Investigation of Student Reasoning as a Significant Predictor of Academic Success in Introductory Genetics

Austin Hammermeister Suger

Knight Lab

Department of Molecular, Cellular & Developmental Biology

University of Colorado Boulder

Defense Date: April 3, 2020

Thesis Advisor: Jennifer Knight, Ph.D.
Department of Molecular, Cellular & Developmental Biology

Defense Committee:

Jennifer Knight, Ph.D. Department of Molecular, Cellular & Developmental Biology

Pamela Harvey, Ph.D. Department of Molecular, Cellular & Developmental Biology

Rolf Norgaard, Ph.D. Program for Writing and Rhetoric

Abstract

Making connections between content knowledge and given information is a critical part of solving complex problems. This study explores the influence of a variety of factors on student academic success in an introductory genetics course using statistical modeling. A student's use of logical reasoning while solving genetics problems, completion of practice problems, and performance on a genetics content knowledge pre-test were all found to be significantly predictive of the student's academic success on the course final examination in linear regression modeling. Data were collected from undergraduate biology students (N = 230) enrolled in an introductory genetics course at the University of Colorado Boulder. During the semester, students answered nine open-response questions in which they were asked to self-document their problem-solving processes. These processes were coded for specific components of problem solving established in prior studies. Using linear regression modeling, it was found that students who frequently made logical connections between information in a problem and their process of reaching an answer (i.e., reasoning) had better final examination performance on average than other students. These "high reasoning" students had on average higher performance on both mid-semester and final examination than students who reasoned less frequently. The findings of this study suggest that reasoning behaviors during problem solving have impacts independent of student practice and prior genetics knowledge on student academic performance. This study also suggests that promoting reasoning behaviors in genetics students could benefit their problem-solving abilities and success in introductory genetics courses.

Introduction

The importance of conducting biology education research is emphasized by the popularity of biological science fields as undergraduate majors. The National Science Board reported in 2017 that 121,742 bachelor's degrees were awarded in biological science fields in the United States (National Science Board, 2019). This makes U.S. undergraduates in biological sciences the single largest population of science majors. Though not all students who graduate with a degree in the biological sciences will enter into a related profession, they likely still encounter challenging problems where their education in the field can be of use, such as making personal medical decisions. Therefore, it is important to help biology students develop skills that will make them competent problem solvers in their specific fields and their lives. Researchers have identified methods that connect a knowledge base to a problem-solving task, such as providing students immediate feedback on their problem solving, as some of the most successful interventions to improve a student's problem-solving performance (Taconis, Ferguson-Hessler, & Broekkamp, 2001). The further development of methods to improve problem solving in undergraduate biology takes into consideration the work done in the field of physics education research (PER) over the past 40 years. As suggested by (Hoskinson, Caballero, & Knight, 2013) scholars in the field of biology education research (BER) may be able to adopt aspects of research-supported approaches because both fields deal with problem-solving that involves making predictions about complex systems.

Researchers in the field of PER have long been interested in the differences in the ways experts and novices solve physics problems. They have focused on how to improve instruction in physics courses so that students can succeed in solving complex physics problems (Redish & Steinberg, 1999). Much of this work began in an effort to determine the differences between experienced and inexperienced problem-solvers in physics. Assessing how expert and novice problem solvers represent physics problems using their domain-specific knowledge has revealed differences in the ways an individual applies knowledge to the problem at hand (Chi, Feltovich, & Glaser, 1981). Researchers determined that while experts abstracted the principles of the problems, novices based their approaches on the literal details in the problems. These types of investigations led others to work on describing scientific methods to approach the improvement of teaching in science (Reif, 1986). They proposed that the problem-solving process to reach an answer could often reveal far more about the student than the answer itself.

Informed by their colleagues in PER, contemporaries in the field of BER worked to characterize the domain-specific processes utilized by students and experts in the biological sciences. A pair of researchers investigated the differences between expert and novice approaches to solving problems in classical genetics (Smith & Good, 1984). They also supported the view that an individual's problem-solving process may be more important than their answer to the specific problem. The studies in both fields found that many of the differences between novice and expert problem-solving came down to how the experts perceived and represented the problem. The BER group compared undergraduates (novices) to graduate students and instructors (experts) as they solved genetics problems out loud. They discussed that problem-solving expertise appears to exist as a continuum that is dependent on a variety of factors. One of their key results was that successful problem-solvers understood that the solution could be found through logical connections of the information provided in the problem. Another key result was that unsuccessful problem-solvers did not recognize that problem solving requires more than the memorization of concepts related to the problem. While an understanding of the concepts underlying problems is important, so too are the skills needed to connect those specific concepts to information and evidence provided by a problem (i.e., reasoning skills).

Researchers became particularly interested in studying how student reasoning was involved in how students understood and solved genetics problems (Cavallo, 1996). They found that reasoning ability predicts student achievement in solving genetics problems. Other researchers have identified that student reasoning plays an important role in student discussions during in-class problem-solving exercises (Knight et al., 2013). They focused on how students used reasoning during group discussions in an upper-division developmental biology course. The data suggested that during discussions of clicker questions, students discussed their answers by sharing reasoning and evidence for their ideas. When the instructor cued students to use more reasoning, the researchers observed students using and sharing more reasoning.

In order to gauge student understanding of core concepts in introductory genetics courses a genetics concept assessment (GCA) was developed (Smith, Wood, & Knight, 2008). The assessment was designed to be taken by students as a pre-test and then as a post-test in order to measure student learning gains. The GCA consists of multiple-choice questions covering a range of fundamental topics in genetics. This assessment has allowed researchers to better understand the concepts with which students struggle the most in introductory genetics courses (Smith & Knight, 2012). It also allows researchers and instructors to identify student learning in genetics concepts which are particularly challenging for students. This information is useful when designing studies around student problem solving on specific topics in genetics.

More recent studies have focused on methodologies to capture a student's written step-by-step descriptions of their problem-solving process as they solve a multiple-choice biology problem (Prevost & Lemons, 2016). The researchers were able to examine the students' written descriptions of their problem-solving to characterize specific procedures used by the students. They found that domain-specific problem-solving procedures were associated with student success in multiple-choice problem-solving. Their data also showed that students used more domain-specific procedures that they categorized at a higher level on Bloom's taxonomy. One of the limitations they acknowledged is that multiple-choice problems do not always bring out all the aspects of critical thinking used by biologists.

There has been growing interest in using statistical models to better understand how student reasoning along with measures of academic preparedness predict success in undergraduate introductory biology courses. A recent study showed relationships between a student's scientific reasoning ability, ACT math score and their performance in an introductory biology course using logistic regression modeling (Thompson et al., 2018). The authors acknowledged that one of their major limitations was that they did not collect other types of student data to generate a holistic model of factors contributing to student success. The present study aimed to incorporate student demographic and academic data into the statistical models in order to generate a more complete understanding of how student reasoning and other factors predict academic success. This study examines student success in solving genetics problems through statistical modeling of student self-documented reasoning frequency in single answer constructed response questions. The study seeks to characterize the relationships between academic success in genetics and the use of reasoning during problem solving for individual genetics students. This research provides genetics instructors with evidence that student reasoning plays an important role in student academic success in genetics. If instructors had a better understanding of the academic benefits students gain when applying reasoning to problems, they might be more likely to emphasize reasoning in the problem-solving process.

