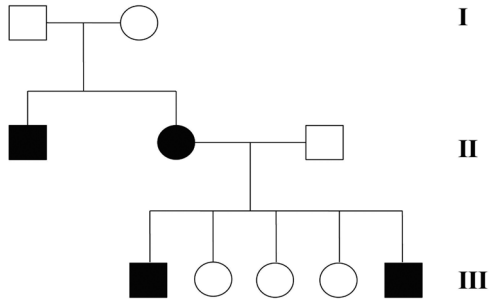
RESEARCH ON LEARNING

An Assessment to Investigate Student Conceptions of Pedigree Analysis

ZACHARY T. GRIMES, NANCY M. BOURY, CHLOE WASENDORF, AUDREY L. MCCOMBS, JOSHUA W. REID, OLENA JAMES, BROCK COUCH, PATRICK I. ARMSTRONG, REBECCA L. SEIPELT-THIEMANN



Abstract

Genetics plays an increasing role in modern life as evidenced by the development of revolutionary techniques such as CRISPR-based genome editing and the rise of personalized genome services. However, genetics is difficult to learn; known issues include its abstract nature, different scales, and technical language. Pedigree analysis is a convergence of these concepts, requiring use of multiple symbolic scales and understanding the relationships and nature of alleles, genes, and chromosomes. To measure student understanding of these concepts, as well as support biology educational reform toward student-centered instruction, we developed a formative assessment to provide reliable and valid evidence of student understanding, learning, and misconceptions for pedigree analysis. Nine multiple choice items targeted to four learning objectives were developed in an iterative process with faculty and student input. We designed distractor answers to capture common student misconceptions and deployed a novel statistical technique to assess the congruence of distractor language with targeted misconceptions. Psychometric analysis showed the instrument provides valid and reliable data and has utility to measure normalized learning gains. Finally, we employed cross-tabulation and distractor progression to identify several stable misconceptions that can be targeted for instructional intervention.

Key Words: *pedigree; assessment; distractor congruence; undergraduate genetics; naive conceptions.*

○ Introduction

Over the past 30 years, genetic technologies have changed modern life in ways no one could have predicted. Sequencing of over 40,000 genomes (Sayers et al., 2019), the rise of personal genetic sequencing companies such as 23 and Me (Phillips, 2016), the construction of genome-edited humans (Cyranoski, 2019), and the development and use of personalized medicine (Gibson, 2019) are but a few examples that have affected modern life. Genetics concepts are both central to and foundational in biology, as evidenced by their position within elementary and secondary curricula (Pruitt, 2014), freshman college survey courses, and as a full course generally targeted to sophomore-level biology major and minor students.

Despite the increasing impact of the field, genetics is not well understood by either the lay person or students of biology. In fact, it is a well-known frustration point for students in biology (Bahar et al., 1999; Çimer, 2012). Genetics presents many problems for student learning due to its abstract nature, the need to consistently utilize different spatial and symbolic scales (Bahar et al., 1999; Johnstone, 1991; Niebert & Gropengiesser, 2015), and heavy reliance on technical language (Batzli et al., 2016; Ramorogo & Wood-Robinson, 1995; Sutton, 1996). For example, when students studied transmission of genetic traits, they were unable to make the connection between a gene and its resulting phenotype (Lewis & Kattmann, 2004) or an allele and its expression (Newman et al., 2021). Naïve conceptions and confusion are further exacerbated by outdated or incomplete information, as well as ambiguity in the subject itself (Gericke & Hagberg, 2010).

Pedigree analysis represents a convergence of many of these difficult concepts in genetics and, therefore, has the potential to help instructors identify and aid in the resolution of multiple student-learning difficulties. Pedigree analysis requires understanding of genetic transmission and meiosis, the dominance and recessiveness of alleles and traits, and the differential inheritance of sex chromosomes and autosomes. Pedigree analysis also requires the ability to use multiple symbolic scales, consolidate and critically evaluate multiple types of evidence, and understand the interconnected nature of rationally linked terms such as *genotype* and *phenotype*, as well as *gene*, *allele*, and *chromosome* (Batzli et al., 2016; Hackling, 1994; Hackling & Lawrence, 1988; Stewart & Kirk, 1990; Timm et al., 2022). Additionally, pedigree charts represent both practical and broadly used applications of biology and are long-lived, having been used for centuries (Resta, 1994) with official records kept as proof of lineage and to avoid inbreeding among purebred species, such as thoroughbred horses. Genetic counselors utilize pedigree charts to provide couples with the likelihood of having an affected offspring based on family history, carrier status, and genetic test statistics (Battista et al., 2012). Since their applied uses are many, pedigree charts provide an excellent mechanism to engage students in learning several challenging and practical aspects of genetics and biology.

The American Biology Teacher, Vol. 84, No. 9, pp. 535–544, ISSN 0002-7685, electronic ISSN 1938-4211. DOI: https://doi.org/10.1525/abt.2022.84.9.535.

Active learning approaches, strongly advocated in the *Vision and Change* report (Brewer & Smith, 2009) and the President's Council report (Gates & Mirkin, 2012), have particularly beneficial effects during the critical first two years of college (Braxton et al., 2008) and for underrepresented groups (Berry, 1991), decrease failure rates, and increase performance in STEM courses (Freeman et al., 2014). However, evidence-based tools for assessing active learning methods are currently lacking. The goal of this project was to construct an assessment tool that would enable faculty who teach at or before this critical second college year to compare student performance against defined external standards (Erlbaum, 1999; Glaser, 1963). These standards are the learning objectives specific to the genetic concepts in pedigree analysis. Thus, this assessment is categorized as a criterion-referenced assessment (CRA). Faculty can use the Pedigree Analysis Criterion-Referenced Assessment (PACRA) to determine student background knowledge by using it prior to instruction. They can also assess the effectiveness of different teaching methods by using the PACRA before and after instruction.

