Chapman University
Chapman University Digital Commons

Biology, Chemistry, and Environmental Sciences Faculty Articles and Research

Science and Technology Faculty Articles and Research

1-16-2019

Concept Inventories as a Resource for Teaching Evolution

Robert E. Furrow
University of California, Davis

Jeremy L. Hsu
Chapman University, hsu@chapman.edu

Follow this and additional works at: https://digitalcommons.chapman.edu/sees_articles

Part of the Biology Commons, Curriculum and Instruction Commons, Educational Methods Commons, Higher Education Commons, Other Ecology and Evolutionary Biology Commons, Other Life Sciences Commons, and the Science and Mathematics Education Commons

Concept Inventories as a Resource for Teaching Evolution

Comments

This article was originally published in *Evolution: Education and Outreach*, volume 12, issue 2, in 2019.

CURRICULUM AND EDUCATION

CrossMark

Concept inventories as a resource for teaching evolution

Robert E. Furrow[1*†] and Jeremy L. Hsu[2†]

Abstract

Understanding evolution is critical to learning biology, but few college instructors take advantage of the body of peer-reviewed literature that can inform evolution teaching and assessment. Here we summarize the peer-reviewed papers on tools to assess student learning of evolutionary concepts. These published concept inventories provide a resource for instructors to design courses, gauge student preparation, identify key misconceptions in their student population, and measure the impact of a lesson, course, or broader curriculum on student learning. Because these inventories vary in their format, target audience, and degree of validation, we outline and explain these features. In addition to summarizing the published concept inventories on topics within evolution, we lay out a flexible framework to help instructors decide when and how to use them.

Keywords: Assessment, Concept inventories, Evolution, Learning goals

Introduction

Facility with evolutionary concepts is foundational to a rich understanding of biology, and several large, collaborative efforts to improve undergraduate education have outlined this importance (American Association for the Advancement of Science 2011; Association of American Medical Colleges and the Howard Hughes Medical Institute 2009; National Research Council 2003, 2009, 2012). *Thinking Evolutionarily,* a report summarizing a convocation organized by the National Research Council and the National Academy of Sciences, lays out the value of and practical approaches to infusing the teaching of evolution throughout biology courses across K-12 and undergraduate curricula (National Research Council 2012). Focusing on undergraduate curricula, the American Association for the Advancement of Science report *Vision and Change* identifies core concepts within evolutionary biology for developing biological literacy (American Association for the Advancement of Science 2011). That succinct description of concepts has since been interpreted and elaborated for specific fields of biology

(American Society of Plant Biologists and the Botanical Society of America 2016; Merkel et al. 2012; Tansey et al. 2013), and translated into a framework to help instructors align their departmental educational goals with *Vision and Change* (Brownell et al. 2014). However, even with clear educational goals in mind, carefully measuring student learning and adjusting teaching practices to achieve these goals is a daunting task (Handelsman et al. 2004).

One way to measure student learning, usually within the context of a single course or module, is by using a concept inventory. Concept inventories are test-based assessments of a concept or set of concepts, usually using multiple-choice questions (D'Avanzo 2008; Knight 2010). The incorrect choices for a question are called distractors, and are ideally based around common student misconceptions (Haladyna et al. 2002; Sadler 1998). For example, to create the Genetic Drift Inventory (GeDI), a concept inventory of genetic drift, the authors used student interviews and built upon previous work to identify six common student misconceptions about genetic drift, then designed many of the inventory's questions to assess these (see Table 3 in Price et al. 2014, as well as Andrews et al. 2012). One misconception they identified was that "Natural selection is always the most powerful

*Correspondence: refurrow@ucdavis.edu
†Robert E. Furrow and Jeremy L. Hsu contributed equally to this work
[1] University of California, Davis, 1544 Newton Court Rm 203, Davis, CA 95618, USA
Full list of author information is available at the end of the article

mechanism of evolution, and it is the primary agent of evolutionary change", and four of the 22 questions on the inventory test some aspect of this misconception.

Despite the growing number of concept inventories assessing topics in evolution, there are many impediments to their widespread use among college instructors. First, the current concept inventories cover only a few of the major topics that may be taught in an undergraduate evolution course. In an analysis of peer-reviewed evolution education research, Ziadie and Andrews (2018) found that the majority of published papers pertaining to assessment of evolutionary concepts relate only to natural selection or phylogenetics (particularly tree-thinking). Many common topics in undergraduate evolution courses had limited or no coverage. In addition, Ziadie and Andrews note there are few literature reviews of such assessments, and that college instructors who wish to use these assessments in their teaching would benefit from a review of evolution-related assessments that summarize both the topics and misconceptions covered and the differences in approach to their development.

Alongside the challenge of uneven coverage, college instructors also face barriers to translating this work into practical use (Anderson 2007). Instructors often have limited time and training to apply new teaching methods (American Association for the Advancement of Science 2011; Henderson et al. 2011; Henderson and Dancy 2007), and may face tensions with professional norms about scientific identity (Brownell and Tanner 2012). In some cases, discipline-based educational research may not be presented in a way that is clearly connected to classroom application (Kempa 2002). In other cases, instructors may not have confidence in the validity of the interpretation of educational research (Herron and Nurrenbern 1999).

Concept inventories avoid some of these concerns, as they are generally designed to be easily used within the current framework of a course. However, there are limitations to their effective use. The target audience is not always clear, and instructors may be unsure of exactly how to interpret results. Furthermore, concept inventories are often limited in their scope and interpretation, and can be influenced by the specific design of the test questions and logistics of test implementation. Understanding how the inventory creators gathered evidence about its validity (Box 1) is critical (Adams and Wieman 2011).

