two of three randomly assigned content areas (probability, nondisjunction, and gel/pedigree), and for the second assignment, students completed questions on both recombination and imprinting.

Experimental Conditions

We developed content-focused prompts (referred to hereafter as "content hints") based on common student errors revealed during in-class questions and previous exams for this course and/or during individual student think-aloud interviews. Each hint addressed the most common student error and contained only a single content idea (Table 1). In each online practice assignment, we randomly assigned students to one of two conditions: an optional content hint when taking the second question of a content trio (hint at Q2) or an optional content hint when taking the third question of a content trio (hint at Q3). The first question (Q1) served as a baseline measure of performance for all students. At Q2, we compared the performance of students in the two conditions to determine the effect of a hint versus practice only. At Q3, we compared the performance of students within each condition with their performance on Q2 to determine whether performance was maintained (for hint at Q2 condition) or the hint improved performance compared with Q2 (for hint at Q3 condition). Using this randomized design, we could examine the differential effect of practice versus a hint, while still giving all students a chance to receive hints.

Either at Q2 or at Q3 (depending on the condition), students were asked to respond to the following question: "Do you want a hint to solve this problem (No penalty)? If so, click here." If they clicked, the hint appeared immediately below the problem, so students could see the hint while solving the problem. By asking students to select the hint rather than just showing it to everyone, we could track who chose to take a hint and thus distinguish between hint takers and non–hint takers. We did not provide real-time feedback to students, because the provided hints were intended to serve as a scaffolding mechanism without individual feedback. In addition, it would have been challenging to provide feedback, because the online platform used did not allow for personalized feedback and because the student answers were constructed response and could not be automatically graded.

Problem-Solving Content and Errors

We instructed students to explain in writing their thinking and the steps they were taking to solve the problem before they provided the final answer to each question (Prevost and Lemons, 2016). Students were not allowed to return to a question once

TABLE 1. Content hints

Content area	Hint
Recombination	The map distance between two genes provides information about the chance that recombination will occur between those two genes.
Probability	For an autosomal recessive disease, a *known non-diseased* child of two carriers has a 2/3 probability of being a carrier because you know that they are not homozygous recessive.
Gel/pedigree	Males have only one X chromosome and thus only one copy of any DNA sequences on the X chromosome.
Nondisjunction	Replicated homologous chromosomes align along the metaphase plate in meiosis I, and sister chromatids align along the metaphase plate in meiosis II.

answered. The instructions at the beginning of the assignment outlined an example of how to do this (see the Supplemental Material), and students were able to reread the instructions and an example, if desired, during the assignment. In this study, we only tracked student performance and their use of language regarding the content hint, not their thinking or problem-solving steps.

We categorized student content-specific errors and omissions and also the use of language related to the content hint. The two authors reviewed a selection of student answers to develop an initial set of codes. We then independently coded, over three iterations, the same 66 of 456 selected answers. After each iteration, we discussed our codes to come to a consensus and revised the coding scheme as needed to represent student answers. We coded an additional 19 answers to reach a final interrater agreement of 85% (Cohen's kappa of 0.83). Because we had coded and agreed upon 19% of the student answers at this point and our agreement was above acceptable levels (Landis and Koch, 1977), we then each coded half of the remaining 371 answers independently and discussed and resolved any concerns.

Statistical Analysis

We scored student answers on the practice assignments as incorrect (0) or correct (1) and used performance data only from students who provided a final answer to all possible questions in one or both assignments. We analyzed data from 233 students: 133 students completed both practice assignments, 54 students completed only the first assignment, and 46 students completed only the second assignment. Where content areas are not specified, we report results on all content areas together. We analyzed patterns at the level of the individual answer and used logistic regressions to compare answer performance between conditions, content areas, and progression groups, treating performance on one content area as independent from another content area. A student's performance within a single content area for Q1, Q2, and Q3 was treated as dependent (i.e., a repeated measure), and we used McNemar's test to analyze differences in percentage correct between questions. To examine trends at the student level, we used ordinary least-squares (OLS) regression analysis.

For the analysis of student content language use and content errors, we excluded any trios in which one answer could not be coded (i.e., no problem solving described: 36 answers) or for which there was not enough explanation to be interpretable (31 answers). A total of 342 answers are discussed in this study. We used logistic regression to compare the presence of content-specific language between differing groups within the same hint condition.

For the GCA and instructor-generated final exam questions, we report performance as percentage correct. We excluded GCA pretest scores for individuals who took less than 6 minutes to complete the online questionnaire with the GCA or did not finish at least 85% of the questions. For both the GCA and the instructor-generated final exam (a total of 150 points), a subset of questions addressed the same content areas as the practice assignment questions and are termed "practice-related" questions in this study. For the GCA, practice-related questions included one multiple-choice question per content area (questions 10, 20, 24, 25) for a total of 8 points. For the

instructor-generated final exam, there were two short-answer questions on nondisjunction and recombination and one multiple-choice question on probability, worth a total of 21 points. We also calculated performance on the remaining questions from the GCA and from the instructor-generated final exam ("practice-unrelated" questions). We used OLS regression analysis to examine the association between a student's practice assignment and exam performance, and we report unstandardized beta coefficients. We used average performance on practice Q3 questions ("practice Q3 correct"), a measure of practice success, as the predictor. We also included average performance on Q1 ("practice Q1 correct") in the regression models. For assessment performance analyses, we examined only students who completed both practice assignments (three total content areas) to ensure that all practice predictor variables were calculated based on the same number of questions (three Q3s and Q1s). Out of 133 students who completed both practice assignments, 109 students completed the GCA pre- and posttest and instructor-generated final exam and thus were included in the OLS models. The OLS regression model was the following for the GCA and instructor-generated exam questions, both practice related and practice unrelated:

$$\text{Exam performance} = \text{Practice Q3 correct}$$
$$+ \text{Practice Q1 correct} + \text{GCA pretest}$$

We also compared assessment outcomes for students who completed the GCA at both time points and the final exam but did not complete any practice assignments ($n = 35$) with those who completed all assessments and practice assignments (via OLS or independent t tests, as indicated). For this analysis, the OLS regression model was the following for the GCA and instructor-generated exam questions, both practice-related and practice-unrelated:

$$\text{Exam performance} = \text{Practice} + \text{GCA pretest}$$

We used Stata v. 15.0 and R v. 3.3.3 (dplyr, VennDiagram, statmod, VGAM, irr packages) for all statistical tests. The cutoff for statistical significance was defined as an alpha of 0.05.

