Article

The Genetic Drift Inventory: A Tool for Measuring What Advanced Undergraduates Have Mastered about Genetic Drift

Rebecca M. Price,* **Tessa C. Andrews,[†]** **Teresa L. McElhinny,[‡]** **Louise S. Mead,[§]**
Joel K. Abraham,[∥] **Anna Thanukos,[¶]** **and Kathryn E. Perez[#]**

*School of Interdisciplinary Arts and Sciences, University of Washington, Bothell, Bothell, WA 98011; [†]Department of Genetics, University of Georgia, Athens, GA 30602; [‡]Department of Zoology and [§]BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI 48824; [∥]Department of Biological Science, California State University, Fullerton, Fullerton, CA 92831; [¶]University of California Museum of Paleontology, Berkeley, CA 94720; [#]Department of Biology, University of Wisconsin–La Crosse, La Crosse, WI 54601

Understanding genetic drift is crucial for a comprehensive understanding of biology, yet it is difficult to learn because it combines the conceptual challenges of both evolution and randomness. To help assess strategies for teaching genetic drift, we have developed and evaluated the Genetic Drift Inventory (GeDI), a concept inventory that measures upper-division students' understanding of this concept. We used an iterative approach that included extensive interviews and field tests involving 1723 students across five different undergraduate campuses. The GeDI consists of 22 agree–disagree statements that assess four key concepts and six misconceptions. Student scores ranged from 4/22 to 22/22. Statements ranged in mean difficulty from 0.29 to 0.80 and in discrimination from 0.09 to 0.46. The internal consistency, as measured with Cronbach's alpha, ranged from 0.58 to 0.88 across five iterations. Test–retest analysis resulted in a coefficient of stability of 0.82. The true–false format means that the GeDI can test how well students grasp key concepts central to understanding genetic drift, while simultaneously testing for the presence of misconceptions that indicate an incomplete understanding of genetic drift. The insights gained from this testing will, over time, allow us to improve instruction about this key component of evolution.

INTRODUCTION

Few investigators have explored how undergraduate students understand nonadaptive evolutionary processes, partly because of the lack of tools to assess student knowledge of

processes such as mutation, genetic drift, and gene flow. Genetic drift seems to be particularly difficult to learn because it combines the conceptual challenges of evolution with the conceptual challenges of understanding randomness. It proves to be confusing for many students (Andrews *et al.*, 2012). However, both evolution and randomness—and by extension thinking about probability—are featured as learning goals in the American Association for the Advancement of Science *Vision and Change* report (AAAS, 2011).

Genetic drift is taught in major introductory biology textbooks (e.g., Freeman, 2005; Campbell *et al.*, 2008) in the context of the Hardy-Weinberg equilibrium, a population genetics model that assumes that no evolution has occurred and, consequently, that the frequency of each allele remains constant. Hardy-Weinberg equilibrium is a theoretical situation in which allelic frequencies in a population do not change because the population has an infinite size, is isolated, has

individuals that exhibit no preference for mates, and does not experience natural selection or mutation. Genetic drift is associated with relaxing the mathematical assumption that the population size is infinite; all actual populations are finite and are therefore subject to genetic drift because of random pre- and postzygotic fluctuations in allelic frequency.

Understanding genetic drift is crucial for establishing a comprehensive understanding of biology (Gould and Lewontin, 1979) in fields as diverse as conservation, molecular evolution, and paleobiology. Imperiled populations have small effective population sizes, so the effects of random sampling are magnified (Masel, 2012). Over time, genetic drift reduces the amount of genetic variation within populations, limiting the potential for adaptive evolution (Freeman and Herron, 2004). Genetic drift also tends to increase the genetic distinction among populations, potentially leading to evolutionarily significant units that require independent protection (Mills, 2007). Genetic drift is the theoretical framework for Kimura's highly influential neutral theory of molecular evolution, which is important for understanding many aspects of molecular biology, bioinformatics, and genetics (Masel, 2012). In paleobiology, understanding genetic drift is essential for distinguishing active macroevolutionary trends driven by natural selection from passive trends driven by random processes (Raup *et al.*, 1973; McShea, 1994).

Despite the importance of the concept of genetic drift across biology, we know little about how students learn genetic drift other than that it is a challenging concept fraught with misconceptions (Andrews *et al.*, 2012). Genetic drift combines two topics that are notoriously difficult in and of themselves: evolution and randomness (Garvin-Doxas and Klymkowsky, 2008; Mead and Scott, 2010). The challenges students face when learning evolution have been studied primarily in the context of natural selection (reviewed by Gregory, 2009) and include thinking that need or desire drives evolutionary change (i.e., teleological thinking) and a tendency to attribute all evolution to natural selection (Hiatt *et al.*, 2013). Similarly, probability and randomness perplex people of all ages (Fischbein and Schnarch, 1997; Mlodinow, 2008). When considering random evolutionary processes, students are challenged by both the terminology (Kaplan *et al.*, 2010; Mead and Scott, 2010) and conceptual complexities (Garvin-Doxas and Klymkowsky, 2008). For example, in student interviews conducted to identify students' ideas about genetic drift, students repeatedly stated that a random process could not account for any directional evolutionary change (Andrews *et al.*, 2012). Furthermore, they found it challenging to recognize that many different processes could cause a random change in allelic frequency (Andrews *et al.*, 2012). Biology students often have a weak understanding of mathematics (Maloney, 1981; Jungck, 1997) that makes it challenging to teach processes that have a random element (Garvin-Doxas and Klymkowsky, 2008; Kaplan *et al.*, 2010). However, probabilistic reasoning is necessary to develop a scientifically accurate understanding of all of the core concepts of the AAAS *Vision and Change* report (AAAS, 2011).

Few evidence-based instructional strategies exist to teach genetic drift, despite its foundational nature. To assess attempts to address this absence, instructors and researchers require tools to measure student learning of genetic drift. This paper describes the development and evaluation of the Genetic Drift Inventory (GeDI), which consists of 22 agree–disagree statements and is intended for upper-division biology majors. The GeDI can be used to assess and therefore improve instruction about genetic drift. As with other biological concept inventories (e.g., Anderson *et al.*, 2002; Smith *et al.*, 2008; Fisher *et al.*, 2011; Perez *et al.*, 2013), this instrument provides a way for researchers to measure student learning and identify which aspects of genetic drift challenge students. The GeDI is one of a growing number of concept inventories assessing evolutionary concepts, including the Concept Inventory of Natural Selection (Anderson *et al.*, 2002), the EvoDevoCI (Perez *et al.*, 2013), the Dominance Concept Inventory (Abraham *et al.*, 2014), and others (Baum *et al.*, 2005; Cotner *et al.*, 2010; Nadelson and Southerland, 2010; Novick and Catley, 2012).