This paper aims to be a resource for college instructors in evolution, helping to minimize the challenges and maximize the benefits of using concept inventories in teaching. We present the logic of why and how an instructor might choose to use a concept inventory in their teaching, and summarize current evolution concept inventories. We also briefly outline the general process of concept inventory validation. To ground the discussion in practice, we explain several ways an instructor might use the inventory to support their teaching, including applications that do not require formal student test-taking.

Why and how to use concept inventories

Many papers have examined the goals and benefits of using concept inventories to inform undergraduate teaching (Adams and Wieman 2011; D'Avanzo 2008; Garvin-Doxas et al. 2007; Knight 2010; Libarkin 2008; Marbach-Ad et al. 2010; Smith and Tanner 2010; Steif and Hansen 2007). Here, we synthesize and build upon these goals, highlighting several key benefits of using concept inventories to inform teaching of evolutionary concepts.

Concept inventories with validity evidence based on test content can inform learning objectives within a course or across a broader curriculum

The majority (14 out of 16) of concept inventories relating to evolution that we identified had empirical evidence for the validity of the test content (see Box 1 and Table 1), meaning that there were several steps in the development of the concept inventory where content experts (i.e. evolution experts) or other sources of expert knowledge (e.g. peer-reviewed literature or textbooks) were consulted. A subset of these concept inventories also attempt to cover all major themes relevant for the given topic assessed in the concept inventory by asking the content experts to delineate main learning goals and concepts related to the topic. As such, these concept inventories can be used to identify potential core ideas related to a topic, which can in turn influence an instructor's preparation for a course. If the instructor follows principles of backward design (Wiggins and McTighe 2005), then these concept inventories provide a ready-made list of learning goals and concepts relevant to the evolutionary topic.

Box 1—validation

A concept inventory is a test to assess conceptual understanding. But what exactly is this test measuring? Validation is the process of gathering evidence about the "the degree to which evidence and theory support the interpretations of test scores for proposed uses" (American Educational Research Association et al. 2014). In other words, validity evidence is critical to ensure that a test is actually assessing student understanding of the concepts that it purports to measure. Many forms of validity evidence can be gathered during a pilot period before the concept inventory is rolled out to a large student population, while others

involve the statistical analysis of student responses to the instrument. Table 1 presents a framework defining different types of validity evidence (American Educational Research Association et al. 2014). These types of evidence have been elaborated and explained in more detail for discipline-based educational assessment (Reeves and Marbach-Ad 2016). Few assessments have all of these types of validity evidence, and a concept inventory may still be useful for teaching and learning even if its validation process was minimal. Evidence based on response processes, particularly from think-aloud student interviews, can be especially useful (Adams and Wieman 2011). This evidence reveals how the students think about answering each question, and provides an instructor with some confidence that the student responses should reflect their underlying understanding of the concepts being tested. However, as with all validation, different student populations may respond differently, so additional evidence should be gathered if high-stakes decisions rest upon the results of the assessment. Validation is an ongoing process; researchers have continued to validate and propose modifications to several concept inventories, including the Concept Inventory of Natural Selection (Furtak et al. 2011; Nehm and Schonfeld 2008) and the Measure of Understanding of Macroevolution (Novick and Catley 2012). Note that test validity is a distinct concept from test reliability, which refers to the consistency of test results over multiple instances of the test application. Reliability is less often analyzed in the creation of concept inventories, though it can provide evidence that test format and other extraneous variables do not have an undue effect on assessment results.

For example, one of the authors (JLH) has used the genetic drift inventory, GeDI (Price et al. 2014), while designing a mid-/upper-level evolution course. During the development of the GeDI, Price et al. surveyed content experts and generated a list of main concepts relevant to genetic drift that the experts identified as appropriate (and relevant) to advanced undergraduates studying evolution (see Table 4 in Price et al. 2014). While not all these concepts were ultimately included in the GeDI, JLH consulted this table during development of his course to cross-reference his own list of topics related to genetic drift and to ultimately generate a list of key learning objectives he wanted his students to be able to master.

Concept inventories can also be used to inform learning objectives about a given topic *across* courses in a curriculum. For instance, Marbach-Ad et al. (2010) created a curricular alignment map based on the list of topics in a concept inventory by surveying instructors in different courses. This alignment allowed the instructors to discuss the progression of learning about the topic across classes, and sparked changes in some of the surveyed courses. Concept inventories can also aid in the planning process for a new series of courses. One author (REF) reviewed the validated biology science quantitative reasoning exam (BioSQuaRE; Stanhope et al. 2017), to create a set of learning goals across multiple introductory quantitative biology courses. Although BioSQuaRE is not exactly a concept inventory, the process of test content validation in the creation of this instrument made a convincing case for their set of quantitative biology learning goals.

Table 1 Types of test validity evidence

Evidence based on…	Overview	Example
… test content	Checking the match between the assessment's content and what it claims to measure	Getting the assessment reviewed by multiple content experts to ensure even and thorough coverage of a concept
… response processes	Making sure that the test takers' responses align with the underlying concepts being measured	Think-aloud interviews with students where they explain the reasoning behind each answer they chose on the test
… internal structure	Analyzing how the assessment questions relate to each other and map onto the underlying concepts being measured	Using a statistical modeling approach like factor analysis to analyze how responses to different questions correlate
… relations with other variables	Relating test scores to external variables. The external variable could be the results of a similar test (convergent evidence), a test expected to be different (discriminant evidence), or non-test outcomes like students' future course enrollment decisions	Comparing the assessment to a previously established assessment of similar concepts, expecting to find a positive correlation in student scores
… consequences of testing	Evaluating the soundness of using the assessment in practice, including potential impacts of the testing itself and appropriateness of interventions made based on student results	For an assessment being used to group students into leveled sections, gathering evidence on the learning outcomes of this intervention

