Accepting and Understanding Evolution: The Development and Evaluation of Measurement Tools in Evolution Education

by

Anastasia Misheva

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved May 2023 by the
Graduate Supervisory Committee:

Sara Brownell, Chair
M. Elizabeth Barnes
James Collins
Katelyn Cooper
Beckett Sterner

ARIZONA STATE UNIVERSITY

August 2023

ABSTRACT

Evolution is a key feature of undergraduate biology education: the American Association for the Advancement of Science (AAAS) has identified evolution as one of the five core concepts of biology, and it is relevant to a wide array of biology-related careers. If biology instructors want students to use evolution to address scientific challenges post-graduation, students need to be able to apply evolutionary principles to real-life situations, and accept that the theory of evolution is the best scientific explanation for the unity and diversity of life on Earth. In order to help students progress on both fronts, biology education researchers need surveys that measure evolution acceptance and assessments that measure students' ability to apply evolutionary concepts. This dissertation improves the measurement of student understanding and acceptance of evolution by (1) developing a novel Evolutionary Medicine Assessment that measures students' ability to apply the core principles of Evolutionary Medicine to a variety of health-related scenarios, (2) reevaluating existing measures of student evolution acceptance by using student interviews to assess response process validity, and (3) correcting the validity issues identified on the most widely-used measure of evolution acceptance - the Measure of Acceptance of the Theory of Evolution (MATE) - by developing and validating a revised version of this survey: the MATE 2.0.

ACKNOWLEDGMENTS

I wish to thank my PhD advisor Dr. Sara Brownell for all of the project guidance and feedback, professional advice, and encouragement that she has provided during my years of working with her. In particular, the combination of understanding, flexibility, and structure that Dr. Brownell provided throughout the long phase of remote work during the pandemic enabled me to remain on my intended schedule for graduation and not fall behind despite all the difficulties that 2020 brought. I also wish to thank my committee member Dr. Elizabeth Barnes for her role as my frequent collaborator and mentor. We began working on a project together when she was a postdoctoral researcher and I was a newly-arrived graduate student in Dr. Brownell's lab, and have remained close colleagues and friends through her transition from postdoc at Arizona State University to faculty member at Middle Tennessee State University. I also thank Dr. Katelyn Cooper, Dr. James Collins, and Dr. Beckett Sterner for their thoughtful feedback on all of the projects included in this dissertation.

Finally, I wish to thank all of my collaborators, whose input made this dissertation possible. Dr. Daniel Grunspan and Dr. Randolph Nesse worked with me on Chapter 2 of this dissertation. Dr. Grunspan taught me how to conduct the full process of assessment validation, while Dr. Nesse provided essential advice and feedback as the senior expert in Evolutionary Medicine. Dr. Elizabeth Barnes worked with me on Chapters 3 and 4 of this dissertation; she taught me how to conduct an interview study and we had extensive theoretical discussions about what it means for a student to accept evolution. Dr. Supriya and Dr. Michael Rutledge worked with me on Chapter 4 of this dissertation. Dr. Supriya assisted with the statistical analyses for survey validation, and Dr. Rutledge provided valuable input on my revision of his previously developed survey tool. And as my advisor, Dr. Brownell provided guidance on each of these three projects.

As a student of ASU, I acknowledge that the Tempe campus sits on the ancestral homelands of those American Indian tribes that have inhabited this place for centuries, including the Akimel O'odham (Pima) and Pee Posh (Maricopa) peoples.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Evolution is key feature of undergraduate biology education. The National Academy of Sciences (NAS) has described evolution as “the central organizing principle of modern biology,” and the American Association for the Advancement of Science (AAAS) identified evolution as one of the five core concepts of undergraduate biology education in its Vision and Change report, which was produced by over 500 biologists and biology educators across the U.S. (American Association for the Advancement of Science, 2011; Brownell et al., 2014; National Academy of Sciences, 2008). This is because evolution is essential for not only those who aim to become evolutionary biologists, but for students planning to go into a wide array of careers in medical practice, biomedical research, public health, K-12 science education, and agriculture (Bull & Read, 2019; Graves et al., 2016; Natterson-Horowitz et al., 2023; Sickel & Friedrichsen, 2013a; Thrall et al., 2011). Notably, evolutionary biologists have called for the inclusion of Evolutionary Medicine (EvMed) in medical and pre-medical education because it applies an evolutionary lens to deepen our understanding of health and disease (Antolin et al., 2012; Graves et al., 2016; Nesse et al., 2010; Power et al., 2020).