METHODS

We used a multistep, iterative process of writing, testing, and revising to develop the GeDI (Table 1). To avoid redundancy, we describe the major components of the development below and restrict presentation of their chronological order to Table 1.

Format

The GeDI is a concept inventory composed of true–false statements. This format minimizes test wiseness compared with a multiple-choice test because students evaluate each statement separately instead of using the process of elimination to select an answer (Frisbie, 1992). Students often have both scientifically accurate and inaccurate ideas about evolution

Table 1. Overview of the multistep process used to develop the GeDI

1. Identified and described students' ideas about genetic drift (see Andrews *et al.,* 2012) and reviewed the literature on common misconceptions about the broader topics of randomness and evolution.
2. Reviewed evolution textbooks and consulted with experts ($n = 86$) to identify a list of key concepts biology undergraduates should understand about genetic drift.
3. Developed and administered a pilot multiple-choice instrument (GeDI-Draft 1) based on known misconceptions and identified key concepts to evaluate difficulty ($n = 136$) and the clarity and appropriateness of the language ($n = 161$).
4. Reworded jargon and unfamiliar phrasing, revised items with low difficulty, and reformatted to true–false format, creating GeDI-Draft 2.
5. Administered and revised GeDI-Draft 2 by administering to students ($n = 85$) to evaluate difficulty and discrimination, completing student interviews ($n = 21$), and getting expert input ($n = 7$).
6. Revised and eliminated statements, creating GeDI-Draft 3.
7. Administered GeDI-Draft 3 to students in upper-division biology courses ($n = 593$) at four institutions to evaluate difficulty and discrimination, interviewed students ($n = 15$) to assess construct validity, and administered to experts ($n = 21$) to make student-to-expert comparison.
8. Eliminated poorly performing and confusing statements, resulting in the GeDI 1.0.
9. Assessed difficulty, discrimination, and internal reliability of GeDI 1.0 with large-scale administration ($n = 661$) and determined test–retest reliability ($n = 51$).
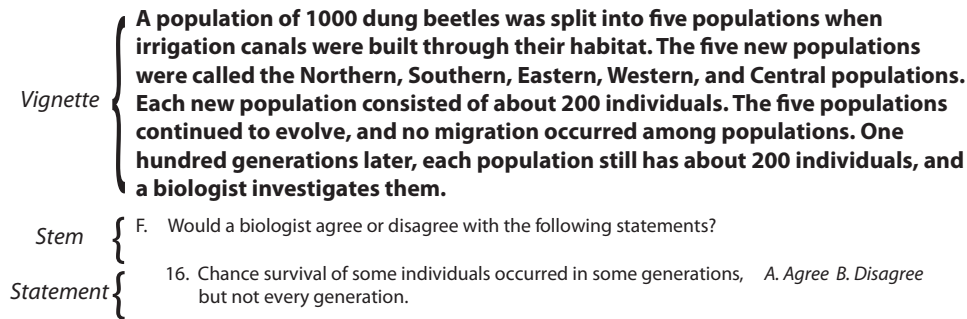
Vignette {
A population of 1000 dung beetles was split into five populations when irrigation canals were built through their habitat. The five new populations were called the Northern, Southern, Eastern, Western, and Central populations. Each new population consisted of about 200 individuals. The five populations continued to evolve, and no migration occurred among populations. One hundred generations later, each population still has about 200 individuals, and a biologist investigates them.

Stem {
F.   Would a biologist agree or disagree with the following statements?

Statement {
16.   Chance survival of some individuals occurred in some generations,   *A. Agree   B. Disagree*
but not every generation.

Figure 1.   Structure of the GeDI. Students evaluate a series of statements that follow a question stem about a scenario presented in a vignette.

(Andrews *et al.*, 2012; see also "heterogeneous understanding" of Nehm and Schonfeld, 2008), so they may consider more than one answer choice in a multiple-choice question accurate (Parker *et al.*, 2012), yet they are forced to choose just one answer choice. Because of this, we miss the opportunity to gather data about a student's ideas regarding each choice in a multiple-choice question. This missed opportunity is particularly unfortunate in concept inventories because they are specifically designed so that the incorrect answer choices express nonscientific ideas commonly held by students (D'Avanzo, 2008; Smith and Tanner, 2010). Concept inventories with true–false statements, in which students must evaluate each statement, have the potential to better capture the diversity, consistencies, and contradictions within students' ideas.

The nomenclature for a true–false test is not standardized, so we define our terms here and illustrate them in Figure 1. Our instrument is arranged as a series of *vignettes*, short stories that contextualize the items. One or more *items*, each of which is composed of a question stem and at least one true–false *statement*, follow each vignette. Although we describe this instrument as a true–false format to emphasize the tradition from which we work, the instrument asks students to determine whether a biologist would agree or disagree with each statement, rather than asking students to determine whether a statement is true or false. This phrasing follows the tradition of the Concept Inventory of Natural Selection, which asks students to "Choose the one answer that best reflects how an evolutionary biologist would answer" (Anderson *et al.*, 2002, p. 974). The wording emphasizes that we are assessing students' understanding of biological concepts, rather than their personal opinions.

Student Participants

Our objective was to create an instrument that would be a useful tool for instruction and education research with upper-division biology students in different institution types and in different courses. Therefore, we broadly sampled from five different campuses around the country to test and revise the instrument. At every step of the process (Table 1), we also made a conscious effort to include a diversity of students within institutions and within targeted courses. We sampled students from different college-achievement levels, ethnicities, races, and genders. Table 2 presents the institutions and student populations used for each stage of the development and revision process. All of the research was conducted with the approval of the appropriate Institutional Review Boards (California State University, Fullerton: HSR-12-0432; Michigan State University: i040365; University of Georgia: 2013-10134-0; University of Washington: 42505; University of Wisconsin–La Crosse: approved, no number assigned).

Identifying Key Concepts

To be useful for college evolution instructors, our instrument needs to address key concepts commonly taught in evolution courses and to match what an evolutionary biologist expects a biology undergraduate to know about genetic drift. We created an initial list of key concepts that we expected undergraduates to understand about genetic drift, using textbooks written for college evolution courses (e.g., Freeman and Herron, 2004; Barton *et al.*, 2007; Futuyma, 2009). In doing so, we recognized that experts define genetic drift as both a process (e.g., sampling error in the production of zygotes from a gene pool) *and* as the pattern resulting from a process (e.g., changes in the frequencies of alleles in a population resulting from sampling error). The fluidity with which experts can think about genetic drift as both a process and a pattern is illustrated by the fact that a single textbook described genetic drift both ways in different parts of the book (e.g., Freeman and Herron, 2004). However, using the term *genetic drift* to refer to both a pattern and a process within one assessment tool has the potential to confuse students. Therefore, we have written our items to be consistent, with genetic drift defined as a *process*. We made this decision for the sake of clarity, not to promote a single definition of genetic drift.