Concept inventories can identify key misconceptions students hold about an evolutionary topic

Most concept inventories are designed specifically to identify student misconceptions; the multiple-choice concept inventories often rely on distractor answer choices that align with common misconceptions. In addition, several of the concept inventory publications we examined directly identify (either with empirical data or by reviewing peer-reviewed literature) common student misconceptions related to that evolutionary topic. Instructors can benefit from knowledge of these common student misconceptions, given the empirical evidence that a powerful and engaging way to promote deep learning is by eliciting and addressing misconceptions in a systematic manner (e.g. Allen and Tanner 2005; Andrews et al. 2011; Gregory 2009; Nelson 2008). By examining the list of misconceptions identified during development of the GeDI (Price et al. 2014), JLH was able to design activities to directly confront these misconceptions, and incorporated a homework assignment where students were asked to reflect upon their own genetic drift misconceptions and explain why they were incorrect. Students were also challenged to explain why several common misconceptions about drift were incorrect. Once these misconceptions are identified, instructors may draw upon articles that provide further insight into these misconceptions (e.g. Andrews et al. 2012; Gregory 2008) and may look into peer-reviewed curricula for activities designed to counter misconceptions about evolution (e.g. Andrews et al. 2011; Govindan 2018; Kalinowski et al. 2013; Meisel 2010).

Concept inventories allow for measuring student knowledge in a topic before a course or module

In addition to the identification of common misconceptions about a given topic, instructors who have students take a concept inventory at the beginning of a course (or before the topic is covered in the course) can better identify the level of expertise the students have on the given topic, thus allowing the instructor to tailor the instruction to the students' background knowledge on the topic. The concept inventory can also identify specific misconceptions that students in the class harbor, again allowing the instructor to design specific learning activities to counter those misconceptions.

Concept inventories can be used to compare students' background knowledge on a topic across different course sections

Concept inventories can be used to compare student levels across different course sections. For instance, one of the authors (JLH) teaches a course that has several lecture sections, with different sections each having a different instructor. The instructors of the course each give a pre-course assessment with questions from several concept inventories. If one section has many more students holding a particular misconception than another section, the instructor of the former can spend more time addressing the misconception while the other instructors may not need to spend as much time. The scores on this standardized pre-course assessment also contextualize scores on other standardized assessments (e.g. mid-semester and final exams) that are shared in common across the course sections. The instructors have found, unsurprisingly, that in years where students have performed significantly lower in the pre-course assessment in one section, those same students tend to perform worse on the standardized mid-semester and final exams. Without these data, the instructors might have mistakenly attributed the differences in scores to differences in grading or teaching. While there might still be differences in these latter categories (despite the instructors' best efforts to standardize teaching and grading), the scores from the pre-course assessment provide greater context on student background levels.

Concept inventories can be used to assess student learning during a course, module, or activity

Many concept inventories can be used for a pre/post assessment, where the concept inventory is given on the first day of class (or is assigned outside of class for homework or a small amount of participation or bonus points) and then again on the last day of class or embedded in the final exam. Use of concept inventories for such pre/post assessment can be used to assess student learning of the particular evolutionary topic, and can also inform the instructor about which misconceptions, if any, the students still hold after the class, module or activity. In addition, there are some concept inventories (e.g. EcoEvo-MAPS; Summers et al. 2018) designed for longitudinal assessment of a given student cohort. Such an assessment can be given at multiple points throughout an undergraduate cohort's college career, and provide valuable information on student learning throughout their time in the undergraduate program. Assessment data is crucial for the process of scientific teaching (Handelsman et al. 2004), and these data can also be used to identify demographic variables (e.g. ethnicity, gender, etc.) that correlate with learning or preparation if the instructors also collect these demographic information (Marbach-Ad et al. 2010).

Concept inventories can inform changes in instruction from year to year

The use of concept inventories to assess student learning in a course, track a cohort's progress throughout their

undergraduate careers, and identify remaining misconceptions can provide valuable feedback to instructors as they reflect on a course. These data can thus help identify both strengths and weaknesses in a given course, module, or activity, and the instructor can use these data to make changes as appropriate to the course. For instance, one of the authors (JLH) has made changes to his mid/upper-level evolution course, spending additional time on activities related to genetic drift, after questions from the GeDI in the first iteration of the course identified that students still harbored major misconceptions about drift and were not mastering the main learning objectives in a way that the instructor had hoped for. These questions from the GeDI will be used this semester to assess the impact of the changes made in the evolution course this year. Similarly, the use of concept inventories in a longitudinal fashion can also inform broader program-wide curricular discussions.

Concept inventories can inspire instructors to create their own activities and assessments

Finally, concept inventories can be a source of inspiration for instructors in terms of designing new activities and assessments. Concept inventories that have evidence of test content have been reviewed by content experts, and looking at the concepts, misconceptions, and question formats can generate new ideas for instruction and assessment.

How to administer the concept inventory as a test

Several of the approaches above do not require you to actually administer the concept inventory as a test. However, you may wish for students to take the concept inventory to measure student learning or background knowledge. At this point several common questions arise. Is it okay to use a subset of the inventory questions? Should students take this in class, or can it be administered online? Will offering extra credit bias the participation? Choosing only a subset of questions may be practical, as it allows a shorter assessment that can be tailored to your course learning goals. However, the process of validation for an inventory is based around the complete question set. You can still learn useful information about student learning, but data cannot be easily compared with other instances of test implementation. When possible, refer to the statistical analyses of a test's internal structure, which may reveal clusters of conceptually related questions that either form a natural subset or provide a basis to select questions that still span some breadth of content. Regarding test location and incentives, Madsen et al. (2017) review many studies of concept inventory implementation, noting that a small amount of extra credit may increase test completion

without unduly influencing scores. Madsen et al. also argue strongly for the assessment to be taken in some supervised setting, though the format could be paper or online. This eliminates concerns about students using outside resources or saving and sharing questions outside of class, and can increase completion rates.