○ Methods

Participants & context

This study was conducted across seven institutions in the United States: Utah Valley University (UVU), Bridgewater State University (BSU), University of Wisconsin at La Crosse (UWL), University of Northwestern at St. Paul (UoN), Iowa State University (ISU), Middle Tennessee State University (MTSU), and University of North Carolina at Asheville (UNC). As of Fall 2022, these universities were of six Carnegie classification types: M3, baccalaureate, M1, M2, R1, R2, and baccalaureate respectively. Student populations ranged from 1800 to nearly 38,000 students, between 11% and 34% non-White, and between 43% and 68% female. We collected multiple choice data between August 2017 and December 2018 and congruence data in the spring of 2020. Student participants were enrolled in general biology, general genetics, or advanced genetics courses (Supplemental Material, Table S1, available with the online version of this article). This study was approved through an Internal Review Board process (MTSU IRB18-1002; ISU (#17-213); UVU: (IRB #01995); BSU (Exempt); UWL (Approved ISU IRB 17-213); UoN (Exempt); UNC (Exempt).

Inventory design

We used established methods to design the instrument with an additional step of congruence testing for distractor reliability (Adams & Wieman, 2011; D'Avanzo, 2008; Kalas et al., 2013; Knight, 2010; Paustian et al., 2017; Rago et al., 2007). We used four development phases: learning objective and question development, question critique and distractor development, inventory performance analyses, and final inventory performance and congruence testing (Figure 1). The research team included faculty with expertise and extensive teaching experience in genetics who had previously created and published concept inventories, four predoctoral scholars in STEM education programs, a predoctoral applied statistics researcher, and a faculty psychometrician.

Learning objectives development

Assessment strategies are more effective when a backward design is used during development, so our first step involved developing learning objectives reflective of current educational goals in both genetics and biology using the backward design model (Brewer & Smith, 2009; Genetics Society of America Education Committee [GSA], 2015). The GSA's Core Categories of particular importance were "Nature of the Genetic Material" and "Patterns of Inheritance." We distributed draft concepts and learning objectives to four experts with extensive experience teaching undergraduate genetics courses. We used their feedback to refine and clarify both the concepts and student learning objectives. Faculty agreed that students should be able to deduce individual genotypes within a pedigree given phenotype information and mode of inheritance (learning objective 1 [LO1]); predict/identify the most likely mode of inheritance represented in the pedigree (LO2); identify and explain how specific individuals can be used to exclude a mode of inheritance (LO3); and identify and explain how mitochondrial inheritance can be detected from pedigree diagrams (LO4) (Table 1). These student learning objectives served as the basis for question development.

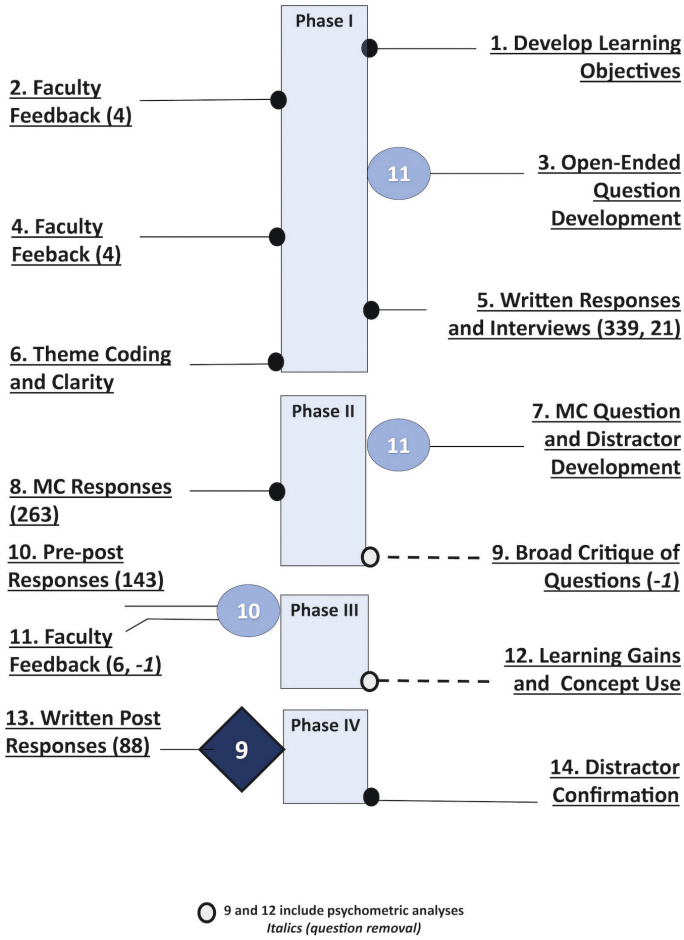


Figure 1. Inventory development schematic. Phases are shown in the center rectangles from top (earliest) to bottom (latest) with numbered events in each phase listed at right and left. Large symbols with numbers inside show the number of inventory items at important development junctures. Item removal is noted by a negative number in italics (steps 9 and 11). Sample size (student responses or faculty input) is noted by numbers inside parentheses and open circles with dashed lines indicate when psychometric analyses were conducted (steps 9 and 12). Phase I: learning objectives and question development. Phase II: question critique and distractor development. Phase III: inventory performance psychometric analyses. Phase IV: final inventory testing, confirming distractors with congruence testing.

Table 1. Pedigree analysis concepts, learning objectives, and alignment to the GSA genetics learning framework.

Concept	Learning Objective	GSA Example Learning Objective
Genotypes can be inferred based on inheritance patterns shown in a pedigree chart.	Deduce individual genotypes within a pedigree given phenotype information and mode of inheritance.	"Calculate the probability that an individual in a pedigree has a particular genotype."
Pedigree charts can be used to predict the most likely mode of inheritance.	Predict/identify the most likely mode of inheritance represented in the pedigree.	"Using pedigrees, distinguish between dominant, recessive, autosomal, X-linked, and cytoplasmic modes of inheritance."
Pedigree charts can be used to investigate relationships between phenotype and allele inheritance patterns.	Identify and explain how specific individuals can be used to exclude a mode of inheritance.	"Using pedigrees, distinguish between dominant, recessive, autosomal, X-linked, and cytoplasmic modes of inheritance."
Pedigree charts can be used to distinguish between Mendelian and mitochondrial inheritance.	Identify and explain how mitochondrial inheritance can be detected from pedigree diagrams.	"Using pedigrees, distinguish between dominant, recessive, autosomal, X-linked, and cytoplasmic modes of inheritance."