If we as educators want our students to use evolution to address scientific challenges post-graduation, our students need to be able to apply evolutionary principles to real-life situations, and accept that the theory of evolution is the best scientific explanation for the unity and diversity of life on Earth. Acceptance – and not just understanding – of evolution is important because individuals who do not accept evolution have little reason to choose to use it once they are no longer in the classroom; for example, studies have found that science teachers who do not accept evolution are less likely to teach it in a thorough and scientifically accurate manner (Berkman & Plutzer, 2011; Plutzer et al., 2020; Sickel & Friedrichsen, 2013a).

Surveys and assessments with validation evidence are essential tools for research on student acceptance and understanding of evolution. These measurement tools are what allow biology education researchers to measure the variables of interest and draw theoretically valid inferences from those measurements (Mead et al., 2019a; Nehm & Mead, 2019). In order to produce a survey or assessment suitable for general research use, researchers assess the validity and reliability of a measurement tool using a variety of qualitative (e.g., expert reviews, student interviews) and quantitative (e.g., Cronbach's α, Classical Test Theory, Rasch analysis) methods (American Educational Research Association et al., 2014; Artino et al., 2014; Campbell & Nehm, 2013; Mead et al., 2019a). Validity evidence for a measurement tool can be gathered using many different sources of evidence and types of analyses; as such, it is common practice for researchers to re-evaluate and revise important measurement tools to strengthen the quality of the inferences that can be draw from the research data (Glaze et al., 2020a; Romine et al., 2017, 2018a; Sbeglia & Nehm, 2018a, 2019).

The goal of this dissertation is to improve the measurement of student understanding and acceptance of evolution by:

(1) Developing a novel EvMed Assessment that measures students' ability to apply the core principles of EvMed to health-related scenarios;

(2) Re-evaluating existing measures of student evolution acceptance using an aspect of survey validation (response process validity) that has thus far been neglected during the development of measures of evolution acceptance; and

(3) Using the results from the previous study to revise an existing measure of student evolution acceptance so as to increase the validity of the inferences that can be drawn from the survey.

CHAPTER 2

THE EVMED ASSESSMENT: A TEST FOR MEASURING STUDENT UNDERSTANDING OF CORE CONCEPTS IN EVOLUTIONARY MEDICINE

INTRODUCTION

Evolutionary medicine (EvMed) is a relatively new and rapidly changing field that applies an evolutionary lens to deepen our understanding of health and disease (Bull & Read, 2019; Moltzau Anderson & Horn, 2020; Natterson-Horowitz et al., 2023). Given its potential to influence medical practice and biomedical research, there has been a call to include EvMed in medical and pre-medical education (Antolin et al., 2012; Basile et al., 2018; Graves et al., 2016; Moltzau Anderson & Horn, 2020; Nesse et al., 2010; Power et al., 2020). Efforts to answer this call can be seen in the growing number of courses teaching EvMed at the postsecondary level (Grunspan et al., 2019; Hidaka et al., 2015). A majority of research-intensive universities in the United States now offer at least one undergraduate course that either focuses on EvMed exclusively, or includes EvMed as a unit within a more general evolution or health-related curriculum (Grunspan et al., 2019). These courses are typically housed within Biology or Anthropology departments, or cross-listed between the two (Grunspan et al., 2019). The increasing prevalence of EvMed instruction at the undergraduate level has created a need for pedagogical resources.

This need has been partially addressed by the growing number of instructor resources for teaching EvMed, including textbooks (Brüne & Schiefenhövel, 2019; Gluckman et al., 2016; Perlman, 2013; Zimmer, 2013), websites with lesson plans or activities (International Society for Evolution, Medicine, and Public Health, n.d.; Michigan State University, n.d.), and journal articles that include insights for teaching EvMed (Graves et al., 2016; V. H. Smith et al., 2015). While these resources focus on *what* to teach in EvMed or *how* to teach EvMed, resources that help assess student understanding of EvMed are largely absent. To fill this gap, I developed and validated the **E**v**M**ed **A**ssessment (EMA), a closed-ended exam that consists of 45 true-or-false items across 11 question sets.