General steps to use concept inventories

While there is no set "formula" for how to use concept inventories, we delineate five general steps for how to use a concept inventory.

1. *Determine your goals for using concept inventories.* In other words, how do you want to use concept inventories to inform your teaching? Which of the above goals do you wish to accomplish, and for which topic within evolution? Which classes are you thinking of using the concept inventory for? Is the class a non-majors class or one for biology majors? Is it an introductory or advanced class? Are you hoping to assess learning throughout the whole course, or for a specific module or activity? Thinking carefully about your goals and objectives is essential before you start looking at specific concept inventories.
2. *Identify and obtain relevant concept inventories.* Once you have thought carefully about your goals, you can now identify any relevant concept inventories to your chosen topic. Table 2 provides a current list of all concept inventories with content relevant to evolution as of the time of publication, as well as how to obtain them. Concept inventories are often, but not always, found in the relevant paper or its supplement.
3. *Review the details of the concept inventory and its development.* We have summarized some features of each concept inventory (e.g. target population, time it takes to complete the concept inventory, types of validation evidence; Table 2). This information can help you check the appropriateness of the concept inventory to your class and your goals. If you plan to administer the concept inventory as a test and use the results to draw conclusions about student learning, make sure that the validation population is similar to your focal student population, and that the evidence the inventory creators present is convincing. When in doubt, consider ways that you might gather additional evidence to strengthen your confidence in the inventory's use. For example, you could conduct student think-aloud interviews or use additional free-response questions (Table 1); Furtak et al. (2011) model this process as they performed additional validation and adjusted the Concept Inventory of Natural Selection (Anderson et al. 2002) for use with high

Table 2 Evolution concept inventories

Assessment name, shorthand (citation)	Content/population	Answer type (time)	Validation population/evidence	How to access in creators' publication/Other notes
Ecology and evolution–measuring achievement and progression in science, EcoEvo-MAPS (Summers et al. 2018)	Ecology and evolution/AY	63 MC (15–35 m)	Multiple/TC, RP, IS	Supplement/Designed for longitudinal use
Biological concepts instrument, BCI (Klymkowsky et al. 2010)	Molecular biology/Y1, Y2	29 MC	Multiple/TC, RP	Appendix
Unnamed instrument (Blacquiere and Hoese 2016)	Phylogenetics/Y1	18 MC (<20 m)	CSU Fullerton/TC, RP, IS, C	Supplement
Phylogenetic assessment test, PhAT (Smith et al. 2013b)	Phylogenetics/Y1	5 open-ended	Michigan State University/TC, RP	Embedded/Rubric published in paper
Tree-thinking concept inventory, TTCI (Naegle 2009)	Tree thinking/AY	29 MC (20–30 m)	Multiple/TC, RP, IS	Contact author/Suggestions for choosing subset of Qs detailed in thesis
Basic tree-thinking assessment (Baum et al. 2005)	Tree thinking/AY	10 MC each	Not stated	Supplement/Two separate quizzes (I for LD, II for UD)
Conceptual inventory of natural selection, CINS (Anderson et al. 2002)	Natural selection/NM	20 MC	Multiple/TC, RP, IS	Supplement
Assessing conceptual reasoning about natural selection, ACORNS (Nehm et al. 2012)	Natural selection/LD	Variable open-ended	The Ohio State University/TC, RP, IS, C	Supplement/Grading with rubric (Nehm et al. 2010; Opfer et al. 2012) or with EvoGrader (Moharreri et al. 2014)
Conceptual assessment of natural selection, CANS (Kalinowski et al. 2016)	Natural selection/LD	24 MC	Montana State University/TC, RP, IS	Supplement
Open response instrument, ORI (Nehm and Schonfeld 2008)	Natural selection/LD	5 open-ended (25 m)	University in urban area of NE US/TC, RP, IS, C, D	Embedded/Built from earlier assessments (Bishop and Anderson 1990; Nehm and Reilly 2007). Scoring guide and rubric (Nehm et al. 2010)
Natural selection misconceptions diagnostic (DeSaix et al. 2011)	Natural selection/LD	12 MC	Not stated	https://evolution.berkeley.edu/evolibrary/teach/evo_misconceps_diagnostic.pdf
Evolutionary development concept inventory, EvoDevoCI (Perez et al. 2013)	Evolutionary development/AY	11 MC	Multiple/TC, RP, IS	Supplement/Details about CI development also published (Hiatt et al. 2013)

Table 2 (continued)

Assessment name, shorthand (citation)	Content/population	Answer type (time)	Validation population/evidence	How to access in creators' publication/Other notes
Host–pathogen interactions concept inventory, HPI-CI (Marbach-Ad et al. 2009)	Host–pathogen interactions/AY	18 MC	University of Maryland/TC, RP, IS	Unclear
Measure of understanding of macroevolution, MUM (Nadelson and Southerland 2009)	Macroevolution/LD	27 MC (plus 1 open-ended)	Not stated/TC, IS	Appendix
Genetic drift inventory, GeDI (Price et al. 2014)	Genetic drift/UD	22 MC	Multiple/TC, RP, IS	Supplement
Dominance concept inventory, DCI (Abraham et al. 2014)	Mendelian genetics/AY	16 MC	Multiple/TC, RP, IS	Supplement/Some questions are two-tiered
Genetics concepts assessment, GCA (Smith et al. 2008)	Genetics/AY	25 MC	Multiple/TC, RP, IS, C	Contact authors
Genetics literacy assessment instrument, GLAI (Bowling et al. 2008)	Genetics/NM	31 MC	University of Cincinnati/TC, RP, IS	https://etd.ohiolink.edu/rws_etd/document/get/ucin1195583851/inline (p94)

Y1 first-year undergraduates, *Y2* second-year, *AY* all years, *LD* lower division majors, *UD* upper division majors, *NM* non-majors, *MC* multiple choice, *m* expected number of minutes to complete inventory, Validity evidence: *TC* test content, *RP* response processes, *IS* internal structure, *C* convergent with other variable, *D* discriminant from other variable

school students. In addition, be sure to review the inventory's associated paper for more details about the concept inventory's development. These details can be a valuable resource to reveal student thinking about the concept.