Item (question & distractors) development

We developed eleven open-ended questions, and faculty not involved with question development reviewed them to verify questions were at an appropriate level, were readable, and involved material covered in their courses. The open-ended inventory was then disseminated across two universities (MTSU and ISU) to 339 students. Student interviews were conducted at MTSU on a self-selected subsample (n = 21) of the overall sample during the spring 2018 semester. The interviews were semistructured using a think-aloud method based on the students' written responses to the open-ended prompts. All interviews had an average length of 20 minutes, were audio recorded, and then were transcribed. First, we coded responses as correct or incorrect. Second, we inspected student answers for difficulty with terminology or item readability. Third, we discussed and categorized incorrect student reasoning, coded them into themes to generate a master code list with example student language, and subsequently used incorrect student reasoning and language as the source of multiple-choice distractors (Adams & Wieman, 2011). As a resource for instructors, we documented examples of student reasoning, many of which involve problems in terminology, scale, mathematics, and the symbolic nature of genetics (Supplemental Material, Table S2).

Inventory performance

The next step of inventory development was to design and test multiple-choice items. We rephrased the open-ended questions as multiple-choice questions and used common misconceptions as distractor answers. We used actual student language whenever possible to construct each item (question plus correct answer plus distractors). We used the inventory as a posttest at ISU and MTSU and involved 263 students from both lower- and upper-division genetics courses. Preliminary validity and reliability statistics were calculated by the faculty psychometrician (data not shown) and discussed among the research team. Based on this discussion, we removed one item and revised the remaining items to shorten and clarify most items. We then tested the inventory in a pre-post design with students across seven colleges and universities (pretest n = 313; posttest n = 280). Of these students, 143 answered both the pretest and posttest. Faculty input and analysis of student answers indicated a problem with one item due to shortening of some answers. This item was removed for a final total of 9 items. Reliability and validity statistics for CRAs—such as item difficulty, discrimination index, point biserial correlation coefficient, and KR20 (a measure similar to Cronbach's alpha that is used for binomial data)—were calculated using the 280-student posttest data set using SPSS (Findley, 1956). We gathered demographic data to determine if there were differences based on gender, year in school, GPA, first generation student, and ESL status. Using JMP Pro 16 (Sall et al., 2017), we ran an ANOVA comparing the posttest scores based on each demographic category (Table S1). The only category demonstrating significant differences was the year in school. We then compared each year in school using Tukey's adjustment for multiple comparisons and a student's t-test.

Distractor congruence

As a final step, we measured the ability of each distractor to accurately capture the rationale for which it was designed using a novel statistical technique described in detail in the Supplemental Material (Supplemental Method) and Wasendorf and colleagues (2022). We calculated for each item the probability that we would see the observed number of congruent self-reported rationales under a null hypothesis that students were randomly guessing both the answer and their rationale for answering the way they did. The reported probabilities are similar to a p-value; see Supplemental Material (Supplemental Method) for more details.

Learning & reasoning change measures

We used paired data (n = 143) to calculate several learning and reasoning change measures: normalized learning gains, distractor frequency change, and mastery analysis. We calculated normalized learning gains using Hake's formula (Hake, 1998; McKagan et al., 2017)—(posttest score – pretest score) / (100 – pretest score)—for each student in JMP Pro 13; the average of gains is presented for each item. We investigated group-level changes in student reasoning by comparing distractor frequencies on the pretest to their frequencies on the posttest. Finally, we used cross tabulation to identify

the following mastery/nonmastery categories: preknowledge (correct on pretest / correct on posttest); learning (incorrect/correct); a mastery gap (incorrect/incorrect); and reversion (correct/incorrect).

○ Results

Instrument design, reliability & validity of the PACRA

After implementing the iterative design strategy (Figure 1) (Adams & Wieman, 2011; D'Avanzo, 2008; Kalas et al., 2013; Knight, 2010; Paustian et al., 2017; Stefanski et al., 2016), the final assessment contained nine items across four learning objectives (Table 1). We initially designed the instrument with eleven items. One was removed from LO2 due to technical error in the item, and another was removed from LO4 due to the high difficulty, and student answers were randomly distributed across all distractors (data not shown). To ensure that the data from this assessment would be reliable (consistent across repeated use) and valid (accurate), we calculated reliability and validity measures appropriate for CRAs including item difficulty, discrimination index, point biserial correlation coefficient, and the Kuder-Richardson index (KR20) of the final nine items (Table 2).

Index of difficulty, which is the fraction of students answering correctly, had an average of 0.54 ± 0.1 and was broadly distributed with a range of 0.42–0.69; a broad distribution is considered optimal when constructing multiple choice items (Kubiszyn & Borich, 1987). Three items were in the range of 0.69–0.60; three items were in the range of 0.59–0.50; and three items were in the range of 0.49–0.40. Next, the discrimination index (D_{27}), a measure that indicates distinction between low-performing students and high performing student, was calculated for each item and was in the range of 0.44–0.77 (average: 0.66 ± 0.12). All were above the acceptable level of 0.3 (Ebel, 1954). Point biserial correlation coefficient, which has an acceptable level of >0.2 (Onwuegbuzie & Daniel, 1999) and is a measure of each item's reliability, was in the range of 0.24–0.50 and averaged 0.39 ± 0.12. This measure was calculated by comparing the student's score on a single item to the student's total score. The overall reliability of the final nine-item inventory, as measured by the KR20 index, was 0.71, which is above the accepted minimum level of 0.7 (Ding et al., 2006; Doran, 1980). Therefore, this CRA meets or exceeds community standards regarding validity and reliability measures (Cohen & Swerdlik, 2018; McCowan & McCowan, 1999).

Distractor congruence

Proportions of congruent answers were higher than proportions of incongruent answers for all items and all answer types (Figure 2). Probabilities of observing the number of congruent answers under the null hypothesis of random guessing fell between <0.001 and 0.093 (Table 3). Our congruence analysis indicated that overall the distractor answers captured the student misconceptions for which they were designed.

Measuring student learning

To ensure this CRA would be useful for measuring student learning, we calculated both the group means and normalized learning gains (NLG) in a paired pretest/posttest sample (n = 143; Table 2). Student responses were higher on the posttest than the pretest with a pretest mean of 3.62 and a posttest mean of 5.23 out of 9, $p <$ 0.0001. Posttest scores were not statistically different among most of the demographic groups tested. However, the first- and second-year students were statistically distinct from the third- and fourth-year

Table 2. Difficulty, reliability, discriminatory power, and utility measures of the PACRA. * $p < 0.01$, ns = not significant, and KR20 = 0.71 (posttest data).