Human Subjects Approval

This work was reviewed by the University of Colorado Institutional Review Board, and the use of human subjects was approved (protocols 16-0511 and 15-0380).

RESULTS

Practice Problem-Solving Performance: Question Difficulty

By randomizing the order in which students answered each question within a content area, we were able to use student performance on the first question to compare the difficulty of each of the three questions. For all content areas except imprinting, the questions were isomorphic (χ^2, $p > 0.05$), and answering the imprinting question did not influence student performance on recombination questions (taking recombination question first vs. second in the practice assignment; logistic regression, $p > 0.05$). Therefore, from this point on, all data presented represent the four remaining content areas: probability, nondisjunction, gel/pedigree, and recombination.

Two hundred thirty-three students answered a total of 553 trios of questions (Q1, Q2, Q3). The number of trios answered varies for each content area, because not all students answered all questions or completed both assignments: In the first assignment, students answered trios in two out of three content areas (randomly assigned), and in the second assignment, all students answered the trio of questions on recombination. We first examined the performance of all students across all four content areas and then for each content area individually (Table 2). For all content areas combined, student performance increased from question 1 (Q1) to questions 2 (Q2) and 3 (Q3). Upon examination of each content area individually, however, we found that the percentage of correct answers increased from Q1 to Q3 in recombination and gel/pedigree, but not for the content areas of nondisjunction and probability. In comparing Q1 performance between content areas, students had a higher percent correct for gel/pedigree and nondisjunction questions than for probability and recombination questions and a higher percent correct for probability than for recombination (Table 2).

Hint Choice

Although all students were given the option to receive a content hint for each content area during practice assignments, they only took this option in 68% of trios overall (Supplemental Table S1). Students who were offered the hint at Q2 were equally likely as those who were offered the hint at Q3 to take a hint for any given content area. For the most difficult content area (recombination), students chose to take a hint more often than for the easier content area of gel/pedigree. When looking at performance across all content areas combined, students who took a hint in a given trio scored significantly lower on all three questions than students who did not take a hint in a given trio (Supplemental Table S2). This pattern, while not always significant, was also seen in each individual content area (Supplemental Table S2). Additionally, across all content areas combined, answers did not show improvement, on average, from Q1 to Q3 in trios in which a hint was not taken, while they did in trios in which a hint was taken. This difference was also significant in the individual content area of recombination, but not the other content areas (Supplemental Table S2). To maintain reasonable sample sizes in our analyses, we combined all content areas together for the remainder of the data in this paper regarding practice performance.

We also characterized students' initial Q1 performance based on frequency of taking a hint. To best represent whether a student had a consistent pattern in hint choice, we focused on only the students who completed questions in both practice assignments (the maximum of three content areas). Of the 133 students who completed both assignments, 14 students never chose to take a hint, 56 students sometimes chose to take a hint, and 63 students always chose to take a hint when offered. Students who never took a hint performed better on Q1 than students who always took a hint (Supplemental Table S3). We have not further analyzed answer trios in which a student chose not to take a hint for several reasons. We did not have a randomization process for hint presentation: all students were given the option, and those who did not take a hint chose not to do so for reasons that we could not directly examine. In addition, because so few of the students in the study chose to never take a hint, and because we were primarily interested in the effect of taking a content hint on student success, we focused on the students who did take a hint, randomized to either Q2 or Q3 within a trio.

Content Hints Help a Subset of Students

To examine the immediate effect of a content hint on student performance, we focused the remainder of our analyses on situations in which students took a hint. We used Q1 as a baseline measure of student performance in a given content area. Because students were offered a hint either at Q2 or at Q3, we compared student performance at Q2 in the presence or absence of a hint for this question. To examine whether performance was maintained (for hint at Q2 condition) or whether the hint improved performance compared with Q2 (for hint at Q3 condition), we examined performance at Q3. For the students who took a hint, we first looked at aggregate data at the level of individual answers, binning answers into Q1 correct versus incorrect and then looking at performance on the subsequent two questions (Figure 2). As shown in Figure 2A, if students answered Q1 correctly within a trio, 15% went on to answer Q2 incorrectly (without a hint), indicating that practice itself may not help these students who initially answer correctly. Students who did receive a hint at Q2 performed the same as those who did not, indicating the drop in performance from Q1 to Q2 was not due to the hint. In a given trio, Q3 performance also did not differ based on when a hint was received, and performance, on

TABLE 2. Performance on practice problem-solving questions[a]

Content area	n	Q1 % correct	Q2 % correct	Q3 % correct
All content areas	553	56	60	60
Recombination	179	38	45	47
Probability	122	50	55	52
Gel/pedigree	127	65	76	76
Nondisjunction	125	76	71	71

[a]Performance at the level of the individual answer for 553 trios of questions (1659 questions total). The number of trios for each content area is represented by n. Practice assignment 2 included the recombination questions; practice assignment 1 included the other three content areas. Across all content areas combined, the Q2 and Q3 percent correct were significantly higher than Q1 (McNemar's test, $p < 0.05$), while there was no difference between Q2 and Q3 (McNemar's test, $p > 0.05$). For recombination and gel/pedigree, the Q3 percent correct was significantly different from Q1 (McNemar's test, $p < 0.05$), while there was no difference in pairwise comparisons between Q1 and Q2 or Q2 and Q3 (McNemar's test, $p > 0.05$). For probability and nondisjunction, there was no significant difference in percent correct among Q1, Q2, and Q3 (McNemar's test, $p > 0.05$). For Q1, gel/pedigree and nondisjunction had higher performance than probability (OR = 1.9, $p < 0.05$, and OR = 3.2, $p < 0.001$, respectively) and also higher than recombination (OR = 3.1, $p < 0.001$, and OR = 5.2, $p < 0.001$, respectively); probability had higher performance than recombination (OR = 1.6, $p < 0.05$).

A.