Methods

Data Collection

At the beginning of the course, the students were given an optional survey containing a pre-test on genetics concepts (the Genetics Concept Assessment, GCA), a demographics survey, and a consent form for the research project. Students who consented to participate in the study agreed to the following data being collected for use in the research project: pre-GCA scores, demographic information (Table 2), quiz scores, course section enrollment (Table 2), co-seminar enrollment status, documented step-by-step problem-solving question responses, attendance data, extra credit practice problem completion, final exam score, and final course grade. Student data were collected as part of a larger study on student problem solving in genetics at the University of Colorado Boulder (IRB #16-0511 and #15-0380; PI: J. Knight). All student data were de-identified with random unique identifiers. Over the course of the semester, students were given the opportunity to practice their problem solving by completing eight extra credit documented step-by-step problem-solving questions on their homework assignments. A record of the number of questions a student attempted was kept with the de-identified student data as a practice measurement out of eight. Students were given examples of step-by-step problem-solving documentation during these extra credit practice problems (Figure 1).

Over the semester six quizzes and a final exam were administered to the students. The documented problem-solving questions on the quizzes were administered online and the student responses were collected and de-identified from the submission data. The eight topics chosen for these questions were: Short Tandem Repeat (STR) analysis, genetic mutations, probability, gel/pedigree analysis, nondisjunction, recombination, Restriction Fragment Length Polymorphism (RFLP) analysis, and X-inactivation. The topics were chosen to cover a wide variety of core content areas covered in the course. These topics were spread out on the six quizzes taken by the students during the semester. Student scores for these six quizzes and the final examination were collected and added to the de-identified student data. The documented problem-solving question on the final examination contained a combination of the probability and gel/pedigree topics. Four of the quizzes and the final exam contained one single correct answer constructed response question that asked students to document their step-by-step process as they solved the problem (Figure 2). Quizzes three and four each contained two of these problem-solving process documentation questions on different topics: probability and gel/pedigree, nondisjunction and recombination, respectively. These questions were graded based on the correctness of a student's final answer and for completion credit based on the student's completion of the step-by-step documentation (Figure 3).

For this question, **please document step-by-step** how you are solving the problem as you do it. Include the steps you are taking and why, what you are thinking about as you solve the problem, and how you reached your final answer. Please number each of these steps. Below is an example of how to do this.

Example of how to Document your Problem Solving:
Question: *A double-stranded piece of DNA has 30% adenine. What percentage of the molecule is guanine?*
Documented Problem Solving Answer Example:

1. *Read the problem.*
2. *Know that I am thinking about DNA composition.*
3. *Re-read the problem and note that I'm given adenine (A) %, and I need to determine guanine (G) %.*
4. *From class, I learned that A binds T, and G binds C.*
5. *I draw a piece of double-stranded DNA with A binding thymine (T) and G binding cytosine (C) so that I can visualize and make sure that I remembered correctly.*
6. *Because of this binding, there should be the same amount of A as T and the same amount of G as C.*
7. *Therefore, there should be 30% A and the same amount, 30%, T.*
8. *I add 30% A and 30% T together to account for 60% of the nucleotides.*
9. *There is 100% total of A, G, T, and C, so 100-60%=40% will be the remaining percent.*
10. *Half of 40% is 20%, so there will be 20% of G and 20% of C.*
11. *FINAL CONCLUSION: 20% of the molecule will be guanine.*

Figure 1. Example of a documented step-by-step problem-solving process given to students on their extra credit homework assignments (Avena & Knight, 2019).

For this question, please document step-by-step how you are solving the problem as you do it. Include the steps you are taking and why, what you are thinking about as you solve the problem, and how you reached your final answer. Please number each of these steps.

Phenylketonuria (PKU) is a disease that is inherited in an autosomal recessive manner. Below is a pedigree of a family with a history of PKU. If II-3 and II-4 have a child, what is the chance their child will have PKU?

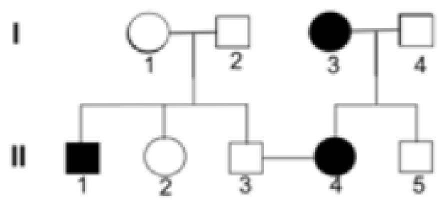

Figure 2. Problem-solving probability question given to students on quiz 3. Students were asked to provide a single correct answer to the question and to document their step-by-step process for solving the problem.

1. Read the problem
2. See who is affected and recognize that PKU is autosomal recessive.
3. Realize that II-3 is not homozygous recessive so he has to have a homozygous dominant or heterozygous genotype.
4. See that II-4 is affected so her genotype is homozygous recessive
5. Realize that I-1 and I-2 are heterozygous because II-1 is affected.
6. That means that II-3 has a 2/3 chance of being a carrier of the gene.
7. If II-3 is a carrier for PKU the child of II-3 and II-4 has a 1/2 chance of getting the disease.
8. Because these are independent events these probabilities need to be multiplied because II-3 needs to be a career. The recessive (II-4) and heterozygous (II-3) parental cross needs to result in an affected individual so the chance of the child being affected is 2/3 times 1/2 which makes the probability the child is affected 1/3.

Figure 3. Example of student documentation of a step-by-step process for solving the probability question given on quiz 3.

The topic for the documented problem-solving question on the final exam was a combination of both the probability and gel/pedigree analysis topics. The documented solutions were collected and de-identified directly from copies of the student final examinations. The final examination consisted of 125 points, 48 of the points were from the 24-multiple choice GCA questions and 77 points were from instructor generated items. A post-GCA measurement was calculated to represent the student's percent score for the GCA questions on the final examination. A final examination instructor generated item component percent (FELP) measurement was calculated to represent the student's percent score for the instructor generated items on the final examination.

After the end of the Spring 2019 term, students who had indicated that they were interested in being interviewed were contacted and asked to participate in an individual interview with a member of the research team. Students who participated in the interviews were compensated for their time with $15 gift cards. Five students agreed to participate in hour-long interviews with members of the research team. The interviews were recorded and then transcribed using the Otter.ai text to speech web application. Interviewed students were asked to verbally document their problem-solving process as they solved genetics problems on topics covered in the course. The students were then asked what problem-solving processes they thought they had used while solving the problem and to define the problem-solving process in their own words. The students were also asked to look at one of their documented step-by-step responses to a problem-solving question during the Spring 2019 semester and tell the interviewer how well they felt their documented response accurately represented their thought process at the time. Students were given the opportunity to discuss aspects of their thought process that they thought may not have been well represented by the written documentation.

Documented Problem-Solving Response Coding

Once student documented step-by-step problem-solving process response data were collected and de-identified the responses were coded for problem-solving processes and content errors. Student documented problem-solving responses were coded for a variety of problem-solving processes (Table 1). Only the "reason correct" and "reason incorrect" process codes were coded for the number of reasoning statements the student made in their documented problem-solving. All other problem-solving process codes were coded as a binary for the presence or absence of coded process within the student's response. A total of 2,070 student responses included in the analysis were coded by between 2 and 5 raters with an average inter-rater agreement of 87.9% and an agreement range between 83.5% and 93.3% on 159 items.