After generating an initial list of key concepts, we asked evolution experts to evaluate these concepts against two criteria: 1) importance of the concept for undergraduate courses and 2) whether instructors emphasized the concept in their own teaching (Supplemental Material, Survey 1). We also solicited any key concepts the experts thought were missing from our list. We recruited evolutionary biologists through the National Evolutionary Synthesis Center and Evolution Directory (EvolDir) listservs, as well as by contacting colleagues. Eighty-six experts completed the survey, and 71 of these individuals taught genetic drift in a college biology course. Using the results of the expert survey, we revised our initial list of 24 key concepts into four main key concepts that are subdivided into a total of 10 subconcepts (see *Results*). As we revised the GeDI to target upper-division students, we narrowed the list of subconcepts from 10 to six.

Table 2. Institutions and students involved in validation

	Institutions	Participants[a]	*n*
GeDI-Draft 1			
Administration	Doctoral-granting institution—Midwest (DG-Mw)	Majors in 300-level genetics	136
Communication validity	DG-Mw	Nonmajors in 200-level integrated science	161
GeDI-Draft 2			
Administration	DG-Mw	Majors in 400-level evolution	85
Interviews	DG-Mw	Majors and nonmajors	7
	Doctoral-granting institution—Southeast (DG-Se)	Majors and nonmajors	6
	Minority-serving, master's-granting institution—West	Majors and nonmajors	5
	Primarily undergraduate institution—Northwest	Majors	3
GeDI-Draft 3			
Administration	DG-Mw	Majors in 300-level genetics	198
	DG-Se	Majors in 300-level genetics	262
	Master's comprehensive university—Midwest	Majors in two sections of 200-level zoology and majors in 400-level evolution	89
			44
Interviews	DG-Mw	Majors in Research Experience for Undergraduates program	15
GeDI 1.0			
Administration	Doctoral-granting institution—Northwest (DG-Nw)	Majors in 300-level cell biology	51
	DG-Nw	Majors in 300-level evolution	91
	DG-Se	Majors in 300-level genetics	318
	DG-Mw	Majors in 400-level evolution	60
	DG-Mw	Majors in 300-level genetics	141
Test–retest	DG-Nw	Majors in 400-level physiology	51

[a]Majors are students in the life sciences, and nonmajors are typically studying fields outside the sciences.

Identifying Misconceptions

The term *misconception* is widely used and often poorly defined. As a result, some researchers have called for the replacement of the term (Maskiewicz and Lineback, 2013). We contend that any other term will be just as fraught with ambiguity and misunderstanding, and so we favor the suggestion that researchers continue to use *misconception* and explicitly define what they mean (Crowther and Price, 2014; Leonard *et al.*, 2014). In this paper, we use the term *misconception* to mean an inappropriate or incomplete idea about a given concept that is commonly held by students. Misconceptions are not inherently inaccurate, but rather are inappropriately or incompletely applied to a given biological context in a way that is not aligned with our current scientific understanding of the world (Leonard *et al.*, 2014). Not all incorrect ideas play a role in constructing an accurate understanding of a topic. By limiting the application of the term misconception to incorrect ideas commonly held by students, we are focusing on ideas that are likely to be an important step on the path to more scientifically accurate understanding of a topic (e.g., Kampourakis and Zogza, 2009).

To create a list of common misconceptions about genetic drift, we consulted previous studies that identified and described common student ideas about genetic drift (Andrews *et al.*, 2012), natural selection (Gregory, 2009), and random patterns and processes in biology (Garvin-Doxas and Klymkowsky, 2008; Klymkowsky and Garvin-Doxas, 2008; Mead and Scott, 2010). We also identified additional misconceptions from student interviews conducted during the process of developing the instrument.

Not all of the misconceptions about genetic drift identified in previous work and in our interviews were appropriate for inclusion in the final instrument. We designed our instrument

for upper-division biology students, so misconceptions primarily expressed by introductory students were outside the scope of this instrument (for a developmental model of genetic drift expertise, see Andrews *et al.*, 2012). Upper-division students were consistently able to identify misconceptions that we initially expected would be common, so we removed statements assessing these misconceptions from the instrument. Our final instrument assesses student ideas about six misconceptions (Table 3).

Writing and Revising the GeDI

GeDI-Draft 1. We began the instrument construction process by writing, testing, and revising a multiple-choice instrument, the GeDI-Draft 1. We administered the GeDI-Draft 1 to upper-division biology students (Table 2). The instructor offered extra credit points in the course as incentive for students to participate. The major finding of this initial administration was that the instrument was too easy; a number of statements were answered correctly by >85% of the students in the course. We therefore substantially revised the instrument, including changing the format to true–false.

We evaluated communication validity of the GeDI-Draft 1 by administering it to 161 nonmajors enrolled in a biology course (Table 2). We divided the GeDI-Draft 1 in half; one group of 80 students took one half, and the remaining 81 students took the other half. Instead of taking the GeDI-Draft 1, these students evaluated each item by circling any words they did not know. We used their feedback to re-examine any words or phrases circled by four or more students. In some cases, we rephrased unfamiliar terms. For example, we changed "disadvantageous" to "harmful." In other cases, we retained unfamiliar terms because we considered them

Table 3. Common misconceptions about genetic drift

Misconception[a]	GeDI statement number[b]
About sampling error	
1. Genetic drift is unpredictable because it has a random component.*	7
Confusing genetic drift with natural selection	
2. Genetic drift is natural selection/ adaptation/acclimation to the environment that may result from a need to survive.	5 6 8
3. Genetic drift is not evolution because it does not lead to directional change that increases fitness.*	2
4. Natural selection is always the most powerful mechanism of evolution, and it is the primary agent of evolutionary change.*	9 12 17 20
Confusing genetic drift with other, nonselective mechanisms of evolution	
5. Genetic drift is random mutation.	14 19 22
6. Genetic drift is gene flow or migration.	11 18 21

[a]Some of these misconceptions were identified from interviews and are reported for the first time in this paper (*); the rest are from Andrews et al. (2012).

[b]See Supplemental Material, The Genetic Drift Inventory 1.0.

crucial to understanding genetic drift. For example, we retained "genetic drift" and "random survival," even though they were unfamiliar to some students. The outcome of these revisions was the GeDI-Draft 2, which included 56 statements.