Answer performance when Q1 was correct
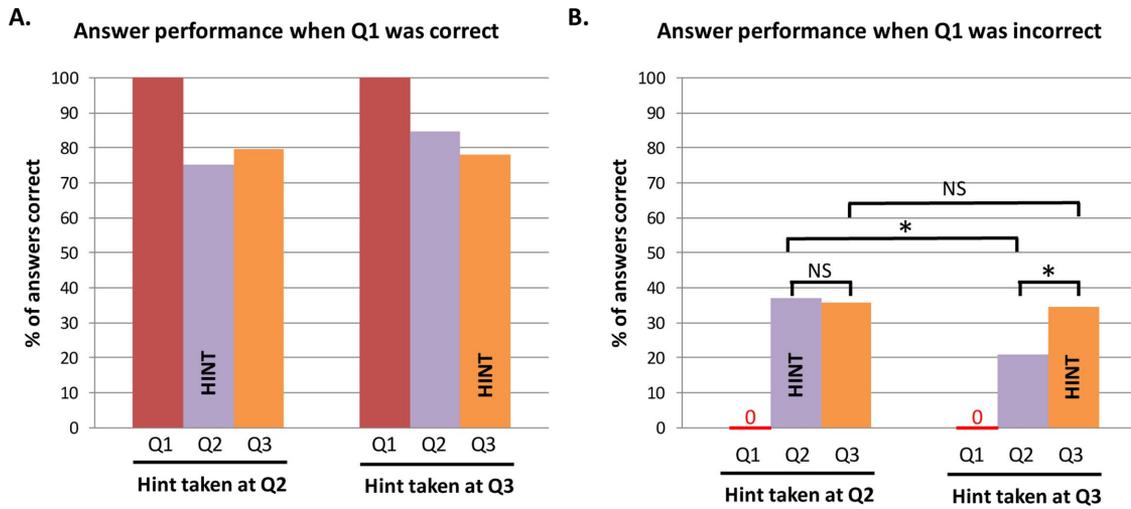

B.

Answer performance when Q1 was incorrect

FIGURE 2. The effect of a hint differs depending on Q1 correctness. (A) Q1 incorrect: the percent of correct answers for Q2 and Q3 is shown for trios in which a hint was taken at Q2 ($n = 84$ trios) or at Q3 ($n = 110$ trios). *, $p < 0.05$; all else NS, $p > 0.05$ (logistic regression between conditions; McNemar's test between Q2 and Q3 for each condition). (B) Q1 correct: the percent of correct answers for Q2 and Q3 is shown for trios in which a hint was taken at Q2 ($n = 89$ trios) or at Q3 ($n = 91$ trios). There were no significant differences between conditions (logistic regression, $p > 0.05$) or between Q2 and Q3 (McNemar's test, $p > 0.05$).

average, did not change from Q2 to Q3, indicating that a hint did not positively or negatively impact performance for these students who initially answer correctly.

If students answered Q1 incorrectly within a trio, 21% went on to answer Q2 correctly without a hint, suggesting that practice alone can help these students who initially answer incorrectly (Figure 2B). However, a significantly higher percent of students answered correctly upon receiving a hint at Q3. Students who took the hint at Q2 were significantly more likely to get Q2 correct than students who had not yet taken a hint, indicating the hint provides an added benefit beyond practice itself. A similar percent of the students who took a hint at Q2 also

answered Q3 correctly, indicating that, on average, they maintained performance on a subsequent question after the hint was taken. By the third question in a content area, all students had received a hint, some at Q2 and some at Q3. Those who took a hint at Q3 performed equivalently on Q3 to those who had taken a hint at Q2, indicating that students benefited similarly at the end of practicing a given content area, regardless of when the hint was received.

To examine how individual students performed sequentially on a trio of questions, we followed the progression of individual students from Q1 to Q3 (Figures 3 and 4). Students took a hint at Q2 in 173 trios of questions (Figure 3). Of these, 49% of
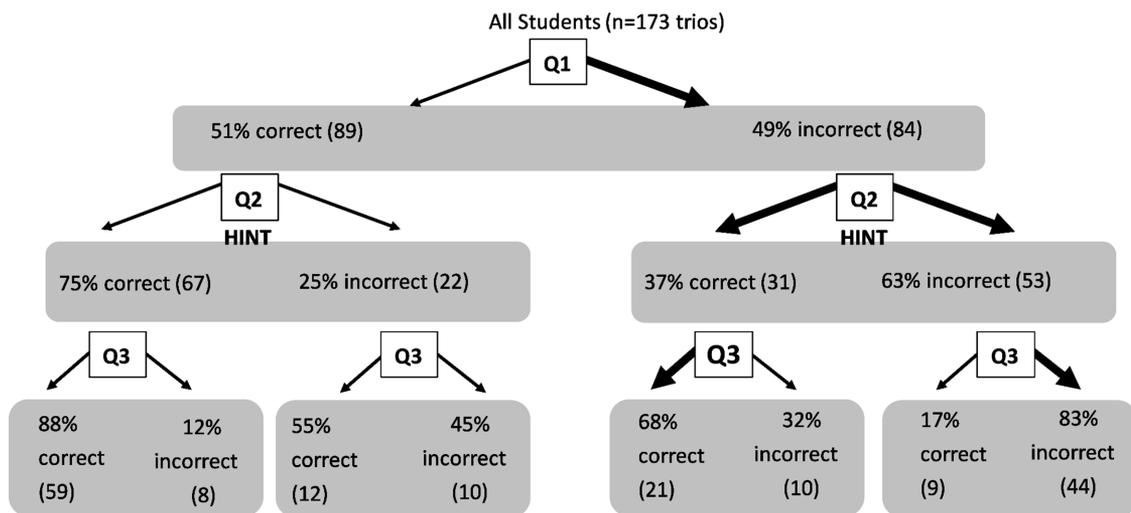