4. *Establish a plan for how and when you will use the concept inventory.* Once you have reviewed this information, you can then establish a plan of how and when you want to use the concept inventory for your class. For example, you might want to use the inventory both before and after a course or set of lessons, or you may only plan to use the assessment at a single time point.

5. *Assess and reflect on your data, if appropriate.* Finally, after implementing your plan, it is vital that you assess and reflect on any data you may have gathered from utilizing concept inventories. These data should allow you to make changes as appropriate to your teaching, and you may then iterate through this process again to continually assess and improve student learning.

Limitations of concept inventories

We hope that concept inventories will prove useful to some readers who had not previously considered their application. However, there are limitations to the use of concept inventories that all instructors should be aware of prior to use. We group these limitations into three main categories: validation-based, cognition-based, and logistical.

For validation-based limitations, concept inventories can be influenced by students' ability to think critically and understand advanced vocabulary and jargon (Knight 2010; Smith and Tanner 2010). While promoting critical thinking and knowledge of evolution vocabulary are important goals, the lack of a foundation in either may confound students taking a concept inventory even if they do have a good conceptual framework of the topic. As such, scores on the concept inventory may not necessarily reflect students' true understanding of the topic. In addition, given that most of these concept inventories rely primarily on multiple choice questions (or agree/disagree questions with even fewer choices), student scores may be artificially inflated by guessing, which can lead instructors to overestimate students' mastery. Several authors of concept inventories (e.g. Price et al. 2014) caution against relying on a single data point of student performance on a concept inventory, and instead advise faculty to focus on comparing student scores across different times (e.g. a pre/post test). Summers et al. (2018) also note that student motivation on a given assessment plays a role in student performance. Instructors are advised to emphasize to students that they should take each assessment seriously, or to use class time or incentives to encourage effortful completion.

In addition, concept inventories may be limited by cognitive biases. Students' mental models of an evolutionary concept may influence the accuracy of the concept inventory as an assessment of skill and knowledge. Novice students who have constructed naïve models of the concept may focus on (and thus be influenced by) surface features of the problem, such as the type of organism, while expert thinkers are able to identify the key biological concepts (Smith et al. 2013a). Studying student open responses to questions about evolutionary change, Nehm and Ha (2011) discovered that students perform worse when asked about evolutionary trait loss versus evolutionary trait gain, despite the two having similar explanations based on natural selection. Many other cognitive biases have been identified, including differences in student performance on questions testing identical evolutionary concepts when using familiar organisms versus unfamiliar taxa or when testing changes between versus within species (Nehm et al. 2012; Novick and Catley 2014; Opfer et al. 2012). Concept inventories that do not draw upon this body of knowledge to shape their design and validation may produce inaccurate results that are influenced by these cognitive factors, and instructors should be aware of these cognitive biases when teaching these subjects and using the concept inventories. For example, one may expect different patterns of student responses from a concept inventory on tree-thinking that uses only familiar organisms in its trees versus one that uses a mix of familiar and unfamiliar organisms.

There are also several logistical challenges to implementing concept inventories. While most of the evolution concept inventories that we identified (13 out of 16) rely on multiple choice questions, some assessments use open-ended questions. These questions require more time to grade, and there may be variation in scoring from one instructor to another, even with a given rubric. Furthermore, some concept inventories are not found in the associated peer-reviewed paper and thus may not be immediately accessible to instructors; we have attempted to alleviate this challenge by providing a column for how to access each concept inventory in Table 2. Despite this, some of the concept inventories require emailing authors, and other concept inventories may have restrictions on how they may be used. Finally, there may be problems with instrument validity if instructors use a partial set of questions from concept inventories, or even if they use questions in a different order (Balch 1989; Federer et al. 2015; Hambleton and Traub 1974), although a study that included analysis of question order did not find an effect for the GeDI (Tornabene et al. 2018). Using a partial set

Table 3 Topic coverage by current evolution concept inventories

Topic	Concept inventories	EcoEvo-MAPS questions	BCI questions
Natural selection	CINS, ACORNS, CANS, Natural Selection Misconceptions Diagnostic, HPI-CI, DCI, ORI	2.3, 2.7, 2.8, 3.3, 3.4, 5.3, 6.6, 7.4	6, 12, 26
Macroevolution	MUM	4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7	4
Speciation		7.5	
Phylogenetics	Unnamed (phylogenetics), PhAT, Tree-thinking Quizzes I and II, TTCI	2.2, 2.3, 6.1, 6.2, 6.3, 6.4, 6.5	
Population genetics	GeDI, DCI	2.5, 5.2, 5.7, 6.7, 7.4, 7.5, 7.6	5, 15, 26, 29, 30
Origin of variation		2.4, 2.6, 5.1, 5.4, 5.6, 7.6	14
Evolution of behavior			
Human evolution			
Molecular evolution			
Sexual selection			
Coevolution	HPI-CI		
Quantitative genetics			
Evolutionary medicine			
Biodiversity		4.1	
EvoDevo	EvoDevoCI		
Human impact			

of questions may still provide valuable information to an instructor. However, it limits the instructor's ability to generalize student performance to a measure of overall student facility with the broader concept, and restricts comparisons with other studies that use the assessment. In many cases this may not be a problem for practical use.