GeDI-Draft 2. We evaluated and revised the 56-statement GeDI-Draft 2 by administering it to upper-division biology students, interviewing undergraduates, and gathering input from experts about the quality of the items. We administered the GeDI-Draft 2 to 85 upper-division biology students (Table 2) and evaluated each statement by calculating difficulty (i.e., how challenging a statement is for students) and discrimination (i.e., how well an item distinguishes between high- and low-performing students). We calculated statement difficulty (P) by dividing the number of correct responses by the number of responses for a particular statement to calculate the proportion of students choosing the correct response. We considered statements that most students answered correctly ($P > 0.80$) too easy for our target population. We calculated statement discrimination ($D = P_H - P_L$) by subtracting a statement's difficulty among the low performers, P_L, from a statement's difficulty among high performers, P_H (Doran, 1980; Haladyna, 2004). We designated students as low-performing if their total scores on the instrument were in the bottom percentile of all total scores. Similarly, high-performing students had total scores in the top percentile of all total scores. The percentiles designated as low- and high-performing were around 33 and 67%, but not always exactly these percentiles. Dividing students equally into thirds would result in some students

with the same score being in one group (e.g., low performers), while other students with the same score would be in another group (e.g., middle performers), so we divided students into three similarly sized groups using natural breaks between scores. Our calculations take this into account because discrimination is calculated by calculating the difference of two proportions (see above equation). We excluded statements that were insufficiently discriminating for our target population; these were statements for which the difference in difficulty between low- and high-performing students was 20% or less ($D < 0.20$; Crocker and Algina, 1986).

We examined the construct validity of the items on the GeDI-Draft 2, using student interviews. Construct validity examines whether an instrument behaves as we would expect it to behave. One way to examine construct validity is to determine whether students respond to an item as we would expect them to, given their knowledge of genetic drift. In other words, do students agree with a scientifically accurate statement because they understand the science? Similarly, do they agree with a misconception because they hold that misconception? Interviewers asked biology undergraduates to read and answer items and then to explain why they choose their answers for each statement. We interviewed 21 students at four institutions; each item was answered and explained by at least four undergraduates (Table 2).

We also examined the face validity of the GeDI-Draft 2. Face validity is the degree to which an instrument is judged by knowledgeable people to measure the concept it purports to measure (Neuman, 2009). Seven evolutionary biologists from five institutions evaluated the GeDI-Draft 2 (Supplemental Material, Survey 2). These experts evaluated the plausibility and clarity of each vignette and item stem. The experts also evaluated each statement that addressed a key concept on: 1) the extent to which the statement addressed the intended key concept, 2) the clarity and accuracy of the statement, and 3) whether it was reasonable for someone with a background in general biology to evaluate the statement correctly. We did not ask the experts to evaluate statements addressing misconceptions because those statements are not scientifically accurate and are often perceived as unclear to experts, who have a much more nuanced perspective than nonexpert students. The expert survey, student interviews, and test administration informed the revisions that led to the GeDI-Draft 3. Revisions at this stage resulted in a 40-statement instrument.

GeDI-Draft 3. The objective for our final round of revisions was to eliminate low-quality statements from the 40-statement GeDI-Draft 3 using a three-step process. First, we administered the GeDI-Draft 3 to 593 students enrolled in four different courses at three institutions (Table 2). We evaluated each of the 40 statements by calculating difficulty and discrimination. All of the courses in which we administered the assessment had already covered genetic drift when we tested students. We determined that 11 statements were either too easy ($P > 0.80$) or did not sufficiently discriminate between high- and low-performing students ($D < 0.2$), so we removed them.

We also flagged seven statements for special attention in student interviews because their quality could not be fully evaluated by calculating difficulty and discrimination. In some cases, difficulty and discrimination do not provide a sufficient assessment of statement quality. For example, a

statement that is highly difficult for both high- and low-performing students will have low discrimination. Such a statement might still be valuable because it represents a key genetic drift concept that undergraduates do not understand. On the other hand, such a statement may be difficult because students do not interpret it as we intended. We evaluated these statements by asking students about their responses to statements in interviews.

These interviews, like the earlier ones, allowed us to evaluate construct validity. We interviewed 15 biology undergraduates participating in a summer research experience (Table 2) and collected three to five student responses for each statement. We eliminated four of the seven statements that had been flagged because students did not interpret them as we intended. We retained the other three flagged statements (16, 18, and 21 in Supplemental Material, Genetic Drift Inventory 1.0), because students interpreted them accurately. We removed three more statements when interviews revealed that students commonly misunderstood them. This left 23 statements.

We then administered the 23-statement instrument to experts so that we could compare the performance between undergraduates (novices) and experts. This provided another test of construct validity because we would expect experts both to perform well on the instrument and to perform substantially better than undergraduates. We recruited experts by emailing listservs of researchers in biology education associated with Doctoral-Granting Institution Northwest and through word-of-mouth. We administered the GeDI to 21 individuals with expert knowledge of evolution: six biology and biology education graduate students, one postdoctoral fellow in biology education research, 11 faculty members who teach evolution, and three faculty members who teach other aspects of college biology. These experts did not necessarily teach or research the topic of genetic drift. We used a Welch's *t* test to compare the mean scores between experts and undergraduates. This analysis excludes one statement (19 in Supplemental Material, Genetic Drift Inventory 1.0) because we included the incorrect stem for one statement in the version of the instrument the experts completed. As expected, the mean score achieved by experts on the GeDI-Draft 3 was significantly higher ($87 \pm 8.0\%$ SD) than the mean score achieved by undergraduates ($58 \pm 17\%$ SD) who had received in-class instruction about genetic drift (Welch's $t = 15.00, p < 0.00001$). Administering the GeDI to experts also led to the exclusion of one statement that was difficult to understand. The resulting instrument has 22 statements and is called the GeDI 1.0. We expect that the GeDI 1.0 will continue to be revised to remain useful for the intended population and to be used for additional populations. This naming system allows for future versions (e.g., the GeDI 2.0).

Final Validation of the GeDI 1.0

The goal of the final stage in the development process was to demonstrate the quality and utility of the GeDI 1.0. We administered the GeDI 1.0 to a total of 661 students in five upper-division biology courses at three institutions (Table 2). We calculated difficulty (*P*) and discrimination (*D*) for each statement administered in each course. For both statement difficulty and discrimination, we fitted a mixed-effects model to test for differences between the way students an-swered statements that address misconceptions versus statements that address key concepts. Mixed-effects models are regression models that can account for data with a nested structure. The mixed-effects models we fitted allowed us to account for the lack of independence among scores from the same statement administered to students in different courses, as well as the lack of independence within the same course (Gelman and Hill, 2007).

We evaluated the reliability of the GeDI 1.0 in two ways. We examined test–retest reliability by administering the GeDI 1.0 to students in an upper-division physiology class. Students completed the GeDI 1.0 twice, with about 1 wk between test dates. We calculated test–retest reliability as the coefficient of stability, which is the correlation between the test and retest scores (Gliner *et al.*, 2009). We also calculated the internal consistency of the GeDI 1.0 for each of the five upper-division biology courses included in our final analysis. We calculated internal consistency as Cronbach's alpha.

We made one small grammatical change to the GeDI 1.0 after validation. In the scenario about nearsightedness, the word "small," which modifies "island," was moved from the third sentence to the second sentence. We made this change to improve clarity, but concluded that it is not substantial enough to necessitate additional validation at this time.