FIGURE 3. Student-level progression across answer trios in which a hint was taken at Q2. Percent of correct answers is shown with the number of answers in each category (e.g., Q1 incorrect) in parentheses. Arrows indicate the percent of answers that track to the next category. Bolded arrows signify categories of trios that were analyzed for content-specific language use and errors/omissions: trios with Q1 incorrect but Q2 and Q3 correct (011 group) and those with all three answers incorrect (000 group).
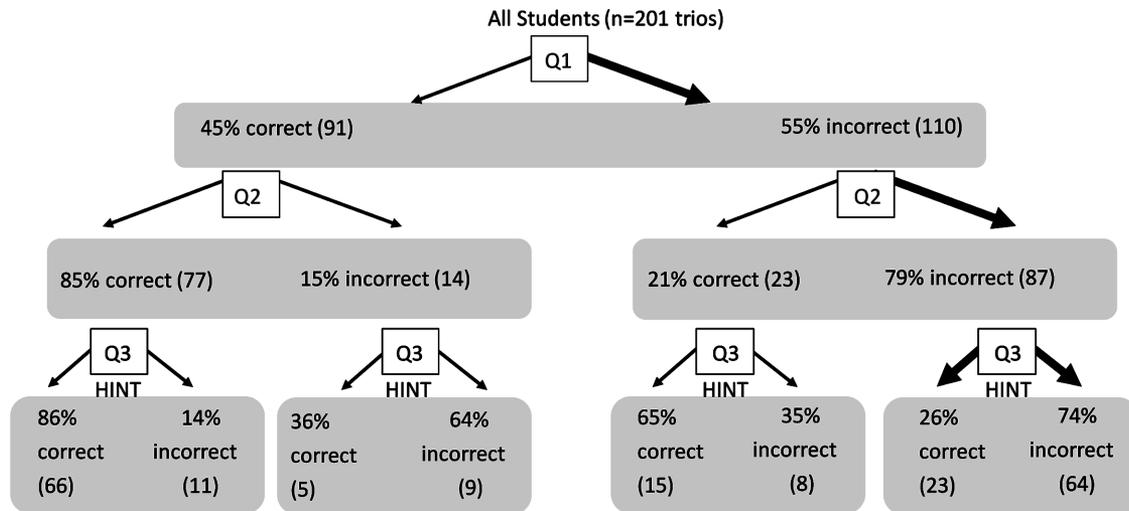
FIGURE 4. Student-level progression for answer trios in which a hint was taken at Q3. Percentage of correct answers is shown with the number of answers in each category (e.g., Q1 incorrect) in parentheses. Arrows indicate the percent of answers that track to the next category. Bolded arrows signify categories of trios that were analyzed for content-specific language use and errors/omissions: trios with Q1 and Q2 incorrect but Q3 correct (001 group) and those with all three answers incorrect (000 group).

students in a given trio answered Q1 incorrectly. Thirty-seven percent of those moved on to get Q2 correct when they received a hint, and then 68% of those went on to get Q3 correct. Thus, the majority, but not all students, maintained this improvement from Q2 to Q3. Students took a hint at Q3 in 201 trios of questions (Figure 4). Of these, 55% of students in a given trio answered Q1 incorrectly. Seventy-nine percent of those also got Q2 incorrect, and then 26% of those moved on to get Q3 correct when they received a hint. As seen in Figures 3 and 4, while a hint helped some students answer a subsequent question correctly, a hint did not help all students; some students answered Q1, Q2, and Q3 incorrectly despite taking a hint.

Content-Specific Language Use and Errors or Omissions

To further explore why the hint did not help some students but did help others, we examined *how* students used the given content hint. We categorized within a student's documented problem-solving answer 1) the presence of language that reflected the content described in the hint (coded as present or absent; Table 3), and 2) the types of content errors and omissions made in solving the problem, tracking both correctness and language use across the three questions (Q1, Q2, Q3) for each content area (Table 4). Only the following selection of students who answered Q1 incorrectly and took a hint were considered for this analysis (see bolded arrows in Figures 3 and 4): students in a given trio who answered Q2 and Q3 correctly after taking a hint at Q2 (defined as 011), those who answered correctly after taking a hint at Q3 (defined as 001), and those who answered incorrectly on all three questions (defined as 000). Students who shifted from incorrect at Q1 to correct at Q2 or Q3 (011 and 001 students, respectively) more often used language associated with the content of the hint than students who answered all three questions incorrectly. In cases in which students took a hint at Q2, 83% of answers in the 011 group contained language reflecting the hint content compared with 55% in the 000 group ($n = 40$ and 74 Q2 and Q3 answers, respectively; logistic regression, odds ratio [OR] = 3.8, $p < 0.01$). Similarly, when

students took a hint at Q3, 91% of answers in the 001 group contained language reflecting the hint content compared with 60% in the 000 group ($n = 23$ and 60 Q3 answers, respectively; logistic regression, OR = 9.2, $p < 0.01$).

Students who continued to answer incorrectly (000 group) displayed a wide variety of content-specific errors and omissions, including multiple errors or omissions within a single answer. Figure 5 shows these errors and omissions for Q1 through Q3 categorized by content area, with each error type or omission represented by different colored circles. For each content area, the orange shading represents an error or omission related to the hint content; the other colors represent different errors or omissions specific to each content area and not related to the content hint. Details for each content area for the 000 group are given in the following sections.

Recombination

In the recombination questions, the most common error in the 000 group was no use of map units to solve the problem (57% of 143 answers; Figure 5A, orange oval). In addition, students

TABLE 3. Presence of language reflecting content in hint criteria, coded only in answers during and after receipt of a hint

Content area	Student language scored as reflecting content hint if contained:
Recombination	Specific mention of map units and/or recombination percentage relevant to the problem
Probability	2/3 as the probability an individual is heterozygous (when known to be unaffected)
Gel/pedigree	Description of males as having only one X chromosome or one allele of a gene on the X chromosome
Nondisjunction	Description of chromosome alignment before separation (1) and/or timing of homologous chromosome separation (2) and/or timing of sister chromatid separation (3)

TABLE 4. Content errors and omissions codes

Content code	Description (in an answer, the student does at least one of the following listed for a given code)
Recombination	
No use of map units	Does not use map units to determine probability
Incorrect use of map units	Uses map units in an unconventional manner (e.g., multiplies map units by Mendelian probabilities or considers map units for a double homozygote)
	Does not divide map units by two to consider only one gamete Performs a calculation error using map units
Gamete type error	Indicates that the desired gamete is parental versus recombinant (or vice versa)
	Assigns the probability of a recombinant gamete to the parental gamete (or vice versa)
Desired genotype incorrect	Incorrectly states desired genotype for the answer
Probability	
Genotype/probability misassignment	Misassigns genotype and/or probability (e.g., does not assign 2/3 to an unaffected child of two heterozygote parents with an affected child)
No/improper use of product rule	Does not use the probabilities of *both* the parents and child to determine the likelihood of a child's phenotype
	Includes the probability of a child's sex when sex is already known
Inaccurate use of inheritance mode or calculations	Uses wrong mode of inheritance
	Combines ideas from multiple modes of inheritance to determine genotypes
	Uses addition instead of using the product rule
	Performs a calculation error
Gel/pedigree	
No discussion of copy number	Does not discuss number of chromosome copies in males (e.g., one X chromosome, two autosomes) and females (two X chromosomes, two autosomes)
No clear use of gel	Uses pedigree, but does not obviously use information in the gel
Incomplete inheritance mode	Mentions one or fewer components of a mode of inheritance (i.e., does not consider *both* of the following: autosomal vs. X-linked and recessive vs. dominant)
Nondisjunction	
Incorrect chromosome definition/separation rules	Reverses definition of sister chromatids and homologues
	Reverses separation rules for homologues and sister chromatids
Misunderstanding divisions	Does not consider meiotic divisions (I and II) in which error occurs
	Does not differentiate meiotic divisions I and II
	Incorrectly differentiates meiotic divisions I and II based on gamete chromosome number
General misunderstanding of meiosis	Incorrectly identifies normally expected chromosome number
	Assumes an atypical process or mutation must occur