		***n* = 280 (validation posttest data)**			***n* = 143 (paired data)**
Learning Objective	**Item**	**Index of Difficulty**	**Discrimination Index**	**Point Biserial Correlation**	**Normalized Learning Gains**
Deduce individual genotypes within a pedigree given phenotype information and mode of inheritance.	1	0.65	0.75	0.50	48.34% *
	2	0.69	0.65	0.42	52.86% *
	3	0.62	0.76	0.50	32.26% *
	4	0.42	0.67	0.42	25.27% *
Predict/identify the most likely mode of inheritance represented in the pedigree.	5	0.53	0.44	0.16	34.64% *
Identify and explain how specific individuals can be used to exclude a mode of inheritance.	6	0.51	0.77	0.46	37.14% *
	7	0.58	0.75	0.46	24.19% ns
	8	0.44	0.61	0.34	1.41% ns
Identify and explain how mitochondrial inheritance can be detected from pedigree diagrams.	9	0.43	0.50	0.24	5.63% ns
Mean values		**0.54 ± 0.10**	**0.66 ± 0.12**	**0.39± 0.12**	
Range		**0.42–0.69**	**0.44–0.77**	**0.24–0.50**	
Acceptable level		**Broad range**	**<0.3**	**>0.2**	

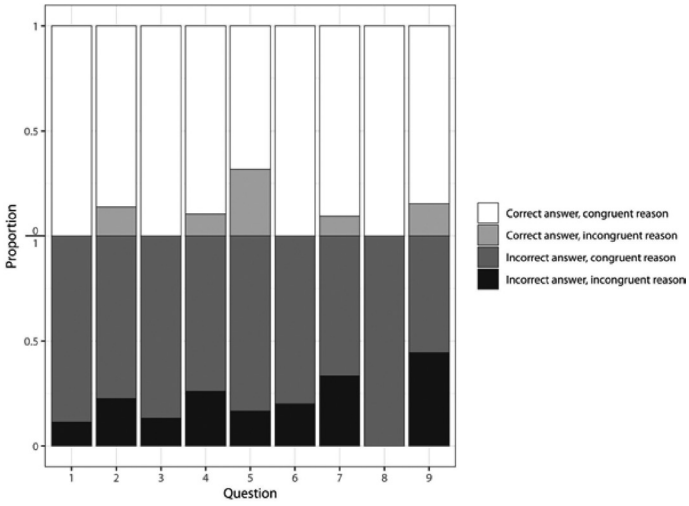


Figure 2. Congruence of student rationale with target rationale. For each item, the proportion of incorrect (bottom) or correct (top) answers that were incongruent (dark) or congruent (light) with target misconception or correct reasoning is displayed as a frequency histogram ($n = 55$).

Table 3. Congruence probabilities.

Item	Probability
1	0.001
2	0.010
3	<0.001
4	0.015
5	0.008
6	0.067
7	0.093
8	0.013
9	0.052

students (Supplemental Material, Table S1). Normalized learning gains, which is a measure of how much learning was observed given how much learning was possible, were significant in six of the nine items (items 1–6) with the highest NLG in items for LO1 (items 1–2, 48% and 53% respectively). Items related to moderately difficult concepts, LO1 (items 3–4), LO2 (item 5), and LO3 (item 6), showed moderate learning gains across the student pool (25–37%). The most difficult concepts (items 7–9) showed no statistically significant learning gain.

To better understand how student thinking changed after instruction, we next compared answer frequencies from pre- to postinstruction (Figure 3). We were most interested in identifying which misconceptions were persistent and resolved postinstruction as a group. Consistent with positive learning gains, correct answer selection shows a steep positive slope for all items except items 7 and 8. While most distractors have a negative slope, indicating that the misconception was resolved to a correct conception (or possibly a different misconception), most items have at least one distractor with a zero or nearly zero slope, indicating a persistent misconception at least at the population level, such as distractors for Mendelian inheritance in items 1, 6, and 7; distractors for terminology/misreading in items 2, 3, and 6; distractors for nature of genes and alleles in item 6; distractors for dominance/recessiveness in item 7; and distractors for mitochondrion-X chromosome conflation and females only acquiring mitochondrial disorders in item 9 (Table 4). Interestingly, two distractors gained in frequency: a distractor for transmission genetics in item 4 and a distractors for conflation of X-linked dominance/recessiveness in item 8. These stable, item-specific misconceptions are good targets for new instructional interventions that can be tested using this CRA.

Since group means, normalized learning gains, and distractor frequency changes are consolidated measures of learning, we next investigated student-specific mastery using cross tabulation to quantify the following: preknowledge (correct answer on pre- and posttest), mastery gap (incorrect answer on pre- and posttests), learning (incorrect answer on pretest and correct answer on posttest), and