RESULTS

Key Concepts and Misconceptions

The results of the initial expert survey (Supplemental Material, Survey 1) helped us identify four main concepts that are required to achieve the level understanding of genetic drift that experts expect of biology majors at graduation (Table 4). These concepts focus on the effects of random sampling in each generation and the fact that genetic drift can result in nonadaptive evolution.

While interviewing students, we identified three misconceptions that have not been previously reported (Table 3). One of these misconceptions is that students conflate randomness with unpredictability (misconception 1 in Table 3). For example, a student stated,

> Because genetic drift is random, you won't know when the genes will drift until it's done, so you can't predict it.

The other two misconceptions relate to natural selection. One is the idea that directional change must result from an increase in fitness (misconception 3 in Table 3). In a student's language,

> It doesn't make sense for one of them [that is, one of the traits] to become fixed, because when I think of something becoming fixed in a population, it becomes that way because it's an advantage.

Note that this misconception is only subtly different from one that has already been reported, that random processes cannot account for directional evolutionary change (Andrews *et al.*, 2012). A related misconception is that natural selection is always the most powerful mechanism of evolution (misconception 4 in Table 3). For example, a typical student response was that the "population is not changing because there is no pressure for change."

Table 4. The concepts students should master for a complete understanding of genetic drift appropriate for undergraduates

Key concept[a]	GeDI statement number[b]
1. Random sampling error happens every generation, which can result in random changes in allele frequency that is called genetic drift.	
1a. Genetic drift results from random sampling error.	Not in GeDI
1b. Random sampling occurs each generation in all finite populations.	16
1c. Random sampling can result in random changes in allelic, phenotypic, and/or genotypic frequency.	Not in GeDI
2. Random sampling error tends to cause a loss of genetic variation within populations, which in turn increases the level of genetic differentiation among populations.	
2a. The processes leading to genetic drift tend to cause a loss of genetic variation within populations over many generations.	3 13
2b. Decreasing genetic variation within populations usually increases genetic differentiation among populations.	Not in GeDI
3. The magnitude of the effect of random sampling error from one generation to the next depends on the population size. The effect is greater when populations have a small effective size, but generally small or undetectable when effective population size is large.	
3a. The effects of genetic drift are larger when the population is smaller.	1
3b. Founding and bottlenecking events are two situations in which the effects of genetic drift are greater because the effective population size is rapidly reduced.	10
4. In populations with small effective sizes, genetic drift can overwhelm the effects of natural selection, mutation, and migration; therefore, an allele that is increasing in frequency due to selection might decrease in frequency some generations due to genetic drift.	
4a. Other evolutionary mechanisms, such as natural selection, mutation, and migration act simultaneously with genetic drift.	Not in GeDI
4b. The processes leading to genetic drift can overwhelm the effects of other evolutionary mechanisms.	15
4c. Random sampling error can result in populations that retain deleterious alleles or traits.	4

[a]Reasons for excluding some of the key concepts from the GeDI 1.0 are detailed in the *Discussion*.

[b]See Supplemental Material, The Genetic Drift Inventory 1.0.

GeDI 1.0

The GeDI 1.0 comprises four vignettes, nine items, and 22 statements (Supplemental Material, The Genetic Drift Inventory 1.0 and Answer Key). Each statement tests biology undergraduates' knowledge of a single key concept or misconception: 15 statements address misconceptions and seven statements address key concepts (Tables 3 and 4). Mean scores for each course ranged from 11.93 to 16.66 out of 22, and students' individual scores ranged from 4 to 22 across all administrations (Table 5).

The GeDI 1.0 produced reliable results for upper-division biology undergraduates. In our test–retest analysis with upper-division physiology students, the coefficient of stability was equal to 0.82. A coefficient of stability equal to 1 would indicate perfect reliability, and values equal to or greater than 0.80 are generally interpreted as indicating that an instrument is sufficiently reliable (Gliner *et al.*, 2009). The internal con-

sistency of the responses provided by undergraduates from five additional upper-division courses ranged from 0.58 to 0.88 using Cronbach's alpha (Table 5), which is similar to the internal consistency of other published instruments designed for instructional and research use with biology undergraduates (e.g., Conceptual Inventory of Natural Selection, alpha = 0.58–0.64, Anderson *et al.*, 2002; Test of Science Literacy Skills, alpha = 0.581–0.761, Gormally *et al.*, 2012; EvoDevoCI, alpha = 0.31–0.73, Perez *et al.*, 2013; Dominance Concept Inventory, alpha = 0.77, Abraham *et al.*, 2014).

The statements included in the GeDI 1.0 challenged upper-division biology students (Figure 2A). Mean statement difficulty ranged from $P = 0.29$ (statement 6) to $P = 0.80$ (statement 1), which means that, on average, only 29% of students answered the most difficult statement correctly, while 80% answered the least difficult statement correctly. Statements addressing misconceptions ($P = 0.58 \pm 0.18$ SD)

Table 5. Performance on the GeDI 1.0 across different courses, described in Table 2, and institutions