made three other types of errors, sometimes in addition to the most common error. In some answers, while map units were used, they were used incorrectly (29%; Figure 5A, blue oval). Students also made errors in gamete-type identification in which they incorrectly assigned the type of gamete (recombinant or parental) or assigned the probability of recombination to the nonrecombinant gamete (22%; Figure 5A, green oval). Less often, students incorrectly identified the desired genotype to solve the problem (4%; Figure 5A, magenta oval). Even after receiving the hint defining map distance, many students made the most common error of not using map units to solve the problem ("No use of map units"; 49% of 67 answers), even though some of these students ($n = 12$) used the content language of the hint.

Probability

In the probability questions in this study, students needed to appropriately assign offspring having a probability of 2/3 for a certain genotype based on information about the parents and the mode of inheritance (due to one possible offspring genotype from a parental mating being eliminated). The two most common errors in the 000 group were incorrectly assigning at least one genotype or probability (which includes not using the 2/3 probability correctly; 81% of 67 answers; Figure 5B, orange circle) and not using or improperly using the product rule for

multiplying multiple independent probabilities (64%; Figure 5B, green circle). These two errors were most commonly present in combination in the same answer (40%; Figure 5B). While not as common, student answers sometimes contained the error of inaccurate use of modes of inheritance or calculations, either alone or in combination with other errors (21%; Figure 5B, blue circle). Even after receiving the hint about the 2/3 probability, many students made incorrect genotype or probability assignments ("Genotype/probability misassignment"; 70% of 33 answers), even though some of these students ($n = 5$) used the content language of the hint.

Gel/Pedigree

Gel/pedigree was one of the two higher-performing categories (the other being nondisjunction), so there are fewer answers in the 000 group. In these problems, students were asked to interpret both a gel and pedigree to determine inheritance patterns. To most accurately answer the gel/pedigree questions, examination of the molecular gel information to inform the number of chromosome copies present was needed. The omission of not discussing the number of alleles per gene in males and females was most common (91% of 23 answers; Figure 5C, orange circle), and while only a few answers contained this single omission, many answers contained this omission in addition to other errors/omissions of not clearly using the provided gel (57% total;
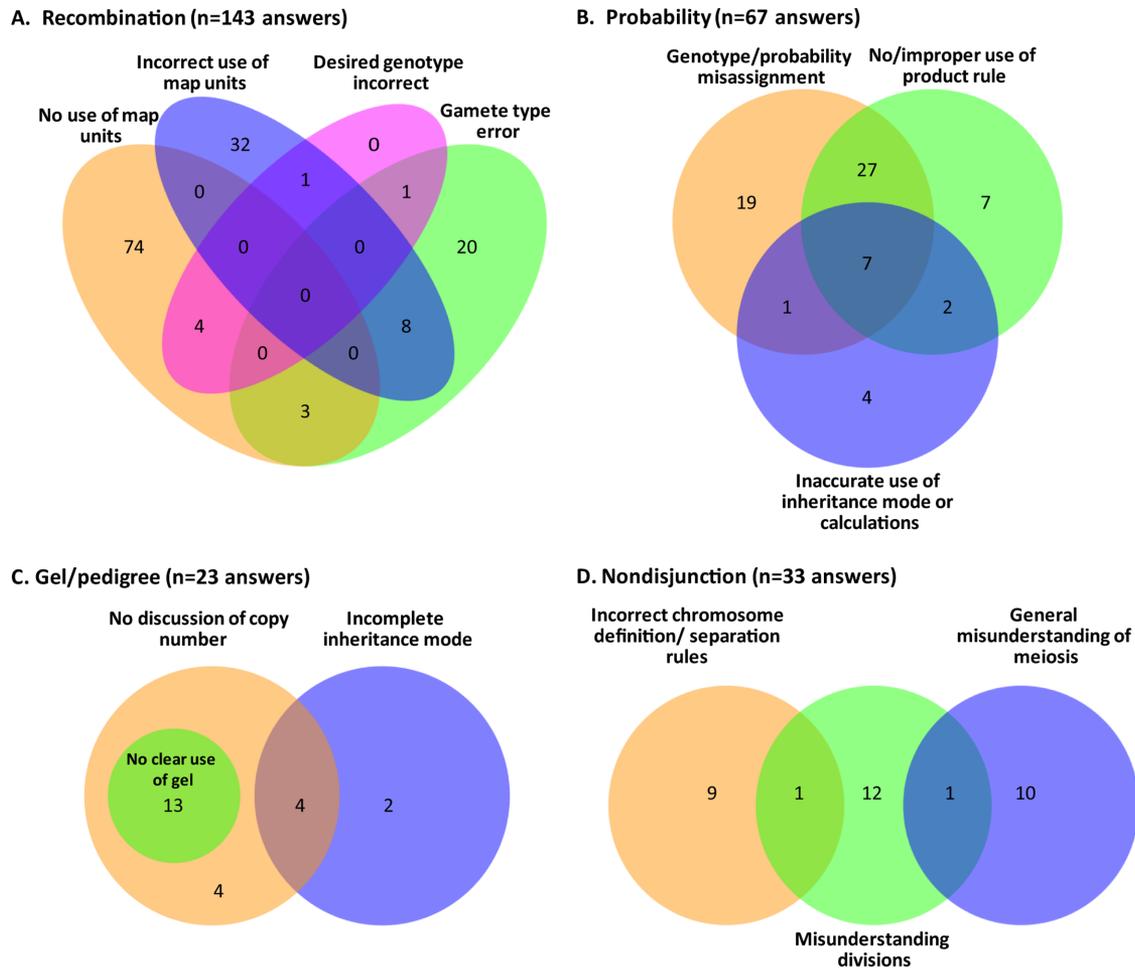
FIGURE 5. Presence of content errors and omissions in incorrect answers in four critical content areas in genetics. The number of answers in which each content error/omission code was observed is shown, with overlap in color indicating the presence of multiple errors/omissions within a single answer. Only 000 progression groups are shown for all questions Q1–Q3. In each case, orange shading indicates an error aligned with the hint content.