Table 1. Coding scheme used to code student self-documentation of step-by-step problem-solving*

Process Code	**Description**	**Example****
Reason Correct (number of uses)	Student provides a correct or logical rationale for a statement/conclusion.	"Realize that I-1 and I-2 are heterozygous **because II-1 is affected.**"
Reason Incorrect (number of uses)	Student provides an incorrect or illogical rationale for a statement/conclusion.	"II-3 must be a carrier for PKU **because both of their parents are heterozygous.**"
Long-term plan (presence or absence)	Student provides a multistep plan of how they will solve the problem. Followed by a documented execution of at least some of the steps in this plan.	First, I will determine the genotypes of grandparents I-1 and I-2 then I will determine the probability that II-3 is a carrier for PKU. I will then use that information to determine the probability that the child of II-3 and II-4 has PKU.

		(student then proceeds to document the execution of some of these steps)
Short-term plan (presence or absence)	Student provides a statement of intended action followed up by a documented execution of the planned action.	Next, I need to determine the probability that II-3 is a carrier for PKU. (student then documents that they executed their planned next step)
Eliminate (presence or absence)	Student rules out a possible final answer to the question.	The probability that II-3 and II-4's child has PKU cannot be 0… (typically followed by reasoning)
Check (presence or absence)	Checking or confirming a conclusion, approach or incorrect (must be reflective).	Looking back on the information in the pedigree I noticed that one of my initial calculations was incorrect.

*Based on a table included in (Avena et al., Submitted)

**Examples used are specifically for the quiz 3 problem-solving question shown above (Figure 2).

Data Analysis

Problem-solving process data from the coding of responses, surveys, and course data were organized and analyzed using the R programming language version 3.5.1 (R Core Team, 2017). Measurements for average correct and incorrect reasoning use were calculated based on the number of reasoning codes for each student on the eight documented step-by-step problem-solving quiz questions. Measurements for the proportion of use for the short-term planning, long-term planning, checking and eliminating processes were also calculated based on the number of documented step-by-step problem-solving student responses in which they were coded. To potentially capture change on questions with similar content, a measurement for the student's change in correct reasoning use over the semester was calculated based upon the difference between the average of the student's reasoning on the two coded quiz 3 documented problem-solving questions and the final examination problem-solving question. The three outcome variables used to represent student academic success at the end of the semester were the total final exam percent (FEP), FELP and post-GCA measures for each student. The problem-solving process measurements were used along with the student practice on extra credit problems, and pre-GCA data as predictor variables first in simple linear regression (SLR) models (N = 199). The variables that were statistically significant predictors of the outcome variables were then used in multiple linear regression (MLR) models (N = 199) for the same outcome variables along with course section enrollment, co-seminar enrollment, class standing, sex, and demographic data variables. The average performance of each reasoning-based student group on individual quizzes and end of semester assessments was then compared between groups via Welch two-sample t-tests.

Setting and Participants

Participants in this study were 230 undergraduate students between the ages of 18 and 35 enrolled in two large lecture sections of a lower-division introductory genetics course during the Spring 2019 semester term at the University of Colorado Boulder. The course is an introductory genetics course taken primarily by a variety of first-year biological science majors. The course covers topics in transmission genetics, molecular genetics, and population genetics. Each of the two lecture sections of the course was taught by a different instructor. Both instructors used the same lecture slides, course materials, and gave the same assessments to their students. An optional genetics co-seminar course associated with the main lecture courses is taught contemporaneously and provides students with additional practice solving genetics problems. During the study period, 37.4% (86) of the study participants were enrolled in both the main genetics course and genetics co-seminar course. Of the 321 students who consented to participate in the study, 230 completed all nine documented step-by-step problem-solving questions given throughout the semester. Only the data from these 230 students were used in the analysis in the study.

Table 2. Participant demographics based on optional beginning of semester survey*

<table>
<tr><td rowspan="2">Course Section Enrolled (n=230)</td><td colspan="3">Course Section 1</td><td colspan="3">Course Section 2</td></tr>
<tr><td colspan="3">54% (124)</td><td colspan="3">46% (106)</td></tr>
<tr><td rowspan="2">Sex (n= 202)</td><td colspan="3">Male</td><td colspan="3">Female</td></tr>
<tr><td colspan="3">22% (45)</td><td colspan="3">78% (157)</td></tr>
<tr><td rowspan="2">Class Standing (n= 203)</td><td>Freshmen</td><td>Sophomore</td><td>Junior</td><td>Senior</td><td>5th year student</td><td>Other</td></tr>
<tr><td>61% (123)</td><td>19% (38)</td><td>11% (23)</td><td>4% (8)</td><td>1% (2)</td><td>4% (9)</td></tr>
<tr><td rowspan="2">Race/Ethnicity (n= 212)</td><td colspan="2">White</td><td colspan="2">Asian</td><td colspan="2">Historically Underserved Groups**</td></tr>
<tr><td colspan="2">66% (140)</td><td colspan="2">11% (23)</td><td colspan="2">23% (49)</td></tr>
</table>

*Students who did not select a choice for a demographic question were omitted from the data shown in this table (Sex n=28, Class Standing n= 27, Race Ethnicity n=18).

** Includes students who self-identified as Black or African American, Hispanic, American Indian or Alaskan Native, Hawaiian or Other Pacific Islander or some combination of those selections.

Results

Distribution of Problem-Solving Process Use and Academic Measurements

The distribution of the average correct reasoning measurements for individual students showed a slightly positively skewed distribution with most students providing 1-2 correct reasoning statements per documented problem-solving process and a few students providing more than 3 (Figure 4). The distribution of the average incorrect reasoning measurements for individual students showed a positively skewed distribution with most students providing 0 incorrect reasoning statements per documented problem-solving process and a few students providing 1-2 over the 8 quizzes. (Figure 4) All other problem-solving processes coded on student responses were seen in a small proportion of responses (Table 3). The change in correct reasoning use measurement showed a distribution centered around -.51 with many students using fewer correct reasoning statements on the final examination than they had on quiz 3 (Figure 5). The end of semester outcome variables all showed symmetric distributions with similar ranges centered between 70 and 73 percent (Figure 6).