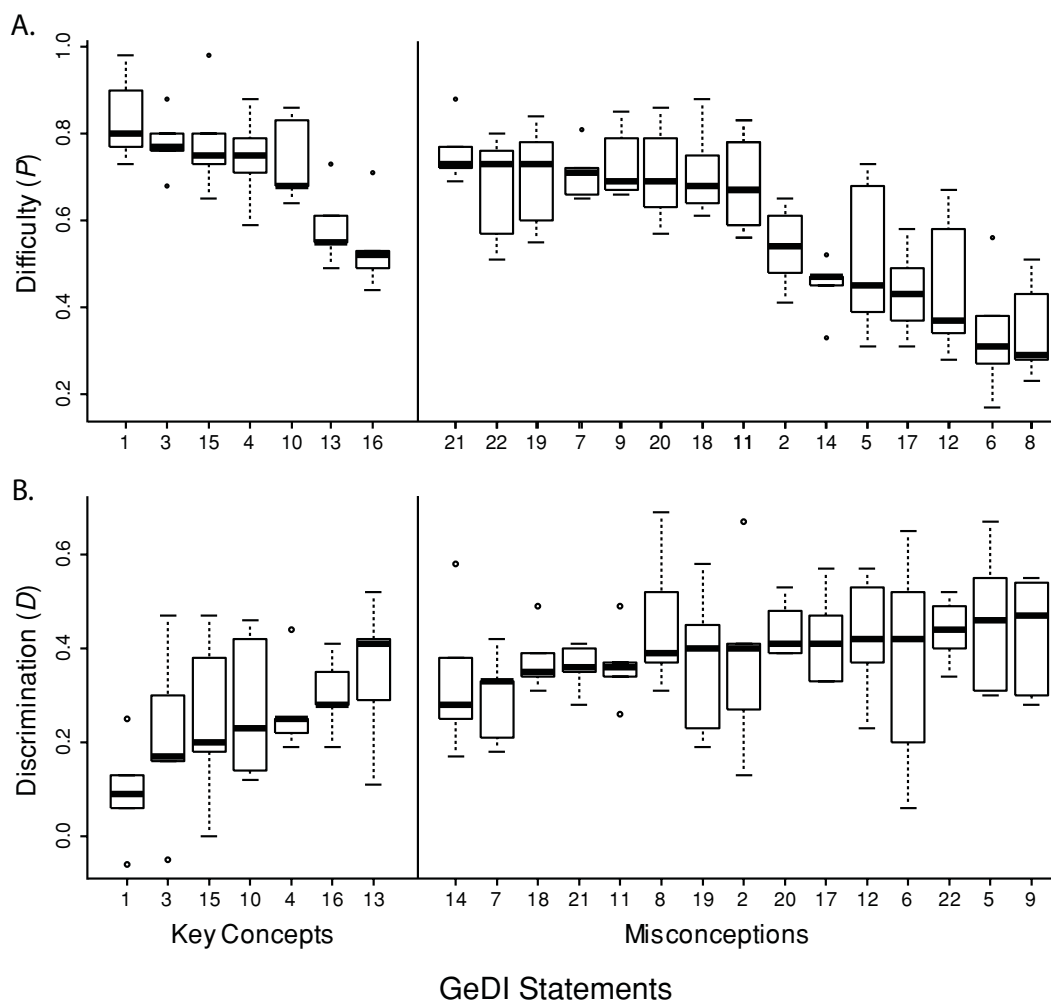|	300-level cell biology, DG-Nw	300-level evolution, DG-Nw	300-level genetics, DG-Se	400-level evolution, DG-Mw	300-level genetics, DG-Mw
Mean of items correct (SD)	13.35 (3.64)	14.47 (3.78)	12.35 (3.29)	16.66(3.44)	11.94 (3.35)
Range of items correct	6–22	8–22	4–22	7–22	4–20
Cronbach's alpha	0.88	0.71	0.58	0.73	0.61

Figure 2. Variability in (A) Difficulty, *P*, and (B) Discrimination, *D*, for each statement in the GeDI 1.0 across the five classes reported in Table 5. Although the mixed models we used to analyze the data statistically were based on means, it is easiest to summarize the differences in classes with box plots. In these box plots, the bars are medians; the width of the box is the interquartile range, demarcated by the 75th percentile (top of the box) and the 25th percentile (bottom of the box); whiskers represent the lowest and highest statement value across the five courses, unless the maximum or minimum is more than 1.5 times the interquartile range; in such cases, the minimum and/or maximum is represented as a dot. In both panels, statements are sorted into key concepts and misconceptions. In (A) statements are ordered from least difficult to most difficult, that is, decreasing *P*. In (B) statements are ordered from least to most discriminating.

tended to be more difficult for students than statements addressing key concepts ($P = 0.71 \pm 0.14$ SD). However, there was a significant interaction between type of statement (i.e., misconception vs. key concept) and course ($F(4, 80) = 6.33$, $p = 0.0002$); in other words, the difference in difficulty for statements addressing misconceptions versus statements addressing key concepts was larger in some courses than in others. Given this interaction, it is not meaningful to draw conclusions about the difference in difficulty between misconception and key concept statements across all courses.

The statements included in the GeDI 1.0 effectively discriminated between low- and high-performing students (Figure 2B). Mean statement discrimination ranged from $D = 0.09$ (statement 1) to $D = 0.46$ (statement 5). These results mean that the high-performing students answered the most discriminating question correctly 46% more of the

time than low-performing students; the high-performing students answered the least discriminating questions only 9% more of the time than low-performing students. We used a mixed model to examine differences in discrimination between statements addressing misconceptions and those addressing key concepts. In this analysis, the interaction between type of statement and course was not statistically significant ($F(4, 80) = 1.50$, $p = 0.209$), which means the pattern of the difference in discrimination score between misconception statements and key concept statements did not vary across courses. We could therefore draw conclusions about differences across courses. Statements addressing misconceptions had higher discrimination scores ($D = 0.39 \pm 0.13$ SD than statements addressing key concepts ($D = 0.25 \pm 0.16$ SD), and this difference was statistically significant ($F(1, 20) = 26.56$, $p < 0.0001$); in other words, statements addressing misconceptions better distinguished between high- and

low-performing students. Statement 1 failed to meet our original cutoff for statement discrimination ($D > 0.20$), probably because it is the least difficult statement in the instrument. However, we retained this statement because it satisfied our validation criteria for earlier drafts of the GeDI.

DISCUSSION

The GeDI 1.0 produces results that are a reliable and valid measurement of what upper-division biology students understand about genetic drift. Because the format is true–false, the GeDI 1.0 can test how well students grasp key concepts central to understanding genetic drift, while simultaneously testing for the presence of misconceptions that indicate an incomplete understanding of genetic drift. This insight will, over time, allow us to assess the effectiveness of different teaching strategies and curricula, as recommended in the AAAS *Vision and Change* report (AAAS, 2011). Furthermore, it will allow us to improve instruction about genetic drift, a key component of evolutionary thinking.

Insights about How Biology Undergraduates Learn Evolution

In another article, we presented a three-stage, testable sequence of how undergraduate students learn genetic drift (Andrews *et al.*, 2012). In stage 1, students with undeveloped conceptions of evolution try to extract a meaning of genetic drift solely from their colloquial understanding of the words *genetic* and *drift*. Students in stage 2 have slightly more knowledge of evolution, but they conflate genetic drift with other mechanisms of evolution; it is this stage that the GeDI 1.0 targets. In stage 3, students who are beginning to understand genetic drift apply inaccurate limitations on when genetic drift occurs.

The results of our field tests and interviews of all three GeDI drafts indicate that introductory students seem typical of stage 1, but advanced students have passed beyond this point. For example, beginning students hold some misconceptions, such as "Genetic drift only occurs when natural selection cannot or is not occurring." A statement in GeDI-Draft 3 that stated that *the fact that natural selection could not occur contributed to genetic drift* was so easy for both low- and high-performing upper-division students ($P = 0.77$) that it was removed. Therefore, the GeDI does not test for understanding of the novice conceptions often held in stage 1.

Upper-division students fall more typically into stage 2, and the GeDI is optimized for measuring understanding in this population. For example, Andrews *et al.* (2012) observed that introductory students hold the misconception that genetic drift is simply a change in allelic frequencies. Through our process of validation, we found that upper-division students easily recognized that idea as inaccurate, and thus we do not include statements addressing this misconception in the GeDI 1.0. On the other hand, upper-division students frequently confuse genetic drift with other evolutionary processes, especially natural selection (misconceptions 2–4 in Table 3). Therefore, the GeDI 1.0 includes a number of statements for determining whether students conflate genetic drift with natural selection, gene flow, mutation, and speciation.