Assessment and core principles in backward design

The EMA was developed as part of a larger effort to assemble pedagogical resources for EvMed using backward design (International Society for Evolution, Medicine, and Public Health, n.d.). Backward design is an educational method that stresses the importance of alignment between all curricular components by moving "backward" when developing a curriculum (Neiles & Arnett, 2021; Reynolds & Kearns, 2017; Wiggins & McTighe, 2005). With backward design, the instructor first specifies the learning goals, then develops assessments to evaluate mastery of the learning goals, and lastly develops instructional materials that match the assessments and learning goals. The key benefit of this approach is that it facilitates internal alignment between what students are taught, what is assessed, and which concepts the instructor deems most important (Biggs, 1996; Orr et al., 2022; Teasdale & Aird, 2023).

To provide EvMed instructors with a resource for setting discipline-relevant learning goals, my collaborators (Daniel Grunspan, Randolph Nesse, and Sara Brownell) previously conducted a study that sought to identify the theoretical principles and key ideas that underlie EvMed. In that study, my collaborators iteratively surveyed a panel of 56 biologists, anthropologists, medical doctors, and other researchers who contribute to the EvMed field through publications and participation in the International Society for Evolution, Medicine and Public Health (ISEMPH) (Grunspan et al., 2018). This survey yielded a list of 14 core principles of EvMed: (1) types of explanation (proximate vs. ultimate causes), (2) evolutionary processes, (3) reproductive success, (4) sexual selection, (5) constrains on adaptation, (6) general evolutionary trade-offs, (7) life history theory, (8) levels of selection, (9) phylogeny, (10) coevolution, (11) developmental plasticity, (12) immune defenses, (13) environmental mismatch, and (14) and the impact of cultural practices on health and disease (Grunspan et al., 2018). The inclusion of each core principle on this list was approved by at least 80% of the panelists. Agreement with the inclusion of each principle ranged from 80% for sexual selection to 100% for types of analyses, evolutionary processes, cultural practices, and life history theory.

These core principles provide a useful guideline for EvMed instructors as they construct course learning goals (Orr et al., 2022) and can offer guidance on what to assess. Although instructors can write their own assessment questions aligned with the core principles, an assessment that has been tested and validated across multiple contexts can be useful for instructors to help align instruction with the core principles.

Assessments with validation evidence

Assessment validation is the process of evaluating the extent to which an assessment is measuring what it claims to be measuring. Assessments with validation evidence have been rigorously tested to demonstrate that the inferences drawn from them are accurate and precise (AERA et al., 2014; Campbell & Nehm, 2013; Mead et al., 2019). Concept inventories (D. L. Anderson et al., 2002; Deane et al., 2014; Kalas et al., 2013; Price et al., 2014) and programmatic assessments (Couch et al., 2019; Semsar et al., 2019; Summers et al., 2018) are two common types of assessments with validation evidence. Concept inventories are designed to measure student understanding of a relatively narrow scientific idea, such as natural selection (D. L. Anderson et al., 2002; Nehm et al., 2012), genetic drift (Price et al., 2014), or macroevolution (Nadelson & Southerland, 2009). Often, students are provided with answer choices that are distractors based on common misconceptions. Concept inventories can be used either at the beginning of a course to help instructors identify misconceptions to focus on in their teaching, or at the beginning and end of a course to measure course efficacy at teaching a certain core concept (Furrow & Hsu, 2019; Knight, 2010). Alternatively, programmatic assessments have been designed for an entire field, such as ecology and evolution (Summers et al., 2018), physiology (Semsar et al., 2019), and general biology (Couch et al., 2019; M. K. Smith et al., 2019). They typically asses a range of core concepts that have been established by a group of disciplinary experts (e.g., *Vision and Change* report on undergraduate biology education (AAAS, 2011; Branchaw et al., 2020; Brownell et al., 2014)) and are not limited to one particular course. Rather, they assess student learning within a field throughout an entire undergraduate program and may be administered repeatedly across multiple courses. The primary focus for both concept inventories and programmatic assessments is on what aggregate results can reveal about a curriculum's efficacy and collective student learning.