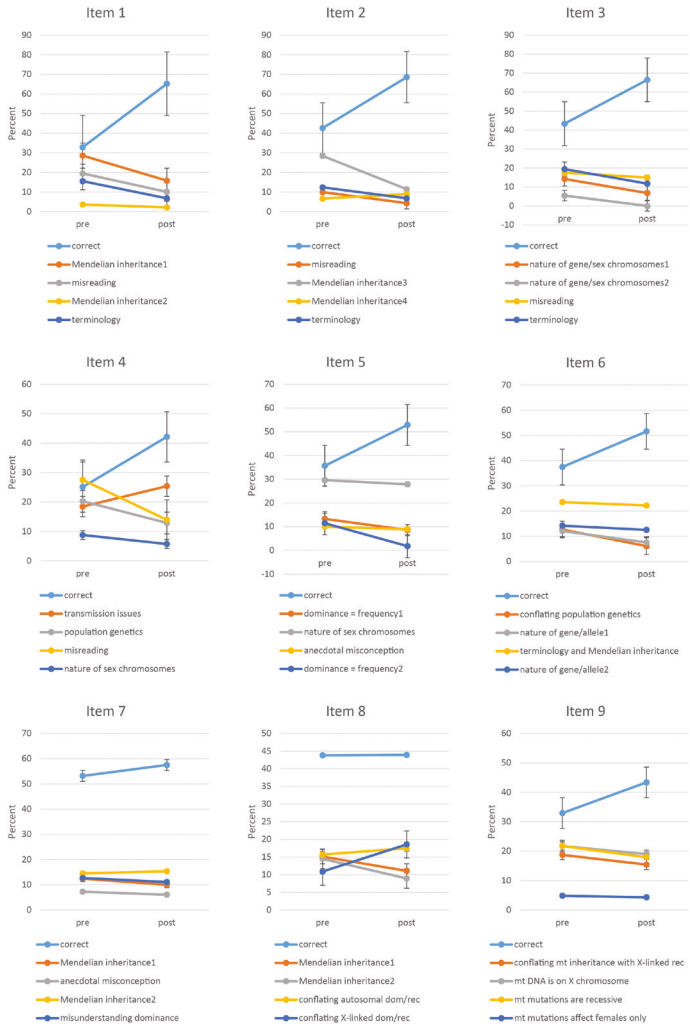


Figure 3. Changes in student answer choice following instruction. The percentage of students selecting each multiple-choice answer was calculated for the pretest and posttest ($n = 143$). The slope of the line connecting pre- and posttest frequencies is an indication of the strength and direction of the change in answer choice from pre- to posttest. Items 1–4 are from LO1; item 5 is from LO2; item 6–8 are from LO3; item 9 is from LO4. A distractor's conceptual error is noted as a phrase below each graph (also see Table 4).

Table 4. Stable misconception categories with student examples.

Distractor Category	Item	Description	Student Example Answer
Mendelian inheritance	1	Student was asked the genotype of an affected male identified in the pedigree, given that it represents an autosomal recessive trait.	Heterozygous dominant.
	6	Student was asked to exclude a mode of inheritance using a falsification strategy.	II-1 wouldn't express the trait. In the second generation one of the offspring doesn't have any black in its genotype so it cannot be recessive.
	7	Student was asked to exclude a mode of inheritance using a specific individual.	If it was a dominant trait, both III-3 and II-3 would also be affected by the same trait that III-4 got.
Terminology/ misreading	2	Student was asked the genotype of an unaffected female identified in the pedigree, given that it represents an autosomal recessive trait.	Dominant because it has the disease.
	3	Student was asked the genotype of an affected male identified in the pedigree, given that it represents an X-linked recessive trait.	The genotype of III-5 would likely be Aa, since it is exhibiting the dominant black color, and the recessive square shape.
	6	Student was asked to exclude a mode of inheritance using a falsification strategy.	III-4 has two homozygous recessive parents.
Nature of genes and alleles	6	Student was asked to exclude a mode of inheritance using a falsification strategy.	Because it is on the first square, meaning that it is a male. If it were X linked, then it would only impact females. Male individuals could be carriers, but would not express the phenotype themselves.
Dominance/ recessiveness	7	Student was asked to exclude a mode of inheritance using a specific individual.	Because it cannot be recessive because both of its parents did not have the trait.
Mitochondria/ X chromosome	9	Student was asked to identify which pedigree more likely represents a disorder due to mutations in the mitochondria.	Pedigree B because mitochondria traits can be traced through X genes and all of the circles are black in one lineage.
Mitochondria/ female	9	Student was asked to identify which pedigree more likely represents a disorder due to mutations in the mitochondria.	Pedigree B shows evidence that is based in mitochondrial mutations instead of nuclear mutations because all females from the same line contain the trait.

reversion (correct answer on pretest and incorrect answer on posttest) within the same paired data set (Figure 4). In this visualization, wide shapes represent mastery, either due to preknowledge or learning, while tall shapes represent nonproficiency, either due to mastery gap or reversion. We see mastery for most items in LO1 and LO3 (items 1–3, 7, 8; Figure 4: A, C). Knowledge gaps (incorrect answers on posttests) are mostly due to a mastery gaps rather than reversion, except for the most conceptually difficult items (Figure 4: C, D). Item 2 showed the largest learning (41%). Items 3 and 7 showed the largest amount of preknowledge by far (39% and 40%, respectively). Item 4 shows the largest mastery gap percentage among all items (46%).

Taken together, these data provide evidence that the PACRA can provide valid and reliable data that can be used to investigate many instructional questions including student misconception prevalence, normalized learning gains, and persistence/resolution of student misconceptions, as well as can provide data in support of evidence-based instructional strategies.

○ Discussion

Although genetics plays an increasingly important role in modern life, it has traditionally been difficult to learn (Karagoz & Cakir,

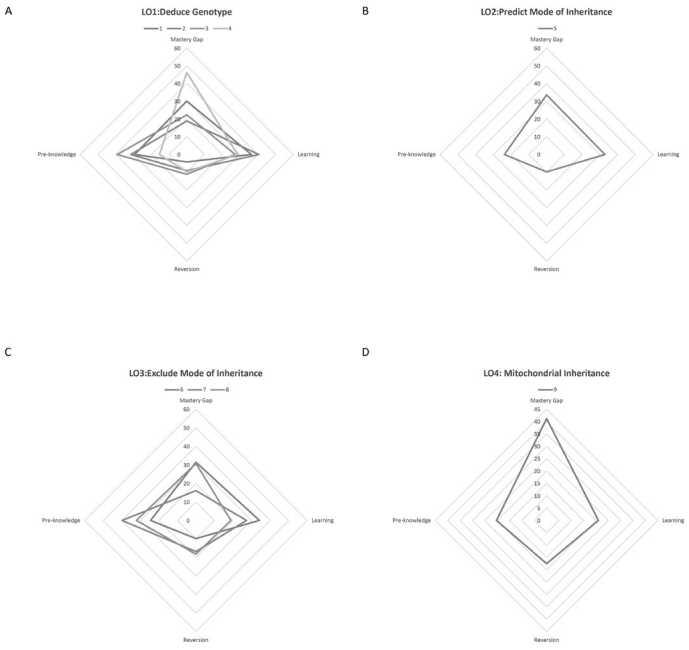


Figure 4. Mastery analysis by cross tabulation. Paired data (n = 143) were organized by the student into one of four groups. The percent of each group is presented in the radial diagram by learning objective and item. Mastery Gap indicates an incorrect response on both the pre- and postassessment. Reversion indicates a correct response in the preassessment and incorrect response on the postassessment. Learning indicates an incorrect response on the preassessment and correct response on the postassessment. Preknowledge indicates correct responses both pre- and postassessment. Shapes that stretch wide (to the right and left) represent mastery, either due to preknowledge (correct/correct) or learning (incorrect/correct), while shapes that stretch tall and long represent nonproficiency, either due to mastery gap (incorrect/incorrect) or reversion (correct/incorrect).