Figure 5C, green circle) and incompletely defining a mode of inheritance (26% total; Figure 5C, blue circle). Even after receiving the hint about X chromosome allele number, many students made the most common omission of not discussing the number of alleles per gene in males and females ("No discussion of copy number"; 88% of 8 answers), and none of these students used the content language of the hint.

Nondisjunction
In the nondisjunction problems, students were asked to identify the cause of an incorrect chromosome number after meiosis. Three errors in understanding of meiosis were present at similar levels in answers in the 000 group, including students not accurately describing homologues versus sister chromatids and/or in what phase they separated at the metaphase plate (30% of 33 answers; Figure 5D, orange circle), students not sufficiently understanding that phases in meiosis (I or II) should be considered and differentiated (42%; Figure 5D, green circle), and students not understanding the typical outcome of meiosis or how errors could occur (33%; Figure 5D, blue circle). After receiving the hint describing chromosome alignment during meiosis, several students still made the error of not accurately describing

homologues versus sister chromatids and/or in what phases they separated in meiosis ("Incorrect chromosome definition/separation rules"; 38% of 13 answers), even though some of these students (n = 3) used the content language of the hint.

Practice Is Associated with Higher Longer-Term Assessment Performance
In addition to the immediate impact of a hint on student performance during a practice assignment, we also examined whether practice itself was associated with longer-term performance on a final exam. Of the 233 students who completed practice assignments, 133 completed both assignments, and 100 completed only one assignment. To ensure that all practice predictor variables were calculated based on the same number of questions (three Q1s and Q3s), we focused on only the students who completed both practice assignments. Of the 133 students who completed both assignments, 109 of these students completed the GCA pre- and posttest and instructor-generated final exam: These are the students included in the final analyses reported in Table 5 and Supplemental Tables S4 and S5. Using the mean performance on Q3 practice questions as a measure of "success" in the practice assignments (Supplemental Table S4), we found

TABLE 5. OLS regression estimates of the association between practice performance and final exam performance[a]

Model	Practice-related β (SE)		Practice-unrelated β (SE)	
	GCA posttest	Instructor-generated final exam	GCA posttest	Instructor-generated final exam
	1	2	3	4
Practice Q3 correct (%)	0.24** (0.08)	0.29*** (0.07)	0.16*** (0.04)	0.16** (0.05)
Practice Q1 correct (%)	0.16 (0.08)	0.06 (0.07)	0.12** (0.04)	0.09 (0.05)
GCA pretest (%)	0.32* (0.16)	0.30* (0.13)	0.28*** (0.07)	0.25* (0.10)
Adjusted R^2	0.2697	0.3077	0.4878	0.2767

[a]Performance is reported as percent correct. Data reported show only students who completed both practice assignments and all assessments included in analyses ($n = 109$). Standard error of β (unstandardized beta coefficient) is shown in parentheses. Interpretation example (Practice Q3 correct, model 2): When controlling for practice Q1 performance and GCA pretest, there is a higher final exam performance of 0.29% for each 1% increase in Q3 performance.
*$p < 0.05$.
**$p < 0.01$.
***$p < 0.001$.

that, for students who completed both practice assignments, success in practice significantly predicted both GCA posttest and instructor-generated question performance for practice-related questions (controlling for mean Q1 performance and GCA pretest performance; Table 5, models 1 and 2). These students also had significantly higher scores on practice-unrelated GCA posttest and instructor-generated questions (Table 5, models 3 and 4).

Finally, we examined whether there was a difference in final exam performance between students who did not complete any practice assignments and those who completed both assignments. There were 35 students who did not complete any practice assignments but did complete the GCA pre- and posttest and instructor-generated final exam. We used GCA pretest scores to control for potential differences in incoming genetics knowledge between the group of students who completed both practice assignments and those who completed none, although we could not control for other factors, such as motivation or interest. There was no significant difference in the GCA pretest scores between these two groups (Supplemental Table S4), but students who completed the practice questions had higher GCA posttest and instructor-generated final exam scores than students who did not practice (Supplemental Table S5).

DISCUSSION
Content Hints Help a Subset of Students during Problem-Solving Practice
We administered genetics practice problems to students on concepts that had already been presented and practiced in class. Overall, we found that some students benefit from this practice, in particular if they initially answer incorrectly. Owing to the design of our study, each student completed at least one question (Q1) within a content area without any assistance. Students then received a hint on one of the subsequent questions. This provided students with the opportunity to struggle through the first question for each concept on their own before receiving assistance. An initial struggle without assistance, followed by feedback, has been shown to help students' future performance (Kapur and Bielaczyc, 2012), and although we did not provide feedback to students about whether they were correct or incorrect in their initial answers, we gave all students a chance to receive scaffolding via a content hint. For students who had initially answered Q1 incorrectly, when they took a content hint while answering Q2, 37% answered correctly, while only 21% of students answered this question correctly if they did not take

a hint at Q2. This difference of 16% indicates that, although practice alone can help, practice with content scaffolding helps more students. In addition, we have demonstrated that students benefit from a content hint regardless of whether they receive that hint at the second question or at the third question. This suggests that students who are learning from the hint at Q2 are able to apply this knowledge in answering the next question. Once they receive a key piece of content, the students who use the hint successfully continue to do so on future problems.