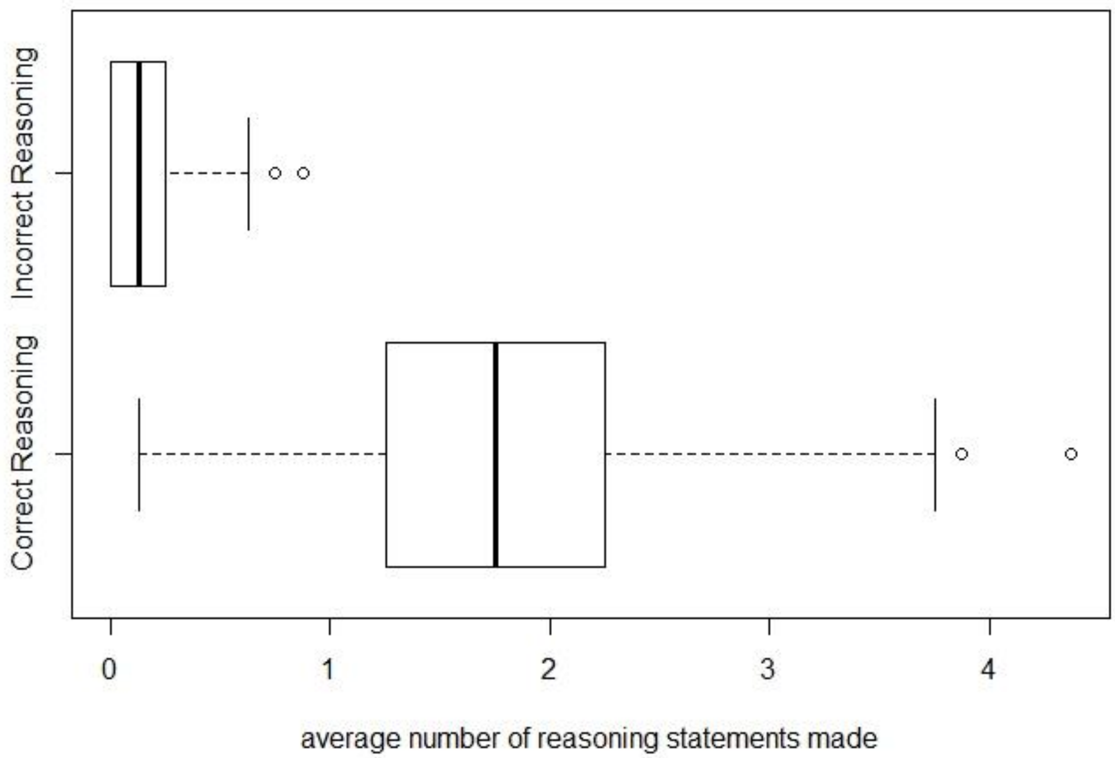

Figure 4. Boxplots showing the distribution of average correct and average incorrect reasoning use for study participants (N = 230). The average correct reasoning distribution had a mean of 1.77 average correct reasoning statements and a median of 1.75 average reasoning statements. The average incorrect reasoning distribution had a mean of .16 average incorrect reasoning statements and a median of 0.13 average reasoning statements.

Table 3. Percentage of student responses coded for other problem-solving processes

Coded Problem-Solving Process	Percentage of Responses with Process Code (number of responses out of 2070)
Short-Term Planning	19.68% (408)
Long-Term Planning	.92% (19)
Eliminate	18.10% (375)
Check	3.15% (65)

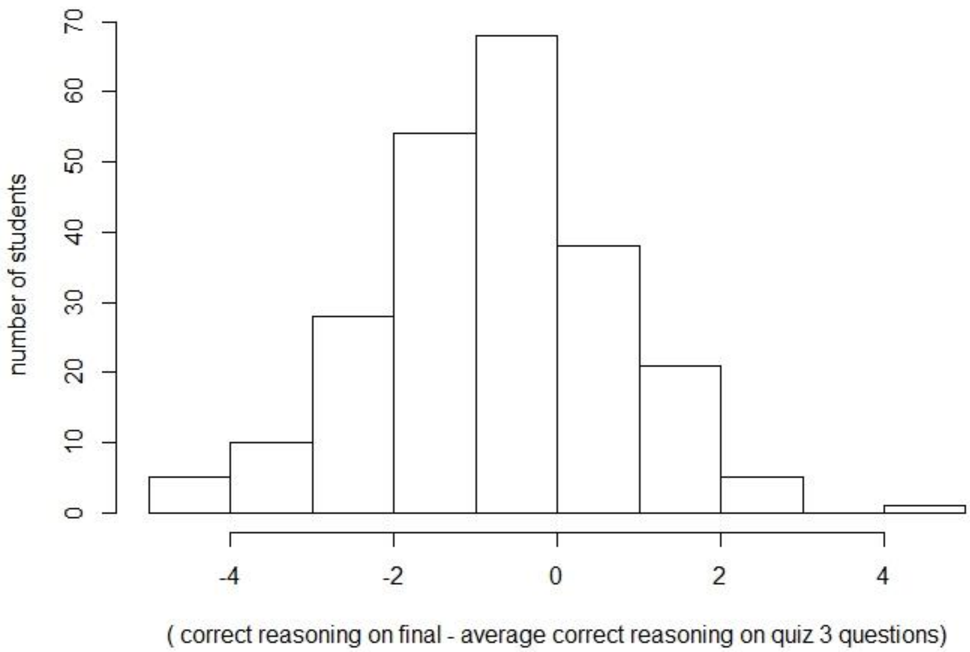

Figure 5. Histogram showing the distribution of correct reasoning use change between quiz 3 and the final exam for study participants (*N* = 230). On average students used .51 fewer correct reasoning statements on the final examination than they did on the quiz 3 questions.

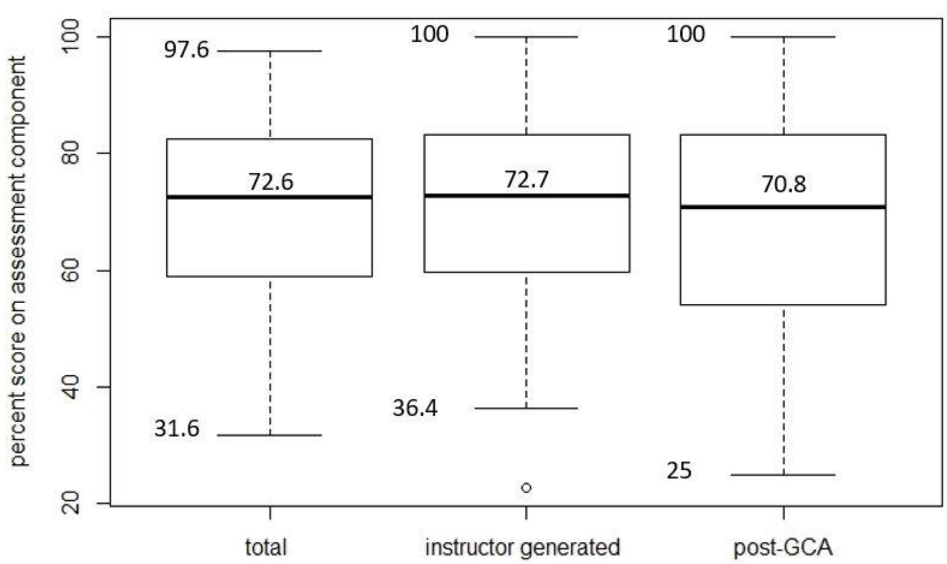

Figure 6. Boxplots showing the distribution of end of semester academic success measurements FEP, FELP, and post-GCA for study participants (*N* = 230). The average total final examination percent (FEP) was 70.7. The average instructor generated final examination percent (FELP) was 71.6. The average GCA component final examination percent (post-GCA) was 69.2.