The GeDI does not test for the misconceptions associated with stage 3 because the ways in which student understanding grows to overcome these misconceptions are more nuanced than this instrument can effectively evaluate. For example, one stage 3 misconception reported in Andrews *et al.* (2012) is that "Genetic drift results only from an isolated event, often a catastrophe." During our validation procedure, students found statements designed to test this misconception to be very easy, even though we found that this misconception is pervasive in open-ended responses (Andrews *et al.*, 2012). Misconceptions such as this are difficult to assess with forced-response questions because they include absolute words such as *only* or *always* that test-wise students associate with false statements (Libarkin, 2008). Thus, even though students use absolute words when describing their own understanding of genetic drift in free-response (Andrews *et al.*, 2012) and interview (this study) questions, they can use test wiseness to recognize that statements assessing those misconceptions are false. For example, both low- and high-performing students were able to easily recognize the correct answers for statements assessing the misconceptions that "Genetic drift only occurs in small populations, because random sampling error does not occur in large populations," "Genetic drift only occurs when natural selection cannot or is not occurring," and "Genetic drift results only from an isolated event, often a catastrophe."

Much of the early literature on evolution education concentrates on the fact that high school and introductory biology students understand disappointingly little about evolution (reviewed in Smith, 2010). Our research indicates that upper-division students are making progress; they understand some concepts integral to understanding genetic drift, while still maintaining some major misconceptions. This finding is highlighted by the fact that we excluded some of the key concepts presented in Table 4. Both low- and high-performing upper-division students achieved high scores on questions related to the key concept that "In populations with small effective sizes, genetic drift can overwhelm the effects of natural selection, mutation, and migration; therefore, an allele that is increasing in frequency due to selection might decrease in frequency [in] some generations due to genetic drift" (key concept 4 in Table 4). In fact, students did so well on questions pertaining to the subconcept "Other evolutionary mechanisms, such as natural selection, mutation, and migration act simultaneously with genetic drift" (4a in Table 4), that statements addressing this subconcept were excluded from the GeDI 1.0. The fact that students did well on statement 4 (Supplemental Material, The Genetic Drift Inventory 1.0) about the evolution of maladaptive traits is noteworthy ("Some harmful traits may have become more common in the island population than the mainland population"; mean difficulty across five courses = 0.71; key concept 4c in Table 4) given that previous research has shown that students struggle to explain the evolution of deleterious traits (e.g., Beggrow and Nehm, 2012). Another statement testing this concept was too easy to include in the GeDI 1.0 and was eliminated because it had low discrimination ("Genetic drift could have resulted in a harmful trait becoming more common"). The fact that upper-division students are already beginning to understand nonadaptive evolution is a noteworthy success indicating that instructors' teaching strategies are partially successful.

Despite some clear gains that students achieve throughout their undergraduate experience, upper-division students still struggle to recognize genetic drift as a distinct evolutionary mechanism separate from natural selection (Freeman and Herron, 2004; Barton *et al.,* 2007; Futuyma, 2009; Table 3). A possible implication of the apparently close conceptual tie between these topics is that teaching genetic drift effectively could be a way to help students address their misconceptions about both natural selection and genetic drift. Another hypothesis stems from the finding that items addressing both drift and selection are among the most difficult: the well-documented challenges of understanding natural selection may hinder learning about other evolutionary topics such as genetic drift. If we could find a way to help students effectively overcome their misconceptions about natural selection, then it would make teaching other topics in evolutionary biology less challenging. Teaching genetic drift effectively may also be one step toward helping students understand probabilistic reasoning, which undergirds so many key scientific concepts. More research is necessary to address both of these intriguing hypotheses.

Guessing Rate

One of the challenges inherent to administering a true–false instrument is that scoring can be biased by a high guessing rate, a problem inherent to any forced-response test. Students who randomly select their answer choices on the GeDI 1.0 will, on average, answer half of the questions correctly. We encourage two strategies to address this weakness. One strategy is for instructors to set a higher-than-usual bar to define successful performance on the GeDI because guesses can inflate scores. Suppose, for example, that a student answers eight questions correctly and guesses on the remaining 14. On average, a student like this will receive a score of 15/22, or 68%. We suggest keeping this in mind when deciding what score will be interpreted as indicating proficiency. Another strategy is to assign a score based on how many points the students earn more than the baseline guess rate of 50%. In that case, the same student would have a score of 4/11, or 36%.

Because of the difficulty in interpreting raw scores of an instrument with a high guess rate, we encourage instructors to focus on the *change* in score on the GeDI 1.0 that students achieve as a result of instruction on genetic drift. This instrument was designed to inform instructors about the diversity and abundance of ideas their students have about genetic drift and to quantify learning gains. It is not intended to be a summative evaluation tool.

Future Uses for the GeDI 1.0

The GeDI can help us, as a community of instructors, build more effective teaching modules. Many different approaches exist for teaching that leads to conceptual change, and the information provided by the GeDI can inform all of them. For instance, taking the approach of Kampourakis and Zogza (2009), we could use the results from pretests of student knowledge with the GeDI to diagnose how misconceptions and key concepts coexist in students' conceptual frameworks and then present experimental data that cannot be explained with the students' current constructs. This approach is designed to encourage students to enter a stage of "conceptual conflict" and to replace their novice construct with one that is more expert-like (Kampourakis and Zogza, 2009).

The obvious next step after developing an instrument for measuring genetic drift understanding with reliable and valid results is to evaluate instruction about genetic drift. Therefore, one of the obvious future directions of this research is to compare the efficacy of different modules for teaching genetic drift. The combination of poor student performance and strong expert performance on the GeDI suggests that, although the GeDI is optimized for advanced undergraduates, it can approximate understanding of genetic drift from introductory to graduate students and university instructors.

REFERENCES

Abraham JK, Perez KE, Price RM (2014). The Dominance Concept Inventory: a tool for assessing undergraduate student misconceptions about dominance in Mendelian and population genetics. CBE Life Sci Educ (*in press*)**.**

American Association for the Advancement of Science (2011). Vision and Change in Undergraduate Biology Education: A Call to Action, Washington, DC.

Anderson DL, Fisher KM, Norman GJ (2002). Development and evaluation of the Conceptual Inventory of Natural Selection. J Res Sci Teach *39*, 952–978.

Andrews TM, Price RM, Mead LS, McElhinny TL, Thanukos A, Perez KE, Herreid CF, Terry DR, Lemons PP (2012). Biology undergraduates' misconceptions about genetic drift. CBE Life Sci Educ *11*, 248–259.

Barton NH, Briggs DEG, E JA, Goldstein DB, Patel NH (2007). Evolution, Cold Spring Harbor, NY: Cold Spring Harbor Press.