Identifying evolution concept inventories
To identify the currently published concept inventories, we conducted a comprehensive literature search with both Google Scholar and PubMed, using the search terms "evolution* 'concept inventory'", and "biology 'concept inventory'". Although this helped us locate many inventories of evolutionary concepts, we continued to find others through published references to other, non-peer-reviewed work. After building the complete list, both authors conducted another search and double-checked each published inventory's references, and the papers citing each inventory, finding no additional evolution concept inventories as of October 24, 2018.

In total, we identified 14 concept inventories assessing specific topics in evolution, 2 broader concept inventories that had some questions assessing evolutionary topics, and 2 genetics concept inventories with questions that may be useful to instructors teaching evolution. Table 2 summarizes these inventories. We categorized each concept inventory by topic, and created a table with inventory details including: target students, question types and number, validation population, and types of validity

evidence. The authors each independently coded each inventory, and any discrepancies were resolved through discussion.

Opportunities for new assessments
Even with 14 evolution-focused concept inventories, coverage across topics was uneven (Table 3). Seven inventories assessed natural selection, four assessed phylogenetics, and other topics generally had coverage by one or no inventories. We also mapped the questions from the two broader inventories, ecology and evolution–measuring achievement and progression in science (EcoEvo-MAPS; Summers et al. 2018) and the Biological Concepts Instrument (Klymkowsky et al. 2010), onto the topics outlined above. The authors of EcoEvo-MAPS also have their own categorization for each of their questions, available by contacting the corresponding author. Natural selection and phylogenetics were similarly well-covered here, as well as macroevolution and population genetics. However, many topics were sparsely or not at all covered by any inventories: speciation, evolution of behavior, human evolution, molecular evolution, sexual selection, quantitative genetics, evolutionary medicine, biodiversity, and human impact. As new concept inventories are created, the process of validation (particularly student think-aloud interviews and other response-process validation) will hopefully continue to reveal new misconceptions and forms of assessment for these less-covered topics.

Conclusion

This paper argues for the varied and flexible potential uses of concept inventories to support undergraduate learning of evolution. Although concept inventories may not always be the ideal assessment instrument for your learning goals, published descriptions of their creation and validation offer a rich additional resource for assessment and curricular development. Despite the large number of topic-specific inventories, many concepts in evolution remain uncovered and could benefit from new assessments. By summarizing the evolution concept inventories and outlining their details and validation approaches, we hope that instructors can quickly identify instruments for further examination. There are surely many other creative ways to use these inventories; usefulness in service of student learning is the key objective.

Authors' contributions
REF and JLH authors performed research, analyzed data, wrote portions of the paper, and edited the writing. Both authors read and approved the final manuscript.

Author details
[1] University of California, Davis, 1544 Newton Court Rm 203, Davis, CA 95618, USA. [2] Schmid College of Science and Technology, Chapman University, One University Dr., Orange, CA 92866, USA.

Competing interests
The authors declare that they have no competing interests.

Availability of data and materials
All analyzed materials are publicly accessible from the paper's citations, or through contact with corresponding authors of the relevant pieces.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
Abraham JK, Perez KE, Price RM. The dominance concept inventory: a tool for assessing undergraduate student alternative conceptions about dominance in Mendelian and population genetics. CBE Life Sci Educ. 2014;13(2):349–58.

Adams WK, Wieman CE. Development and validation of instruments to measure learning of expert-like thinking. Int J Sci Educ. 2011;33(9):1289–312.

Allen D, Tanner K. Infusing active learning into the large-enrollment biology class: seven strategies, from the simple to complex. Cell Biol Educ. 2005;4(4):262–8.

American Association for the Advancement of Science. Vision and change in undergraduate biology education: a call to action. Derwood: American Association for the Advancement of Science; 2011.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing. Standards for Educational and Psychological Testing. Derwood: American Educational Research Association; 2014.

American Society of Plant Biologists and the Botanical Society of America. Core Concepts and Learning Objectives in Plant Biology for Undergraduates. Derwood: American Society of Plant Biologists and the Botanical Society of America; 2016.

Anderson TR. Bridging the educational research-teaching practice gap. The importance of bridging the gap between science education research and its application in biochemistry teaching and learning: barriers and strategies. Biochem Mol Biol Educ. 2007;35(6):465–70.

Anderson DL, Fisher KM, Norman GJ. Development and evaluation of the conceptual inventory of natural selection. J Res Sci Teach. 2002;39(10):952–78.

Andrews TM, Kalinowski ST, Leonard MJ. "Are humans evolving?" A classroom discussion to change student misconceptions regarding natural selection. Evolution. 2011;4(3):456–66.

Andrews TM, Price RM, Mead LS, McElhinny TL, Thanukos A, Perez KE, et al. Biology undergraduates' misconceptions about genetic drift. CBE Life Sci Educ. 2012;11(3):258–9.

Association of American Medical Colleges and the Howard Hughes Medical Institute. Report of Scientific Foundations for Future Physicians Committee. 2009. http://www.hhmi.org/grants/pdf/08-209_AAMC-HHMI_report.pdf.

Balch WR. Item order affects performance on multiple-choice exams. Teach Psychol. 1989;16(2):75–7.

Baum DA, Smith SDW, Donovan SSS. The tree-thinking challenge. Science. 2005;310(5750):979–80.