Simple Linear Regression Models for Academic Success Outcome Variables

No significant predictive relationships were found between the planning (short and long term), checking or eliminating process measurement variables, independently, and any of the three academic success outcome variables The following were all significantly predictive ($p<.001$), in individual models (*N* = 199), of final exam percent: correct and incorrect reasoning, practice problem completion and pre-GCA measurement variables (Figure 7). The SLR model using the correct reasoning measurement variable to predict total final examination showed a significant positive correlation ($r = .49$) between a student's use of correct reasoning and their final examination score (Figure 7A). The SLR model using the incorrect reasoning measurement variable to predict total final examination showed a significant negative correlation ($r = -.51$) between a student's use of incorrect reasoning and their final examination score (Figure 7B). The SLR model using the practice problem completion measurement variable to predict total final examination showed a significant positive correlation ($r = .53$) between a student's completion of practice problems and their final examination score (Figure 7C). The SLR model using the pre-GCA percent measurement variable to predict total final examination showed a significant positive correlation ($r = .45$) between a student's pre-test score and their final examination score (Figure 7D).

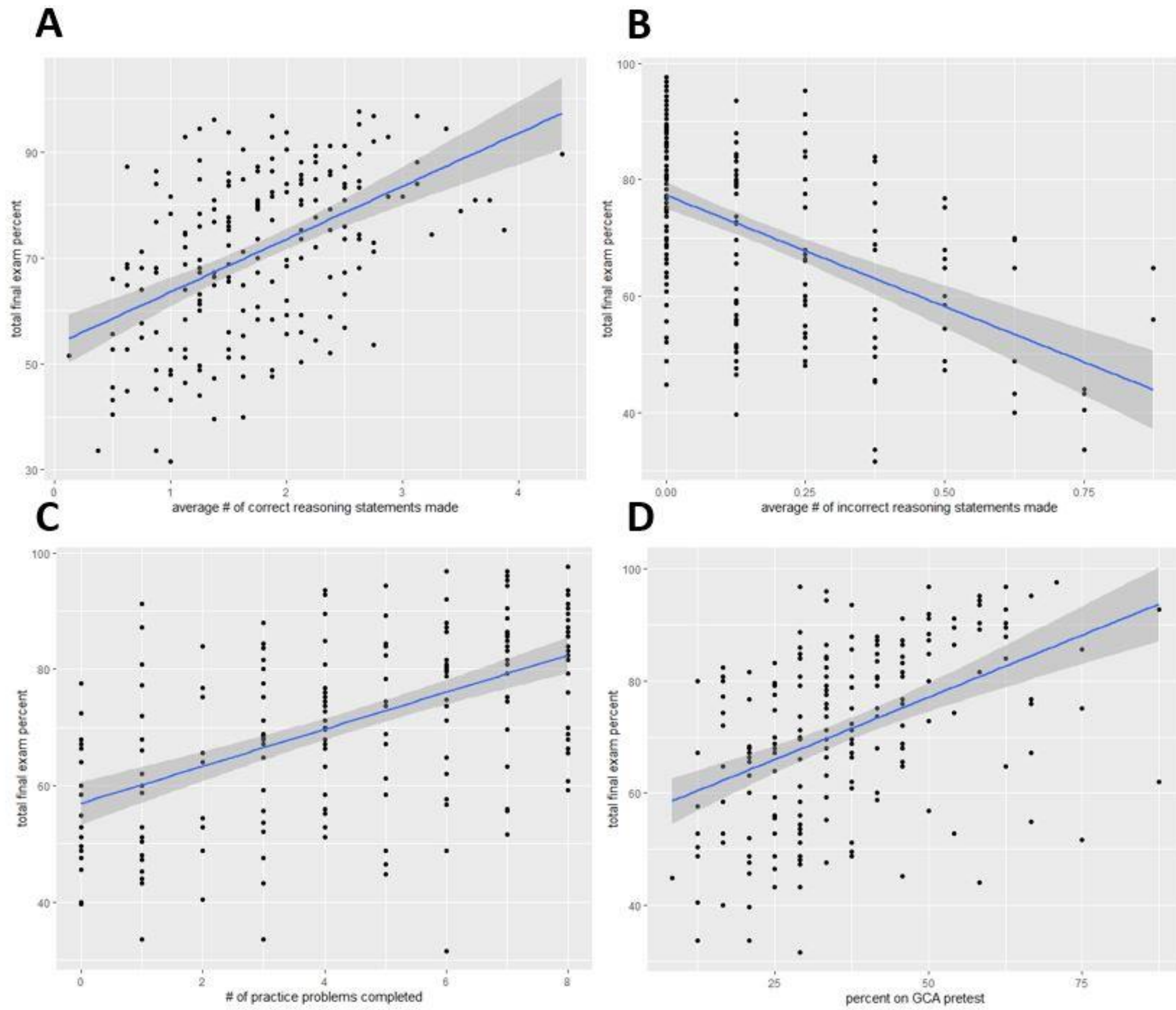

Figure 7. Plots of SLR models predicting total final exam percent. Average correct reasoning measure shows a significant positive relationship (β= 9.993, t = 7.848, p <.001) with total final exam percent (A). Average incorrect reasoning measure shows a significant negative relationship (β= -38.170, t = -8.28, p <.001) with total final exam percent (B). Number of extra credit practice problems measure shows a significant positive relationship (β= 3.184, t = 8.853, p <.001) with total final exam percent (C). Pre-GCA measure shows a significant positive relationship (β= .444, t = 6.993, p <.001) with total final exam percent (D) Regression line predictions are plotted in blue for total final exam percent (*N* = 199). Shaded regions represent 95 percent confidence intervals for regression line predictions.

Multiple Linear Regression Models for Academic Success Outcome Variables

The first MLR model for the FEP measurement (*N* = 199) produced statistically significant estimates for a student's final examination percentage (Table 4). The model indicated that average correct reasoning, average incorrect reasoning, practice problems completed, and pre-

GCA as significant predictors. The second MLR model for the FELP measurement (N =199) produced statistically significant estimates for a student's final examination instructor generated item percent (Table 4). The model indicated that average correct reasoning, average incorrect reasoning, practice problems, and pre-GCA score as significant predictors. The third MLR model for the post-GCA measurement (N = 199) produced statistically significant estimates for a student's final examination GCA item percentage (Table 4). The three models indicated that average correct reasoning, average incorrect reasoning, practice problems completed, and pre-GCA score as significant predictors for the student's score on all components of the final examination.