Hint Choice
Not all students in this study chose to take an offered hint when solving practice problems. Students who did not take a hint for a particular trio had a higher Q1 score than students who did take a hint. Along with these baseline differences in performance, several possible factors could have influenced students' choices. One component of student choice could relate to self-regulatory capacity in monitoring their understanding (Aleven *et al.*, 2003). Students who did not take a hint may have felt confident in their problem-solving ability and thus chose not to view additional information they felt they already knew. In a study that examined students' use of three-dimensional molecular models to assist in drawing molecular representations, some students did not use models even when the models were placed directly into their hands (Stull *et al.*, 2012). Some of these students reported thinking they did not need the models to answer the given questions (Stull *et al.*, 2012). This supports the idea that students who do not use provided hints may simply feel they do not need them. On the other hand, 29% of the students in our study who did not take a hint answered the first question incorrectly, indicating their confidence was misplaced. Similarly, in a study that offered computer-tailored hints for solving problems, even students predicted to benefit from hints did not always take them (Aleven *et al.*, 2006). In the current study, due to the constructed-response nature of the questions, students could not receive immediate feedback on whether they correctly answered a question. Thus, there would be value in examining whether immediate feedback on performance would influence students' future choices. Because we could not discover students' rationales for not taking a hint in this study, we cannot make any further conclusions about their choices.

Utility of a Single Content Idea
We showed that the inclusion of just one content idea as a hint helped some initially struggling students understand a concept,

potentially by activating their prior knowledge related to the hint content. In looking at these students' problem solving, we found that students who improved in a given content trio (011 and 001 groups) more often used language similar to the content of the hint than students who consistently answered incorrectly in a given trio (000 group). Thus, for students helped by the hint, this particular piece of content likely was critical for correctly solving the problem. Adding to previous frameworks (Alexander and Judy, 1988; Alexander *et al.*, 1989), we suggest that this declarative (content) knowledge is the component of domain-specific knowledge that is needed to effectively apply procedural (e.g., strategic) knowledge to accurately solve a problem. In future studies, we plan to further explore the details of students' procedural processes during problem solving and to determine whether a student's inability to recall a piece of information is the main reason for an incorrect answer or whether there are additional higher-order cognitive skills and processes required for correct problem solving.

Some students continued to answer all questions in a content trio incorrectly (000 group) despite a content hint. These students often had multiple gaps in content knowledge or made content errors or omissions not related to the content hint. In future studies, students could receive tailored content hint(s) to match all errors that are present; this could allow us to determine whether the lack of content is the reason for incorrect answers, rather than a lack of procedural process skills. In one previous study, a computer program for solving problems that provides tailored hints and feedback was used to specifically assist in genetics problem solving, providing up to four hints specific to each component of a given problem (the Genetics Cognitive Tutor; Corbett *et al.*, 2010). The authors found a significant improvement in learning from pre- to postcompletion of this program (Corbett *et al.*, 2010).

In cases in which students consistently answered incorrectly (000 group), some used language related to the content hint but made errors when trying to apply the hint in their explanations. If students have inaccurate knowledge on how to apply content, even when correct content ideas are provided, a hint may be insufficient. Indeed, Smith (1988) found that unsuccessful problem solvers can often identify important pieces of information but do not know how to apply this information. In this case, providing more scaffolding to a student, such as by providing students with worked examples of similar problems (e.g., Sweller and Cooper, 1985; Renkl and Atkinson, 2010) or providing more guided hints and feedback via a cognitive tutor (e.g., Corbett *et al.*, 2010), may be needed.

These students who consistently answer incorrectly may also be lacking critical problem-solving skills. In this study, we focused on the use and application of content knowledge, but in future studies, we will examine the problem-solving processes taken by students who answer correctly and compare these with the processes used by students who answer incorrectly. Certain skills may be particularly critical, such as displaying metacognitive ability (the knowledge and regulation of one's own cognition). Activating prior knowledge by identifying similarities between problems is an effective metacognitive skill to help orient oneself to a problem (Gick and Holyoak, 1980; Smith, 1988; Meijer *et al.*, 2006), and using this behavior in combination with several other metacognitive skills, including planning and checking work, can improve problem-solving abil-

ity (Mevarech and Amrany, 2008). Thus, a prompt that asks students to explain how the content in a hint is related to information the student has used previously to solve a problem may be helpful, as it may elicit their prior knowledge of solving similar problems.

Content-Specific Errors and Omissions

Recombination. For the topic of recombination, students who answered consistently incorrectly (000 group) did not often use map units to determine the probability of offspring when considering two linked genes; instead, many students attempted to solve the problem using Punnett squares and/or the logic of solving a probability question for genes on different chromosomes. Even when students used map units, they often either performed incorrect calculations or assigned recombinant probabilities to the incorrect genotypes. This suggests that the conceptual idea behind calculating probability of inheritance using linked genes is challenging.

Probability. Students struggled in calculating the probability that an unaffected child of two heterozygotes would be a heterozygote. Instead of considering information in the pedigree that would allow them to eliminate one of the genotype possibilities (homozygous recessive), students often assumed that the probability of a heterozygote offspring of carriers would be 1/2 rather than 2/3. For students who answered these questions consistently incorrectly (000 group), the most common error included the combination of not using the probability of 2/3 with failing to use the product rule appropriately to account for multiple generations. This suggests that struggling students do not understand the broader concept of how to consider multiple generations when determining probability and thus have difficulty integrating multiple ideas into their solutions. Indeed, previous work has shown that many students have difficulty in using both of these types of calculations (Smith, 1988; Smith and Knight, 2012).

Gel/pedigree. Students who answered consistently incorrectly (000 group) most frequently displayed difficulty in reading the gel to identify the number of allele copies and then connecting this information to the pedigree. In this course, students were taught that, although gels are not always quantitative, one can use the thickness of bands on a DNA gel to determine the relative amounts of DNA present in a sample. Despite being taught this convention, students still had difficulty applying the concept of both allele number (e.g., only one X chromosome allele for a male) and amount of DNA (e.g., a thicker band representing two of the same alleles for an individual). Thus, students need more practice understanding the concept of interpreting information on gels.

Nondisjunction. In nondisjunction questions, students who consistently answered incorrectly (000 group) had a diversity of misunderstandings about meiosis, with three errors being most common. The nondisjunction questions explicitly asked students to specify a phase in meiosis, if any, that was affected. However, students often failed to consider in which meiotic division, I or II, an error could occur, or they expressed uncertainty about differentiating between the two phases of meiosis. Students also struggled with identifying when during meiosis

homologous versus sister chromatids separate; they sometimes attempted to identify the type of chromosome that was failing to separate or to state when each would normally separate, but they were often incorrect. The third error students made represented a general misunderstanding of meiosis in which students incorrectly identified the number of each chromosome that should be present in a gamete, or students assumed an atypical event, such as multiple rounds of replication, must have occurred to produce a gamete with one extra chromosome. Previous work on this topic also found that students demonstrate many errors when depicting meiosis, including incorrect chromosome alignment during metaphase (Wright and Newman, 2011; Newman *et al.*, 2012).