2011; Knippels, 2002; McElhinny et al., 2014; Tibell & Rundgren, 2010). Despite the urgent need to understand genetics, calls for biology education reform, and clear evidence for the value of particular teaching practices (Freeman et al., 2014), change has been slow (Brewer & Smith, 2009). Some barriers include lack of training and time (Brownell & Tanner, 2012) and prioritization of personal experience over education research findings (Andrews & Lemons, 2015). While larger assessments that include concepts across the genetics curriculum are important for evaluating student learning over a course or curriculum (Bowling et al., 2008; Smith et al., 2008), these tools lack the specificity and depth needed to inform instructors about specific concepts. Assessments focused on a distinct concept have great utility for an individual instructor to investigate student learning and, over time, to explore and modify personal teaching practices regarding a distinct concept to find a reasonable integration of student learning and personal teaching style. With this in mind, we developed and tested a short assessment that is easily deployed and can provide empirical evidence regarding the utility of teaching practices and insight into student misconceptions.

This assessment, a CRA that is focused on pedigree analysis, is one of many recently published inventories designed to investigate specific learning targets (Abraham et al., 2014; Anderson et al., 2002; Kalas et al., 2013; Price et al., 2014; Stefanski et al., 2016). The final pedigree analysis CRA was iteratively developed with faculty and student input. Psychometric analysis demonstrated that the CRA distinguishes between low- and high-performing students and that data obtained from its use can be considered reliable and valid in a variety of college classrooms (Table 2). Additionally, we report a new method that our group recently developed (Wasendorf et al., 2022) to demonstrate distractor efficacy, which is an indication that multiple-choice distractors accurately capture the misconceptions for which they were designed. Our congruence analysis both supports and is supported by the traditional psychometric measures of item discrimination, KR-20, and point biserial (Ebel, 1954). Based on item response theory (Baker, 2001), and through paired pre- and posttest data analyses, the CRA demonstrated utility both as a tool for assessing student conceptual understanding of pedigree analysis and as a tool for interrogating student learning postinstruction.

In this study, we further investigated student difficulties regarding pedigree analysis using mastery analysis and distractor progression. For mastery analysis, two of the items (4, 9) show mastery gap percentages above 40% and four of nine items show mastery gaps above 30%. Item 4, which requires students to use evidence from multiple generations rather than a single symbol or family unit to determine genotype, shows the largest gap in mastery among all items (46% of students). This problem, which is also found in secondary students' reasoning (Timm et al., 2022), confirms that the ability to follow and integrate genetic transmission across generations into an argument is a persistent difficulty. Item 9, which requires students to identify a pedigree consistent with mitochondrial inheritance, supports another persistent problem related to the nature of different chromosome types, such as problems in distinguishing X-linked inheritance previously observed (Hackling, 1994). Items 7–9, which are conceptually more difficult items, show the highest reversion, suggesting that student knowledge for these items is still fluid. Mastery analysis combined with NLG show that learning was observed for most learning objectives with items 7–8 (to provide evidence to refute a hypothesis, also known as falsification [Hackling & Lawrence, 1988]) and with item 9 (mitochondrial inheritance) showing the smallest to no significant amount of learning. These items are necessarily good targets for altering instruction because they have both a high mastery gap and high reversion. These analyses also suggest that current instruction is reasonably sufficient for high-preknowledge/high-learning items (1–3, 5–7); however, there is room for improvement. Additionally, distractor progression suggests important targets for instruction. In particular, two distractors have a positive slope (one distractor each for items 4 and 8). Item 4 has a distinctly positive slope for the correct answer, so students are likely moving from one distractor (population genetics, misreading, nature of sex chromosomes) to errors related to transmission genetics. Item 8, on the other hand, has a nearly zero slope for the correct answer in combination with an increase in the error related to conflating X-linked dominance and recessiveness. When these distractors are examined closely, we can see that they require the student to not only use information about transmission of alleles from multiple generations but also to use logic to eliminate allele options. It is reasonable that students who had learned to interpret pedigree symbols as indicating a phenotype, but not learned how to use them to follow allele transmission, may have selected these incorrect answers postinstruction.

Deeper analysis of student rationale provided common themes particularly related to the use of symbols across a variety of scales, inappropriate application of mathematical concepts, and logic errors. It also included such categories as mistaken terminology, inability to read a pedigree chart, difficulties with transmission genetics concepts, conflation of population genetics with pedigrees, and issues related to the nature and relationships of genes, alleles, chromosomes, and sex chromosomes. First, terminology, the ability to gather the appropriate information, and the symbolic nature of pedigree analysis were major challenges for students and represent major, crosscutting difficulties with many concepts in genetics. The inability to either follow and/or describe transmission of alleles was also a common issue. Students often used mathematical concepts inappropriately, such as frequency as evidence of dominant disorders, and used reasoning associated with population-level genetics rather than individual probability of inheritance as rationale. Students' answers also revealed difficulties understanding the nature and inheritance differences in autosomes, X chromosomes, and mitochondrial chromosomes, as well as basic allele-gene-chromosome relationships. We also frequently observed mistakes related to the nature of mitochondrial inheritance, that is, the conflation of mitochondrial inheritance with X-linked inheritance and the conceptual error that mitochondrial disorders only affect females. Finally, a more subtle and sophisticated problem occurred when focused too narrowly, such as looking at a single generation or family, when a deeper investigation of the data and use of logic were called for to correctly answer the question.

Looking forward, this CRA provides valid and reliable data and can be used to inform instruction and develop active learning modules specifically related to misconceptions surrounding pedigree analysis. Confronting misconceptions is particularly important because student misconceptions persist after new knowledge is accrued through passive learning and without reflection (Engelmann & Huntoon, 2011; Tanner & Allen, 2005). If misconceptions are left uncorrected, they lead to poor performance and misunderstanding of higher-level concepts (Engelmann & Huntoon, 2011; Fotou & Abrahams, 2016; Kalinowski et al., & Snodgrass, 2012), and in the long-term they can negatively impact degree attainment and career success.

Also, by examining the breakdown of pretest distractors and correct answers by learning objective over multiple semesters or by student year in school, instructors can better understand larger conceptual difficulties or redundancy in instruction that is best addressed at the curriculum level. In summary, this work provides an assessment tool for pedigree analysis and an accompanying psychometric analysis that demonstrates its reliability and validity. The assessment provides a tool for powerful and reliable inferences related to initial student understanding and misconceptions as well as postinstruction learning gains, guiding instruction in near real time, providing evidence for the relative effectiveness of various teaching practices, and possibly guiding curriculum reform.