Table 4. MLR models predicting academic success outcome variables

	Model 1: Outcome FEP	**Model 2: Outcome FELP**	**Model 3: Outcome Post-GCA**
Model Adjusted R^2	$.56^{***}$	$.49^{***}$	$.5^{***}$
Average correct reasoning	4.94^{***}	5.92^{***}	3.37^{***}
Average incorrect reasoning	-21.54^{***}	-21.18^{***}	-22.11^{***}
Practice problems completed	2^{***}	1.76^{***}	2.39^{***}
Pre-GCA	$.263^{***}$	$.2^{**}$	$.37^{***}$

Unstandardized β reported for each predictor in the model. $^{***}p<.001$, $^{**}p<.01$, $^{*}p<.05$

Academic Performance of Students Grouped by Average Correct Reasoning

Students were grouped by their average correct reasoning measurement from their documented problem-solving on quizzes during the semester. These groups were designated as high intermediate and low reasoning. The high reasoning group consisted of students with an average reasoning measurement above one standard deviation from the class average (N = 31). The intermediate reasoning group consisted of students with an average reasoning measurement within one standard deviation from the class average (N = 158). The low reasoning group consisted of students who had an average reasoning measurement below one standard deviation from the class average (N = 41). Students in the high reasoning group had significantly higher average scores on all components of the final examination than students in either the intermediate or low reasoning groups (Figure 8). Students in the intermediate reasoning groups had significantly higher average scores on all components of the final examination than students in the low reasoning group. Students in the high reasoning group also had significantly higher average quiz scores than students in either the intermediate or low reasoning groups across all six quizzes during the semester (Figure 9). Students in the intermediate reasoning group had significantly higher average quiz scores than students in the low reasoning group for every quiz except quiz 2.

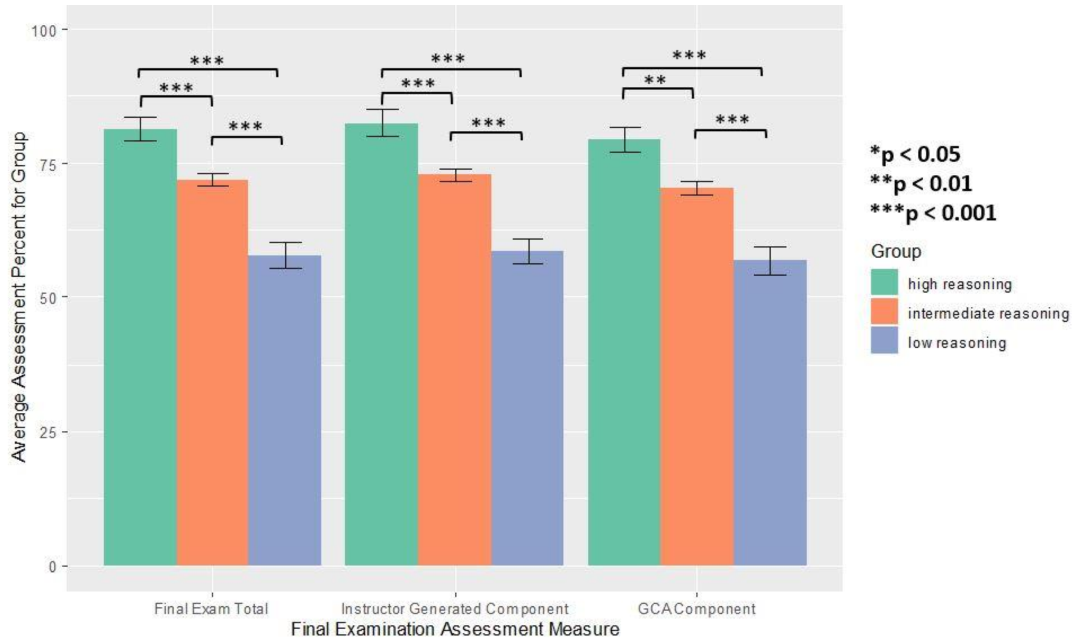

Figure 8. Average percent score on components of the final assessment for students grouped by the average number of correct reasoning statements they made across the semester. High reasoning group (green) included students (*N* = 31) who had average reasoning above one standard deviation (average number of correct reasoning statements > 2.51). Intermediate reasoning group (orange) included students (*N* = 158) who had average reasoning within one standard deviation (2.51 > average number of correct reasoning statements > 1.03). Low reasoning group (blue) included students (*N* = 41) who had average reasoning below one standard deviation (average number of correct reasoning statements < 1.03). Standard error is plotted on each bar in the graph. Welch two-sample t-tests were used to test the significance of the differences between group scores

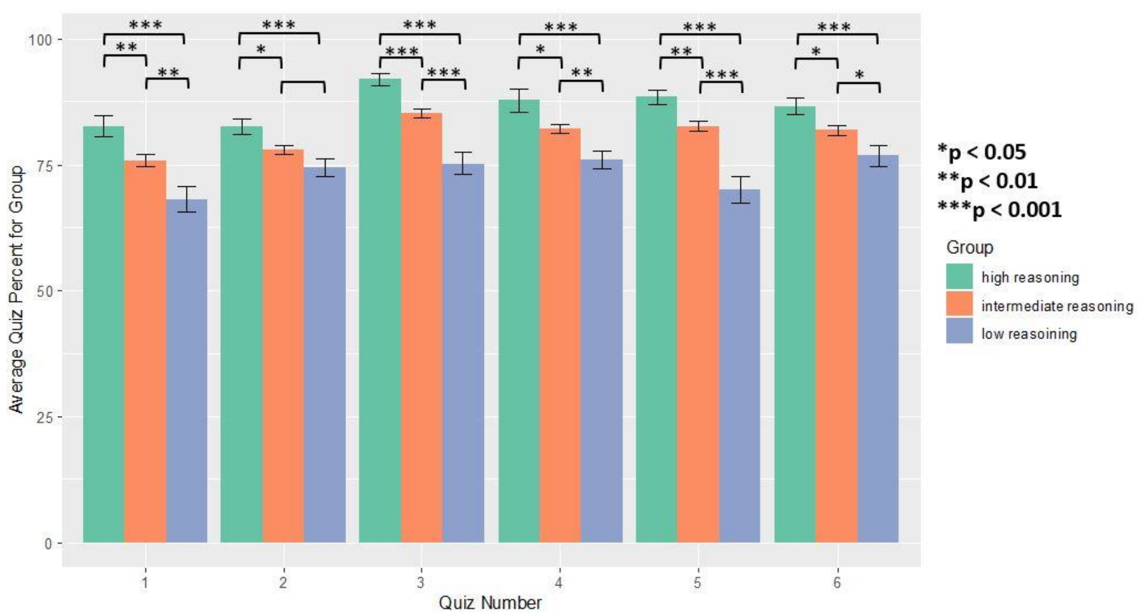

Figure 9. Average percent score on quizzes for students grouped by the average number of correct reasoning statements they made across the semester. High reasoning group (green) included students (*N* = 31) who had

average reasoning above one standard deviation (average number of correct reasoning statements > 2.51). Intermediate reasoning group (orange) included students (N = 158) who had average reasoning within one standard deviation (2.51 > average number of correct reasoning statements > 1.03). Low reasoning group (blue) included students (N = 41) who had average reasoning below one standard deviation (average number of correct reasoning statements < 1.03). Standard error is plotted on each bar in the graph. Welch two-sample t-tests were used to test the significance of the differences between group scores.