Bishop BA, Anderson CW. Student conceptions of natural selection and its role in evolution. J Res Sci Teach. 1990;27:415–27.

Blacquiere LD, Hoese WJ. A valid assessment of students' skill in determining relationships on evolutionary trees. Evolution. 2016;9(1):5.

Bowling BV, Acra EE, Wang L, Myers MF, Dean GE, Markle GC, et al. Development and evaluation of a genetics literacy assessment instrument for undergraduates. Genetics. 2008;178(1):15–22.

Brownell SE, Tanner KD. Barriers to faculty pedagogical change: lack of training, time, incentives, and tensions with professional identity? CBE Life Sci Educ. 2012;11(4):339–46.

Brownell SE, Freeman S, Wenderoth MP, Crowe AJ. BioCore Guide: a tool for interpreting the core concepts of vision and change for biology majors. CBE Life Sci Educ. 2014;13(2):200–11.

D'Avanzo C. Biology concept inventories: overview, status, and next steps. Bioscience. 2008;58(11):1079.

DeSaix J, Katcher J, Urry L, Young C, Bridges C, Frost J. Natural selection misconceptions diagnostic. 2011. https://evolution.berkeley.edu/evolibrary/teach/evo_misconceps_diagnostic.pdf https://evolution.berkeley.edu/evolibrary/search/lessonsummary.php?&thisaudience=13-16&resource_id=426. Accessed 1 Jan 2018.

Federer MR, Nehm RH, Opfer JE, Pearl D. Using a constructed-response instrument to explore the effects of item position and item features on the assessment of students' written scientific explanations. Res Sci Educ. 2015;45(4):527–53.

Furtak EM, Morrison D, Iverson H, Ross M, Heredia S. A conceptual analysis of the Conceptual Inventory of Natural Selection: improving diagnostic utility through within item analysis. Presented at the National Association of Research in Science Teaching Annual Conference. 2011. https://www.researchgate.net/publication/266347441 http://spot.colorado.edu/furtake/Furtak_etal_NARST2011_FINAL.pdf.

Garvin-Doxas K, Klymkowsky M, Elrod S. Building, using, and maximizing the impact of concept inventories in the biological sciences: report on a National Science Foundation-sponsored conference on the construction of concept inventories in the biological sciences. CBE Life Sci Educ. 2007;6:277–82.

Govindan B. Bacterial Survivor: an interactive game that combats misconceptions about antibiotic resistance. J Microbiol Biol Educ. 2018. https://doi.org/10.1128/jmbe.v19i3.1675.

Gregory TR. Understanding evolutionary trees. Evolution. 2008;1(2):121–37.

Gregory TR. Understanding natural selection: essential concepts and common misconceptions. Evolution. 2009;2(2):156–75.

Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. Appl Meas Educ. 2002;15(3):309–33.

Hambleton RK, Traub RE. The effects of item order on test performance and stress. J Exp Educ. 1974;43(1):40–6.

Handelsman J, Ebert-May D, Beichner R, Bruns P, Chang A, DeHaan R, et al. Scientific teaching. Science. 2004;304(5670):521–2.

Henderson C, Dancy MH. Barriers to the use of research-based instructional strategies: the influence of both individual and situational characteristics. Phys Rev Special Top. 2007;3(2):020102.

Henderson C, Beach A, Finkelstein N. Facilitating change in undergraduate STEM instructional practices: an analytic review of the literature. J Res Sci Teach. 2011;48(8):952–84.

Herron JD, Nurrenbern SC. Chemical education research: improving chemistry learning. J Chem Educ. 1999;76(10):1353.

Hiatt A, Davis GK, Trujillo C, Terry M, French DP, Price RM, et al. Getting to evo-devo: concepts and challenges for students learning evolutionary developmental biology. CBE Life Sci Educ. 2013;12(3):494–508.

Kalinowski ST, Leonard MJ, Andrews TM, Litt AR. Six classroom exercises to teach natural selection to undergraduate biology students. CBE Life Sci Educ. 2013;12(3):483–93.

Kalinowski ST, Leonard MJ, Taper ML. Development and validation of the conceptual assessment of natural selection (CANS). CBE Life Sci Educ. 2016;15(4):ar64.

Kempa R. Research and research utilisation in chemical education. Chem Educ Res Pract. 2002;3(3):327–43.

Klymkowsky MW, Underwood SM, Garvin-Doxas RK. Biological concepts instrument (BCI): a diagnostic tool for revealing student thinking. 2010. http://arxiv.org/abs/1012.4501.

Knight JK. Biology concept assessment tools: design and use. Microbioly Aust. 2010;31:1–4.

Libarkin J. Concept inventories in higher education science. National Research Council. 2008. p. 1–13. http://www7.nationalacademies.org/bose/Libarkin_CommissionedPaper.pdf.

Madsen A, McKagan SB, Sayre EC. Best practices for administering concept inventories. Phys Teac. 2017;55(9):530–6.

Marbach-Ad G, Briken V, El-Sayed NM, Frauwirth K, Fredericksen B, Hutcheson S, et al. Assessing student understanding of host pathogen interactions using a concept inventory. J Microbiol Biol Educ. 2009;10(1):43–50.

Marbach-Ad G, McAdams KC, Benson S, Briken V, Cathcart L, Chase M, et al. A model for using a concept inventory as a tool for students' assessment and faculty professional development. CBE Life Sci Educ. 2010;9:408–16.

Meisel RP. Teaching tree-thinking to undergraduate biology students. Evolution. 2010;3(4):621–8.

Merkel S, Reynolds J, Siegesmund A, Smith A, Chang A. Education Heidi Smith A. Recommended curriculum guidelines for undergraduate microbiology education. JMBE. 2012;13(1):32.