○ Acknowledgements

This project was supported in part by a National Science Foundation grant to Nancy Boury and Rebecca Seipelt-Thiemann (NSF Award Number 1710277) and is based on work supported by the National Institute of Food and Agriculture, USDA, Hatch project IOW04008, Accession # 1010715. A portion of student work was supported by the Middle Tennessee State University Mathematics and Science Education PhD Program. We thank faculty who provided valuable faculty feedback and/or also participated in concept inventory testing. Specifically, Heather Wilson Ashworth and Eric Domyan (Utah Valley University), Joanna Klein (University of Northwestern St. Paul), Robyn Puffenbarger (Bridgewater College), Jelena Kraft, Steven Rodermel, Erik Vollbrecht (Iowa State University), Jackie Wisinski (University of Wisconsin at La Crosse), and Moh Salem and Ashley Morris (Middle Tennessee State University).

References

Abraham, J. K., Perez, K. E., & Price, R. M. (2014). The dominance concept inventory: A tool for assessing undergraduate student alternative conceptions about dominance in Mendelian and population genetics. *CBE-Life Sciences Education*, *13*(2), 349–58.

Adams, W., & Wieman, C. (2011). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education*, 1289–1312.

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching*, *39*(10), 952–78. https://doi.org/10.1002/tea.10053.

Andrews, T. C., & Lemons, P. P. (2015). It's personal: Biology instructors prioritize personal evidence over empirical evidence in teaching decisions. *CBE—Life Sciences Education*, *14*(1), ar7.

Bahar, M., Johnstone, A., & Hansell, M. (1999). Revisiting learning difficulties in biology. *Journal of Biological Education*, *33*(2), 84–86.

Baker, F. B. (2001). The basics of item response theory. ERIC.

Battista, R., Blancquaert, I., Laberge, A.-M., Van Schendel, N., & Leduc, N. (2012). Genetics in health care: An overview of current and emerging models. *Public Health Genomics*, *15*(1), 34–45.

Batzli, J. M., Knight, J. K., Hartley, L. M., Maskiewicz, A. C., & Desy, E. A. (2016). Crossing the threshold: Bringing biological variation to the foreground. *CBE—Life Sciences Education*, *15*(4), es9.

Berry, L., Jr. (1991). *Collaborative learning: A program for improving the retention of minority students*. https://eric.ed.gov/?id=ED384323

Bowling, B. V., Acra, E. E., Wang, L., Myers, M. F., Dean, G. E., Markle, G. C., Moskalik, C. L., & Huether, C. A. (2008). Development and evaluation of a genetics literacy assessment instrument for undergraduates. *Genetics*, *178*(1), 15–22. https://doi.org/10.1534/genetics.107.079533.

Braxton, J. M., Jones, W. A., Hirschy, A. S., & Hartley III, H. V. (2008). The role of active learning in college student persistence. *New Directions for Teaching and Learning*, *2008*(115), 71–83.

Brewer, C. A., & Smith, D. (2009). *Vision and change in undergraduate biology education*. American Association for the Advancement of Science.

Brownell, S. E., & Tanner, K. D. (2012). Barriers to faculty pedagogical change: Lack of training, time, incentives, and…tensions with professional identity? *CBE-Life Sciences Education*, *11*(4), 339–46.

Çimer, A. (2012). What makes biology learning difficult and effective: Students' views. *Educational Research and Reviews*, 7(3), 61.

Cohen, R. J., & Swerdlik, M. (2018). *Psychological testing and assessment* (9th ed.). McGraw-Hill Education.

Cyranoski, D. (2019). The CRISPR-baby scandal: What's next for human gene-editing. *Nature*, *566*, 440–42. https://doi.org/10.1038/d41586-019-00673-1

D'Avanzo, C. (2008). Biology concept inventories: Overview, status, and next steps. *BioScience*, *58*(11), 1079–85.

Ding, L., Chabay, R., Sherwood, B., & Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical Review Special Topics—Physics Education Research*, *2*(1), 010105.

Doran, R. L. (1980). *Basic measurement and evaluation of science instruction*. ERIC.

Ebel, R. L. (1954). Procedures for the analysis of classroom tests. *Educational and Psychological Measurement*, *14*(2), 352–64.

Engelmann, C. A., & Huntoon, J. E. (2011). Improving student learning by addressing misconceptions. *Eos, Transactions American Geophysical Union*, *92*(50), 465–66.

Erlbaum, M. R. L. (1999). Test theory: A unified treatment. [Google Scholar].

Findley, W. G. (1956). A rationale for evaluation of item discrimination statistics. *Educational and Psychological Measurement*, *16*(2), 175–80.

Fotou, N., & Abrahams, I. (2016). *Students' ideas in novel situations: Misconceptions or fragmented pieces of knowledge?*. Paper presented at conference. New Perspectives in Scienze Education.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–15.

Gates, S. J., & Mirkin, C. (2012). Engage to excel. *Science*, *335*(6076), 1545.

Genetics Society of America Education Committee. (2015). *Genetics learning framework*. https://genetics-gsa.org/education/genetics-learning-framework.

Gericke, N. M., & Hagberg, M. (2010). Conceptual variation in the depiction of gene function in upper secondary school textbooks. *Science & Education*, *19*(10), 963–94.

Gibson, G. (2019). Going to the negative: Genomics for optimized medical prescription. *Nature Reviews Genetics*, *20*(1), 1–2. https://doi.org/10.1038/s41576-018-0061-7.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist, 18*(8), 519.

Hackling, M. W. (1994). Application of genetics knowledge to the solution of pedigree problems. *Research in Science Education*, *24*(1), 147–55.

Hackling, M. W., & Lawrence, J. A. (1988). Expert and novice solutions of genetic pedigree problems. *Journal of Research in Science Teaching*, *25*(7), 531–46.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64–74.

Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning*, *7*(2), 75–83.

Kalas, P., O'Neill, A., Pollock, C., & Birol, G. (2013). Development of a meiosis concept inventory. *CBE-Life Sciences Education*, *12*(4), 655–64.