Student Definitions and Perceptions of Problem-Solving Processes from Interviews

Five female students participated in interviews after the completion of the semester (Table 5). As the students worked through genetics questions interviewers recorded the student's use of planning, reasoning and checking problem-solving processes. The use of the reasoning process was observed in all five of the students interviewed, the planning process was observed in four of the five, while the checking process was only observed in two of the students. When asked to define the process of reasoning students described it as the way they thought through solving a problem. Student 1 defined the process of reasoning as "an explanation for why you think what you think", Student 2 defined it as "thinking through everything and trying to find things to backup all of the answers or steps that you come to". When asked to define the process of planning students described it as a method to map out or organize the information needed to solve a problem. Student 1 defined the process of planning as a way to "map out what you're going to do before you do it", Student 5 defined it as "laying out like a rough sketch of what I'm going to do". When asked to define the process of checking students described it as the way they made sure all of the information in the problem aligned with their selected answer. Student 5 defined the process of checking as when "[I] go back and even information that I don't need to solve the problem, putting that in and making sure that that lines up with everything else". Most of the interviewed students said that they thought they had been using both reasoning and planning prior to beginning the genetics course. They also described that it was a fundamental part of their problem-solving process. When the students were asked to look back at their own documented problem-solving responses, some discussed that they may have been doing certain things uncaptured by their self-documentation. Student 1 discussed a lack of planning in their self-documentation: "I guess I like to plan it before I do it, but I guess I didn't think about writing down that I planned it". Student 3 mentioned that they felt there were some things they did that they did not write down when solving the problem: "I probably would have been thinking like the steps of meiosis in my head to try to figure it out".

Table 5. Interviewed student data*

Student	Final Examination Score	Pre-GCA Score	Number of Practice Problems Completed	Average Correct Reasoning Use	Average Incorrect Reasoning Use	Planning Process Percentage	Checking Process Percentage
1	91.2	54.2	8	2.25	0.25	0%	11%
2	91.2	45.8	1	2.5	0	33.3%	0%
3	76.8	20.8	6	Student did not complete all documented problem-solving questions on quizzes			

4	58.4	20.8	8	Postbaccalaureate student excluded from main dataset
5	90	37.5	8	Student enrolled in course section not included in main dataset

*All students interviewed had completed the course with the same materials but not all were included in statistical modeling and data analysis due to failure to meet dataset inclusion criteria.

Discussion

Importance of Academic Preparedness and Content Practice in Student Academic Success

Numerous factors impacted the academic success of individual students in the course as demonstrated by the regression modeling of data collected from the students. These models highlighted key factors that had significant impacts on how students performed on the final examination. The regression models show that a student's prior knowledge of genetics concepts measured in the GCA assessment does have a significant relationship with their final exam performance. A recent study found prior knowledge to be correlated with academic performance for both biology and physics undergraduate students (Binder et al., 2019). It has also been shown that students who appear to have deficits in genetics content knowledge are more likely to make content specific errors when trying to solve complex genetics problems (Avena & Knight, 2019). Although student prior knowledge about genetics concepts was gauged by the GCA assessment at the beginning of the semester there was no measurement of a student's prior academic success. A measurement of a student's high school GPA or first-semester college GPA might have provided some additional information on a student's general preparedness and success as a student before they enrolled in the genetics course. This type of measurement could have been used to control for additional variance in overall academic performance on the final examination. The number of optional practice problems a student completed was also shown to be a significant predictor of how well they would do on their final examination. Previous studies have shown that practice test-taking not only improves a student's current understanding of the material but also increases how well they are able to retain that information for use on later examinations (Roediger & Karpicke, 2006; Avena & Knight, 2019). Analysis of these two factors suggests that a student's academic success is related to their prior knowledge of genetics concepts and the number of genetics practice problems they completed.

Importance of Reasoning During Problem-Solving in Student Academic Success

The study showed significant associations between academic success and aspects of the student problem-solving process. The linear regression models generated from the student data suggest that a student's effective use of reasoning plays an important role in their overall success in the genetics course. These reasoning problem-solving variables were significant even in the presence of the other preparedness and practice variables mentioned above. This provides evidence that the reasoning a student is doing when they solve problems can have an independent effect on their academic success. The average correct reasoning frequency measurement had positive correlations with total final exam percent while the average incorrect reasoning frequency measurement had negative correlations with total final exam percent. This suggests that the application of valid and logical reasoning to problem-solving had a positive impact on a student's success in the genetics course. Previous work on student problem solving in genetics on a previous semester of the same course had shown that the process of reasoning during problem solving was a major predictor of correctness on the problem (Avena et al., Submitted). Other work done on the relationship between student argumentation practices while drawing from multiple sources of information and the student's ability to build their explanations of the genetic interactions they were examining supports the importance of reasoning in student academic success (Ageitos et al., 2019).

Another finding of this study was that the student's use of reasoning was a significant predictor of performance on both the instructor generated and GCA items on the final examination. The unstandardized β prediction for the average correct reasoning predictor variable was larger in the model predicting instructor generated item score (5.92) than it was in the model predicting GCA item score (3.37). This suggests that further investigation of associations between student reasoning and academic performance on different types of genetics questions could prove interesting. This study suggests that the amount of reasoning used to work through genetics problems is highly significant in a student's summative performance in an introductory genetics course. An earlier study examined the importance of student reasoning ability in eleventh-grade students and found that the reasoning measurements accounted for a significant portion of student understanding of genetics concepts (Kılıç & Sağlam, 2014). When students in the current study were grouped by their reasoning use, significant differences in academic performance were observed between all groups for all components of the final examination and almost all quizzes. This suggests that reasoning is critical for the success of students learning to solve complex problems in genetics.

Complexities in Measuring the Problem-Solving Process of Reasoning

In past studies, researchers have focused on how student reasoning influences student perceptions of self-efficacy and achievement in introductory biology (Lawson et al., 2007). The authors examined trends in student reasoning and self-efficacy, finding that both increased over the course of the semester. The current study examined how student's use of correct reasoning on their documented problem-solving changed between the two questions they were given on quiz 3 and the question they were given on the final examination. These comparison points were chosen to best control for the impact of question content on reasoning because the final examination question was a combination of the two topics from the quiz 3 questions. No significant increase in reasoning use coded on questions given to students on quiz 3 and the final examination was found. This could be partially due to the significant time limitations placed on the students when they had to complete their documented problem-solving while taking the in-class final examination. In contrast, the students were able to complete their documented problem-solving for the quiz questions on the online portion of each quiz. If reasoning measurements had been compared between quizzes with similar questions at the beginning and end of the semester, we may have observed an average increase in reasoning use instead of a small decrease.

The questions students were asked to answer during the study were all questions with only one correct answer. Although the questions did allow students to construct their own answers and allowed the students more latitude in their answers than multiple choice questions, the questions used still only had single correct answers. It is possible that the nature of the question being asked has some impact on the type and or amount of reasoning observed in the students. If students were allowed to fully construct their responses with a small range of correct answers, we may have seen a different distribution of reasoning. This idea is supported by a recent study on the impact of situational features of reasoning tasks in genetics (Shea et al., 2015). The authors discussed a model for understanding what they called genetics literacy which includes student argumentation ability, content knowledge use and the connection between the two. The content knowledge of genetics literacy was the component the present study aimed to capture using the GCA. The connection between argumentation and content knowledge was the component that the present study aimed to measure by coding student connections between evidence and claims as reasoning. A measurement of argumentation ability or complexity of student reasoning could have been useful in more holistically capturing the process of student reasoning.