Moharreri K, Ha M, Nehm RH. EvoGrader: an online formative assessment tool for automatically evaluating written evolutionary explanations. Evolution. 2014;7(1):15.

Nadelson L, Southerland S. Development and preliminary evaluation of the measure of understanding of macroevolution: introducing the MUM. J Exp Educ. 2009;78(2):151–90.

Naegle E. Patterns of Thinking about Phylogenetic Trees: A Study of Student Learning and the Potential of Tree Thinking to Improve Comprehension of Biological Concepts (Ph.D. Thesis). Idaho State University; 2009. .

National Research Council. BIO2010: transforming undergraduate education for future research biologists. New York: Academies Press; 2003.

National Research Council. A New Biology for the 21st century. 2009. http://nap.edu/12764.

National Research Council. Thinking evolutionarily: evolution education across life sciences. Summary of a convocation. Steve Olson, Rapporteur. Planning Committee on Thinking Evolutionarily: making Biology Education Make Sense. Board on Life Sciences, Division on Earth and Life Studies, National Research Council, and National Academy of Sciences.

Washington, DC: The National Academies Press; 2012. http://www.nap.edu/catalog/13403.

Nehm RH, Ha M. Item feature effects in evolution assessment. J Res Sci Teach. 2011;48(3):237–56.

Nehm RH, Reilly L. Biology majors' knowledge and misconceptions of natural selection. Bioscience. 2007;57(3):263–72.

Nehm RH, Schonfeld IS. Measuring knowledge of natural selection: a comparison of the CINS, an open-response instrument, and an oral interview. J Res Sci Teach. 2008;45(10):1131–60.

Nehm RH, Ha M, Rector M, Opfer J, Perrin L, Ridgway J, et al. Scoring Guide for the Open Response Instrument (ORI) and Evolutionary Gain and Loss Test (EGALT) Draft 2. Tech Repf Natl Sci Found. 2010;1:1.

Nehm RH, Beggrow EP, Opfer JE, Ha M. Reasoning about natural selection: diagnosing contextual competency using the ACORNS instrument. Am Biol Teach. 2012;74(2):92–8.

Nelson CE. Teaching evolution (and all of biology) more effectively: strategies for engagement, critical reasoning, and confronting misconceptions. Integr Compar Biol. 2008;48:213–25.

Novick LR, Catley KM. Assessing students' understanding of macroevolution: concerns regarding the validity of the MUM. Int J Sci Educ. 2012;34(17):2679–703.

Novick LR, Catley KM. When relationships depicted diagrammatically conflict with prior knowledge: an investigation of students' interpretations of evolutionary trees. Sci Educ. 2014;98(2):269–304.

Opfer JE, Nehm RH, Ha M. Cognitive foundations for science assessment design: knowing what students know about evolution. J Res Sci Teach. 2012;49(6):744–77.

Perez KE, Hiatt A, Davis GK, Trujillo C, French DP, Terry M, et al. The EvoDevoCI: a concept inventory for gauging students' understanding of evolutionary developmental biology. CBE Life Sci Educ. 2013;12(4):665–75.

Price RM, Andrews TC, McElhinny TL, Mead LS, Abraham JK, Thanukos A, et al. The genetic drift inventory: a tool for measuring what advanced undergraduates have mastered about genetic drift. CBE Life Sci Educ. 2014;13(1):65–75.

Reeves TD, Marbach-Ad G. Contemporary test validity in theory and practice: a primer for discipline-based education researchers. CBE Life Sci Educ. 2016;15(1):rm1.

Sadler PM. Psychometric models of student conceptions in science: reconciling qualitative studies and distractor-driven assessment instruments. J Res Sci Teach. 1998;35(3):265–96.

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

Smith MK, Wood WB, Knight JK. The genetics concept assessment: a new concept inventory for gauging student understanding of genetics. CBE Life Sci Educ. 2008;7:422–30.

Smith JI, Combs ED, Nagami PH, Alto VM, Goh HG, Gourdet MAA, et al. Development of the biology card sorting task to measure conceptual expertise in biology. CBE Life Sci Educ. 2013a;12(4):628–44.

Smith JJ, Cheruvelil KS, Auvenshine S. Assessment of student learning associated with tree thinking in an undergraduate introductory organismal biology course. CBE Life Sci Educ. 2013b;12(3):542–52.

Stanhope L, Ziegler L, Haque T, Le L, Vinces M, Davis GK, et al. Development of a biological science quantitative reasoning exam (BioSQuaRE). CBE Life Sci Educ. 2017;16(4):ar66.

Steif PS, Hansen MA. New practices for administering and analyzing the results of concept inventories. J Eng Educ. 2007;96(3):205–12.

Summers MM, Couch BA, Knight JK, Brownell SE, Crowe AJ, Semsar K, et al. EcoEvo-MAPS: an ecology and evolution assessment for introductory through advanced undergraduates. CBE Life Sci Educ. 2018;17(2):ar18.

Tansey JT, Baird T, Cox MM, Fox KM, Knight J, Sears D, et al. Foundational concepts and underlying theories for majors in "biochemistry and molecular biology". Biochem Mol Biol Educ. 2013;41(5):289–96.

Tornabene RE, Lavington E, Nehm RH. Testing validity inferences for genetic drift inventory scores using Rasch modeling and item order analyses. Evolution. 2018. https://doi.org/10.1186/s12052-018-0082-x.

Wiggins GP, McTighe J. Understanding by design. Alexandria: Association for Supervision and Curriculum Development; 2005.

Ziadie MA, Andrews TC. Moving evolution education forward: a systematic analysis of literature to identify gaps in collective knowledge for teaching. CBE Life Sci Educ. 2018;17(1):ar11.