Practice Is Associated with Higher Longer-Term Assessment Performance

As with previous studies that report on the testing effect (e.g., Adesope *et al.*, 2017; Butler, 2010), we found that practice was associated with later assessment performance. Regardless of practice Q1 performance and GCA pretest performance, student success in practice predicted students' longer-term performance, both practice related and practice unrelated, on their instructor-generated final exam and GCA scores in a course. We also showed that those students who completed both practice assignments performed better than students who did not complete any practice assignments, controlling for GCA pretest performance. Because we could not randomize students into practice or no-practice conditions, we caution that, even though we used the GCA pretest as a proxy for incoming ability, there are likely many other factors influencing these students' performance. Other factors shown to relate to success include student motivation, interest, and metacognition (e.g., Pintrich and de Groot, 1990; Schiefele *et al.*, 1992; Young and Fry, 2008).

Limitations

Our study addressed four critical content areas in genetics with which we know students struggle. However, students likely have additional or different difficulties on other genetics content. In addition, these questions had only a single correct answer and thus may have been limited in their ability to test student problem-solving skills. In the future, we would like to examine more ill-defined questions with multiple possible solutions (National Research Council, 2012).

While we anticipated that most students would take the option to receive a hint, only 68% of students did so. To provide an accurate representation of the influence of a hint, we had to limit our analyses to those who chose to take a hint. As seen in Stull and colleagues' (2012) work on molecular model use and as suggested by our data examining use of the content language reflected in the hint, not all students are likely to use hints, even when hints are easily available. However, it would be interesting to know why students who choose not to take a hint make that decision and whether this decision is based on high confidence or fear that the hint may confuse them.

We also could not test directly whether students who took a hint performed better than those who did not take a hint in longer-term performance, as the only way to measure this is to randomize the students who do and do not receive a hint. We chose not to take this approach, because we felt it was important for student success to give everyone the same access to information.

Implications for Instruction

This study suggests that, after learning a topic in class, a subset of students who initially give incorrect answers to problems on these topics can improve after receiving a single content idea that may fill a knowledge gap. Some students may generally understand how to solve these problems but lack one or two pieces of information; providing the missing piece allows them to apply their knowledge and solve the problem. For these students, reviewing certain pieces of genetics content, which we describe in this study, may be enough to help them solve such problems correctly. Furthermore, we suggest emphasizing the importance of practicing, as this study showed that success at the end of practice predicts longer-term performance in a class, regardless of initial understanding of genetics topics. Even if a student initially struggles with an answer, this "productive failure" can be beneficial to the student's learning (Kapur and Bielaczyc, 2012). Students who continue to struggle despite content hints likely lack content knowledge as well as problem-solving skills. We plan to further examine such deficits in how students solve problems in order to provide suggestions that are focused on the logical steps and metacognitive processes necessary for solving problems. Such instruction may be most beneficial after students have an initial chance to practice problems, so that they have a chance to challenge themselves before receiving hints.

REFERENCES

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659–701. https://doi.org/10.3102/0034654316689306

Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help-seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education*, 16, 101–130.

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, 73(3), 277–320. https://doi.org/10.3102/00346543073003277

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

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

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: Oxford University Press.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. M., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York: David McKay.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of*

Verbal Learning and Verbal Behavior, *11*(6), 717–726. https://doi.org/10.1016/S0022-5371(72)80006-9

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *36*(5), 1118–1133. https://doi.org/10.1037/a0019902

Chi, M. T. H. (1981). Knowledge development and memory performance. In Friedman, M. P., Das, J. P., & O'Connor, N. (Eds.), *Intelligence and learning* (pp. 221–229). Boston: Springer US. https://doi.org/10.1007/978-1-4684-1083-9_20

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

Corbett, A., Kauffman, L., Maclaren, B., Wagner, A., & Jones, E. (2010). A Cognitive Tutor for genetics problem solving: Learning gains and student modeling. *Journal of Educational Computing Research*, *42*(2), 219–239.

Dooling, D. J., & Lachman, R. (1971). Effects of comprehension on retention of prose. *Journal of Experimental Psychology*, *88*, 216.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*(3), 306–355. https://doi.org/10.1016/0010-0285(80)90013-4

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, *21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

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

Newman, D. L., Catavero, C. M., & Wright, L. K. (2012). Students fail to transfer knowledge of chromosome structure to topics pertaining to cell division. *CBE—Life Sciences Education*, *11*(4), 425–436. https://doi.org/10.1187/cbe.12-01-0003

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: Cambridge University Press.

Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40. https://doi.org/10.1037/0022-0663.82.1.33

Pol, H. J., Harskamp, E. G., Suhre, C. J. M., & Goedhart, M. J. (2008). The effect of hints and model answers in a student-controlled problem-solving program for secondary physics education. *Journal of Science Education and Technology*, *17*(4), 410–425. https://doi.org/10.1007/s10956-008-9110-x

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

Renkl, A., & Atkinson, R. K. (2010). Learning from worked-out examples and problem solving. In Plass, J. L., Moreno, R., & Brünken, R. (Eds.), *Cognitive load theory* (pp. 91–108). New York: Cambridge University Press.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In Renninger, K. A., Hidi, S., & Krapp, A. (Eds.), *The role of interest in learning and development* (pp. 183–212). Hillsdale, NJ: Erlbaum.

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*(1), 21–32. https://doi.org/10.1534/genetics.111.137810

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

Stull, A. T., Hegarty, M., Dixon, B., & Stieff, M. (2012). Representational translation with concrete models in organic chemistry. *Cognition and Instruction*, *30*(4), 404–434. https://doi.org/10.1080/07370008.2012.719956

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2*(1), 59–89. https://doi.org/10.1207/s1532690xci0201_3

Wright, L. K., & Newman, D. L. (2011). An interactive modeling lesson increases students' understanding of ploidy during meiosis. *Biochemistry and Molecular Biology Education*, *39*(5), 344–351. https://doi.org/10.1002/bmb.20523

Young, A., & Fry, J. D. (2008). Metacognitive awareness and academic achievement in college students. *Journal of the Scholarship of Teaching and Learning*, *8*(2), 1–10.