Baum DA, Smith SD, Donovan SSS (2005). The tree-thinking challenge. Science *310*, 979–980.

Beggrow EP, Nehm RH (2012). Students' mental modes of evolutionary causation: natural selection and genetic drift. Evol Educ Outreach *5*, 429–444.

Campbell NA, Reece JA, Urry LA, Cain ML, Wasserman SA, Minorsky PV, Jackson RB (2008). Biology, San Francisco, CA: Pearson.

Cotner S, Brook DC, Moore R (2010). Is the age of the earth one of our "sorest troubles"? Students' perceptions about deep time affect their acceptance of evolutionary theory. Evolution *64*, 858–864.

Crocker L, Algina J (1986). Introduction to Classical and Modern Test Theory, Orlando, FL: Holt, Rinehart and Winston.

Crowther GJ, Price RM (2014). Re: misconceptions are "so yesterday." CBE Life Sci Educ *13*, 3–5.

D'Avanzo C (2008). Biology Concept Inventories: overview, status, and next steps. BioScience *58*, 1079–1085.

Doran R (1980). Basic Measurement and Evaluation of Science Instruction, Washington, DC: National Science Teachers Association.

Fischbein E, Schnarch D (1997). The evolution with age of probabilistic, intuitively based misconceptions. J Res Math Educ 28, 96–105.

Fisher KM, Williams KS, Lineback JE (2011). Osmosis and diffusion conceptual assessment. CBE Life Sci Educ 10, 418–429.

Freeman S (2005). Biological Science, Upper Saddle River, NJ: Pearson.

Freeman S, Herron JC (2004). Evolutionary Analysis, Upper Saddle River, NJ: Pearson.

Frisbie DA (1992). The multiple true-false item format: a status review. Educ Meas 11, 21–26.

Futuyma DJ (2009). Evolution, Sunderland, MA: Sinauer.

Garvin-Doxas K, Klymkowsky MW (2008). Understanding randomness and its impact on student learning: lessons learned from building the Biology Concept Inventory (BCI). CBE Life Sci Educ 7, 227–233.

Gelman A, Hill J (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models, New York: Cambridge University Press.

Gliner JA, Morgan GA, Leech NS (2009). Research Methods in Applied Settings: An Integrated Approach to Design and Analysis, 2nd ed., New York: Routledge.

Gormally C, Brickman P, Lutz M (2012). Developing a Test of Scientific Literacy Skills (TOSLS): measuring undergraduates' evaluation of scientific information and arguments. CBE Life Sci Educ 11, 364–377.

Gould SJ, Lewontin RC (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc R Soc London B 205, 581–598.

Gregory TR (2009). Understanding natural selection: essential concepts and common misconceptions. Evol Educ Outreach 2, 156–175.

Haladyna TM (2004). Developing and Validating Multiple-Choice Exam Items, Mahwah, NJ: Erlbaum.

Hiatt A, Davis GK, Trujillo C, Terry M, French DP, Price RM, Perez KE (2013). Getting to evo-devo: concepts and challenges for students learning evolutionary developmental biology. CBE Life Sci Educ 12, 494–508.

Jungck JR (1997). Ten equations that changed biology: mathematics in problem-solving biology curricula. BioScene 23, 11–36.

Kampourakis K, Zogza V (2009). Preliminary evolutionary explanations: a basic framework for conceptual change and explanatory coherence in evolution. Sci Educ 18, 1313–1340.

Kaplan JJ, Fisher D, Rogness N (2010). Lexical ambiguity in statistics: how students use and define the words: association, average, confidence, random and spread. J Stat Educ, 18. www.amstat.org/publications/jse/v18n2/kaplan.pdf (accessed 28 January 2014).

Klymkowsky M, Garvin-Doxas K (2008). Recognizing student misconceptions through Ed's Tools and the Biology Concept Inventory. PLoS Biol 6, 14–17.

Leonard MJ, Kalinowski SK, Andrews TC (2014). Misconceptions yesterday, today, and tomorrow. CBE Life Sci Educ (in press).

Libarkin J (2008). Concept inventories in higher education science. Paper presented at the National Research Council Promising Practices in Undergraduate STEM Education Workshop 2, held 13–14 October 2008, in Washington, DC.

Maloney DP (1981). Comparative reasoning abilities of college students. Am J Phys 49, 784–786.

Masel J (2012). Rethinking Hardy–Weinberg and genetic drift in undergraduate biology. BioEssays 34, 701–710.

Maskiewicz AC, Lineback JE (2013). Misconceptions are "so yesterday!" CBE Life Sci Educ 12, 352–356.

McShea DW (1994). Mechanisms of large-scale evolutionary trends. Evolution 48, 1747–1763.

Mead LS, Scott EC (2010). Problem concepts in evolution. Part II: Cause and chance. Evol Educ Outreach 3, 261–264.

Mills LS (2007). Conservation of Wildlife Populations, Malden, MA: Blackwell.

Mlodinow L (2008). The Drunkard's Walk: How Randomness Rules Our Lives, New York: Pantheon.

Nadelson LS, Southerland SA (2010). Development and preliminary evaluation of the Measure of Understanding of Macroevolution: introducing the MUM. J Exp Educ 78, 151–190.

Nehm RH, Schonfeld IS (2008). Measuring knowledge of natural selection: a comparison of the CINS, an open-response instrument, and an oral interview. J Res Sci Teach 45, 1131–1160.

Neuman WL (2009). Understanding Research, Boston, MA: Pearson.

Novick LR, Catley KM (2012). Assessing students' understanding of macroevolution: concerns regarding the validity of the MUM. Int J Sci Educ 34, 2679–2703.

Parker JM, Anderson CW, Heidemann M, Merrill J, Merritt B, Richmond G, Urban-Lurain M (2012). Exploring undergraduates' understanding of photosynthesis using diagnostic question clusters. CBE Life Sci Educ 11, 47–57.

Perez KE, Hiatt A, Davis GK, Trujillo C, French DP, Terry M, Price RM (2013). The EvoDevoCI: a concept inventory for gauging students' understanding of evolutionary developmental biology. CBE Life Sci Educ 12, 665–675.

Raup DM, Gould SJ, Schopf TJM, Simberloff DS (1973). Stochastic models of phylogeny and the evolution of diversity. J Geol 81, 525–542.

Smith JI, Tanner K (2010). The problem of revealing how students think: concept inventories and beyond. CBE Life Sci Educ 9, 1–5.

Smith MK, Wood WB, Knight JK (2008). The Genetics Concept Assessment: a new concept inventory for gauging student understanding of genetics. CBE Life Sci Educ 7, 422–430.

Smith MU (2010). Current status of research in teaching and learning evolution. II. Pedagogical issues. Sci Educ 19, 539–571.