The measurements of student reasoning during problem-solving in this study relied only on a measurement of reasoning frequency. This limited the information about a student's reasoning behavior that was incorporated into the reasoning problem-solving process measurement. A measurement of the quality, complexity and or clarity of the student's reasoning would have provided additional information about how the student was using reasoning as they solved the genetics problems. Prior studies on student reasoning during in-class discussions have used quality of reasoning measurements within group contexts to investigate how instructional cues can promote student reasoning (Knight et al., 2013; Knight et al., 2015). This study did not use a quality of reasoning metric due to difficulty in the construction of a reliable measure of rating quality of individual student reasoning across a wide range of questions on different genetics topics. The generation of a quality of reasoning metric that could be standardized to each question could have been useful in assessing student reasoning in combination with the frequency of reasoning measurement. An understanding of how students use different levels of reasoning may better describe how a student's reasoning during their problem-solving relates to their solution to the problem.

Limitations

Due to the limited sample size, and sampling from a single semester the results of this study may not be entirely representative of students across different semesters of the genetics course or of students enrolled in genetics courses at different universities. Investigation of data from multiple semesters of the genetics course at the University of Colorado Boulder would provide additional support to the results of this study. The problem-solving process measurements relied on the students accurately documenting their own internal problem-solving processes. The data from the interviews conducted with students suggest that certain problem-solving processes such as planning may not be self-documented reliably by the students. While the students could have been deciding on what steps they would have to take internally they may not have written down these steps in their self-documentation. Both interviewed Students 1 and 2 discussed parts of the problem-solving process they believed they had done while solving the problem that were absent from their written self-documentation. This suggests that the student self-documentation may be under capturing or missing certain aspects of the student problem-solving process. The relatively low frequency of planning and checking codes in student responses supports this idea. It is possible that certain components of the problem-solving process are more difficult for students to document because they are so intuitive. Changes to the documented problem-solving question prompts or could have resulted in better capture of the planning and checking process by more directly prompting students to think about those processes.

Implications for Genetics Instruction

Instructors are unable to control how much a student already knows about genetics topics, how prepared they are for their coursework in general, or how much they practice with the course material on their own time. However, instructors may be able to encourage the use of reasoning during problem-solving by showing students how to utilize reasoning during class time. The data suggest that the amount of reasoning a student uses when solving problems can have an independent impact on student performance on the final examination. The study suggests that genetics instructors could have significant impacts on how students perform in their course by helping their students increase their reasoning use during problem solving. The data suggest that this improvement in performance based on reasoning behavior may be independent of a student's prior knowledge of genetics content and the amount of practice the student completes. A past study suggested that statistical modeling of student performance and reasoning may allow instructors to assist their students in recognizing effective problem-solving strategies (Stevens et al., 2005). A better understanding of how components of the student problem-solving process relate to student academic performance could help instructors promote problem-solving behaviors in their students. Instructors who are able to foster the importance of justifying claims answers with reasoning may be able to provide their students a significant increase in academic performance.

References

Ageitos, N., Puig, B., & Colucci-Gray, L. (2019). Examining Reasoning Practices and Epistemic Actions to Explore Students' Understanding of Genetics and Evolution. *Science & Education*, *28*(9), 1209–1233. https://doi.org/10.1007/s11191-019-00086-6

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

Avena, J. S., McIntosh, B. B., Whitney, O., Weins, A., & Knight, J. K. (Submitted). *Successfully solving genetics problems requires reasoning*.

Binder, T., Sandmann, A., Sures, B., Friege, G., Theyssen, H., & Schmiemann, P. (2019). Assessing prior knowledge types as predictors of academic achievement in the introductory phase of biology and physics study programmes using logistic regression. *International Journal of STEM Education*, *6*(1), 33. https://doi.org/10.1186/s40594-019-0189-9

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

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.

Hoskinson, A.-M., Caballero, M. D., & Knight, J. K. (2013). How Can We Improve Problem Solving in Undergraduate Biology? Applying Lessons from 30 Years of Physics Education Research. *CBE—Life Sciences Education*, *12*(2), 153–161. https://doi.org/10.1187/cbe.12-09-0149

Kılıç, D., & Sağlam, N. (2014). Students' understanding of genetics concepts: The effect of reasoning ability and learning approaches. *Journal of Biological Education*, *48*(2), 63–70. https://doi.org/10.1080/00219266.2013.837402

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

Lawson, A. E., Banks, D. L., & Logvin, M. (2007). Self-efficacy, reasoning ability, and achievement in college biology. *Journal of Research in Science Teaching*, *44*(5), 706–724. https://doi.org/10.1002/tea.20172

National Science Board. (2019). *Higher Education in Science and Engineering. Science and Engineering Indicators 2020.* https://ncses.nsf.gov/pubs/nsb20197/executive-summary

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

R Core Team. (2017). *R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna: Austria.*

Redish, E. F., & Steinberg, R. N. (1999). Teaching Physics: Figuring out What Works. *Physics Today*, *52*(1), 24–30. https://doi.org/10.1063/1.882568

Reif, F. (1986). Scientific Approaches to Science Education. *Physics Today*, *39*(11), 48–54. https://doi.org/10.1063/1.881044

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

Schmiemann, P., Nehm, R. H., & Tornabene, R. E. (2017). Assessment of Genetics Understanding. *Science & Education*, *26*(10), 1161–1191. https://doi.org/10.1007/s11191-017-9925-z

Shea, N. A., Duncan, R. G., & Stephenson, C. (2015). A Tri-part Model for Genetics Literacy: Exploring Undergraduate Student Reasoning About Authentic Genetics Dilemmas. *Research in Science Education*, *45*(4), 485–507. https://doi.org/10.1007/s11165-014-9433-y

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. (2016). Teaching Genetics: Past, Present, and Future. *Genetics*, *204*(1), 5–10. https://doi.org/10.1534/genetics.116.187138

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. https://doi.org/10.1187/cbe.08-08-0045

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

Stevens, R., Johnson, D. F., & Soller, A. (2005). Probabilities and Predictions: Modeling the Development of Scientific Problem-Solving Skills. *Cell Biology Education*, *4*(1), 42–57. https://doi.org/10.1187/cbe.04-03-0036

Taconis, R., Ferguson-Hessler, M. G. M., & Broekkamp, H. (2001). Teaching science problem solving: An overview of experimental work. *Journal of Research in Science Teaching*, *38*(4), 442–468. https://doi.org/10.1002/tea.1013

Thompson, E. D., Bowling, B. V., & Markle, R. E. (2018). Predicting Student Success in a Major's Introductory Biology Course via Logistic Regression Analysis of Scientific Reasoning Ability and Mathematics Scores. *Research in Science Education*, *48*(1), 151–163. https://doi.org/10.1007/s11165-016-9563-5