Kalinowski, S. T., Andrews, T. M., Leonard, M. J., & Snodgrass, M. (2012). Are Africans, Europeans, and Asians different "races"?: A guided-inquiry lab for introducing undergraduate students to genetic diversity and preparing them to study natural selection. *CBE-Life Sciences Education*, *11*(2), 142–51.

Karagoz, M., & Cakir, M. (2011). Problem solving in genetics: Conceptual and procedural difficulties. *Educational Sciences: Theory and Practice*, *11*(3), 1668–74.

Knight, J. (2010). Biology concept assessment tools: Design and use. *Microbiology Australia*, *31*(1), 5–8.

Knippels, M.-C. P. J. (2002). *Coping with the abstract and complex nature of genetics in biology education: The yo-yo learning and teaching strategy*.

Kubiszyn, T., & Borich, G. (1987). *Educational testing and measurement*. Scott, Foresman.

Lewis, J., & Kattmann, U. (2004). Traits, genes, particles and information: Re visiting students' understandings of genetics. *International Journal of Science Education*, *26*(2), 195–206.

McCowan, R. J., & McCowan, S. C. (1999). Item analysis for criterion-referenced tests. *Online Submission*.

McElhinny, T. L., Dougherty, M. J., Bowling, B. V., & Libarkin, J. C. (2014). The status of genetics curriculum in higher education in the United States: Goals and assessment. *Science & Education*, *23*(2), 445–64.

McKagan, S., Sayre, E., & Madsen, A. (2017). Normalized gain: What is it and when and how should I use it?. https://www.physport.org/recommendations/Entry.cfm?ID=93334.

Newman, D. L., Coakley, A., Link, A., Mills, K., & Wright, L. K. (2021). Punnett squares or protein production?: The expert–novice divide for conceptions of genes and gene expression. *CBE—Life Sciences Education*, *20*(4), ar53.

Niebert, K., & Gropengiesser, H. (2015). Understanding starts in the mesocosm: Conceptual metaphor as a framework for external representations in science teaching. *International Journal of Science Education*, *37*(5–6), 903–33.

Onwuegbuzie, A. J., & Daniel, L. G. (1999). *Uses and misuses of the correlation coefficient*. ERIC.

Paustian, T. D., Briggs, A. G., Brennan, R. E., Boury, N., Buchner, J., Harris, S., … Massimelli, M. J. (2017). Development, validation, and application of the microbiology concept inventory. *Journal of Microbiology & Biology Education*, *18*(3).

Phillips, A. M. (2016). Only a click away—DTC genetics for ancestry, health, love…and more: A view of the business and regulatory landscape. *Applied & Translational Genomics*, *8*, 16–22. https://doi.org/10.1016/j.atg.2016.01.001.

Price, R. M., Andrews, T. C., McElhinny, T. L., Mead, L. S., Abraham, J. K., Thanukos, A., & Perez, K. E. (2014). The genetic drift inventory: A tool for measuring what advanced undergraduates have mastered about genetic drift. *CBE-Life Sciences Education*, *13*(1), 65–75.

Pruitt, S. L. (2014). The next generation science standards: The features and challenges. *Journal of Science Teacher Education*, *25*(2), 145–56.

Rago, C., Vogelstein, B., & Bunz, F. (2007). Genetic knockouts and knockins in human somatic cells. *Nature protocols*, *2*(11), 2734. https://www.nature.com/articles/nprot.2007.408.pdf.

Ramorogo, G., & Wood-Robinson, C. (1995). Botswana1 children's understanding of biological inheritance. *Journal of Biological Education*, *29*(1), 60–71.

Resta, R. G. (1994). Historical origins of pedigree charts. *Journal of genetic counseling*, *3*(2), 165–67.

Sall, J., Stephens, M. L., Lehman, A., & Loring, S. (2017). *JMP start statistics: A guide to statistics and data analysis using JMP*. Sas Institute.

Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Clark, K., … Hefferon, T. (2019). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *47*(Database issue), D23.

Smith, M., Wood, W., & Knight, J. (2008). The genetics concept assessment: A new concept inventory for gauging student understanding of genetics. *CBE Life Sciences Education*, *7*(4), 422–30. https://doi.org/10.1187/cbe.08-08-0045.

Stefanski, K. M., Gardner, G. E., & Seipelt-Thiemann, R. L. (2016). Development of a lac operon concept inventory (LOCI). *CBE-Life Sciences Education*, *15*(2), ar24.

Stewart, J., & Kirk, J. V. (1990). Understanding and problem-solving in classical genetics. *International Journal of Science Education*, *12*(5), 575–88.

Sutton, C. (1996). Beliefs about science and beliefs about language. *International Journal of Science Education*, *18*(1), 1–18.

Tanner, K., & Allen, D. (2005). Approaches to biology teaching and learning: understanding the wrong answers—teaching toward conceptual change. *Cell Biology Education*, *4*(2), 112–17.

Tibell, L. A., & Rundgren, C.-J. (2010). Educational challenges of molecular life science: Characteristics and implications for education and research. *CBE—Life Sciences Education*, *9*(1), 25–33.

Timm, J., Wools, K., & Schmiemann, P. (2022). Secondary students' reasoning on pedigree problems. *CBE—Life Sciences Education*, *21*(1), ar14.

Wasendorf, C., Reid, J. W., Seipelt-Thiemann, R., Grimes, Z. T., Couch, B., Peters, N. T., … Boury, N. (2022). The development and validation of the mutation criterion referenced assessment (MuCRA). *Journal of Biological Education*. In press.

In the Math and Science Education Program at Middle Tennessee State University, Murfreesboro, ZACHARY T. GRIMES (zgrimes@crc.edu), JOSHUA W. REID (Joshua.reid@mtsu.edu), OLENA JAMES, and BROCK COUCH are doctoral students, and REBECCA L. SEIPELT-THIEMANN (rebecca.seipelt@mtsu.edu) is a full professor (and a full professor in the Biology Department).

At Iowa State University, Ames, NANCY M. BOURY (nan1@iastate.edu) is an associate professor in the Department of Plant Pathology, Entomology, and Microbiology. CHLOE WASENDORF is a doctoral student in the Interdepartmental Microbiology Graduate Program. AUDREY L. MCCOMBS is a doctoral student in the Department of Statistics, Ecology and Evolutionary Biology Program. PATRICK I. ARMSTRONG is an associate professor in the Department of Psychology.