While assessments with validation evidence share structural similarities with instructor-generated course exams, these assessment types differ in both their primary purpose and their design process. Instructor exams are typically written by the instructor or sampled from a test bank, have minimal to no validation evidence, and cover a portion of the course, which may or may not be based on specific learning goals (Wright et al., 2018). While concept inventories and programmatic assessments primarily measure collective student learning, instructor-generated exams primarily measure an individual student's learning to provide a summative assessment (i.e., a grade). Due to their different functions, instructor-generated exams and published assessments with validation evidence are held to different standards of validity and reliability (Campbell & Nehm, 2013; Knight, 2010). Validity and reliability are frameworks for evaluating the quality of the inferences that can be drawn from a measurement tool, such as an assessment or survey (Campbell & Nehm, 2013). Validity addresses the question of whether an assessment truly measures what researchers want it to measure. This is typically evaluated using a mix of quantitative and qualitative methods that examine (1) the extent to which the assessment accurately represents the relevant knowledge domain, (2) the extent to which it requires students to use the skills and thought processes that researchers want them to use, and (3) whether assessment items that are designed to measure the same concept get answered the same way (AERA et al., 2014; Artino et al., 2014;

Campbell & Nehm, 2013; Mead et al., 2019). Meanwhile, reliability addresses the consistency of an assessment's measurements. This is typically evaluated using quantitative methods that look at test-reset consistency across individual students or at the consistency of students' answers across items within a single assessment (AERA et al., 2014; Artino et al., 2014; Campbell & Nehm, 2013; Mead et al., 2019). Creating an exam with this level of validation is time consuming, and thus not practical for typical instructor generated exams.

Many of the EvMed core principles are general evolutionary concepts that are covered within existing concept inventories. For example, phylogenetic tree-thinking can be measured using an assessment made by Kummer et al. (2019), understanding of coevolution can be measured using the Host-Pathogen Interactions Concept Inventory (Marbach-Ad et al., 2009), and understanding of evolutionary processes can be measured using a variety of existing assessments (D. L. Anderson et al., 2002; Kuschmierz, Beniermann, et al., 2020; Nehm et al., 2012). Yet while existing concept inventories can measure student understanding of general evolutionary concepts in EvMed courses, the application of evolution to novel contexts in human health and disease may influence how students operationalize these concepts, which creates a need for a new assessment. For example, affective and cognitive components of student learning in evolution are influenced by the species context during lessons and assessment (Grunspan et al., 2021; Nettle, 2010; Shtulman & Schulz, 2008). Some of the unique applications of evolutionary principles in EvMed may elicit similar differences. Additionally, EvMed specifies certain ideas and concepts from evolution that differ from those emphasized in currently available concept inventories, such as a focus on different levels of selection to explain the behavior of human cancer cells under chemotherapy or the necessary consideration of human cultural practices.

Project Goals

My goal was to provide EvMed instructors with an assessment resource with validation evidence that uses questions at high levels on Bloom's taxonomy (L. W. Anderson & Krathwohl, 2001) to measure student understanding of the EvMed core principles by requiring students to apply the core principles to a range of novel health-related scenarios. The EvMed Assessment shares similarities with both concept inventories and programmatic assessments, and has differences from both. The EMA is like a programmatic assessment in that it covers multiple core concepts within a field, rather than a single relatively narrow idea. The EMA is also like a concept inventory in that I foresee it being used primarily for individual EvMed courses, rather than across multiple courses in an undergraduate program, due to EvMed being a somewhat niche field compared to ecology or physiology. Furthermore, while the EMA strives to identify common student misconceptions (as concept inventories and programmatic assessments do (Furrow & Hsu, 2019)), student misconception have not yet been well-studied for core concepts that are unique to EvMed. I acquired validity evidence for this assessment using expert feedback, student interviews, and by administering the assessment in 11 different EvMed courses in universities throughout the United States.

DEVELOPMENT

Assessment format and question development

The EMA uses a modified version of the multiple-true-false (MTF) format, in which each question contains an informational premise (stem) followed by several true-or-false statements (items) (Figure 1) (Couch et al., 2019; Semsar et al., 2019; Summers et al., 2018). The advantages of the MTF format over the standard multiple-choice format are that (1) students can answer more items in a set amount of time, (2) the assessment can be used to explore student understanding of multiple principles as applied to the same scenario, and (3) the MTF format has been shown to mimic free-response reasoning by revealing individual misconceptions within a chain of reasoning (Brassil & Couch, 2019; Couch et al., 2019; Frisbie, 1992; Kalas et al., 2013; Price et al., 2014). Similar to several existing assessments, I modified the MTF format by replacing the options of 'true' and 'false' with 'likely' and 'unlikely' (Semsar et al., 2019; Summers et al., 2018). This helped accommodate the inherent uncertainty of applying evolutionary principles to health-related topics and made it easier to create items that were scientifically accurate and sufficiently difficult. This approach reflects the tentative nature of science (Mueller & Reiners, 2022; Popper, 1959; University of California Berkeley, 2021a) and has previously been used in evolution and ecology assessment (Summers et al., 2018).

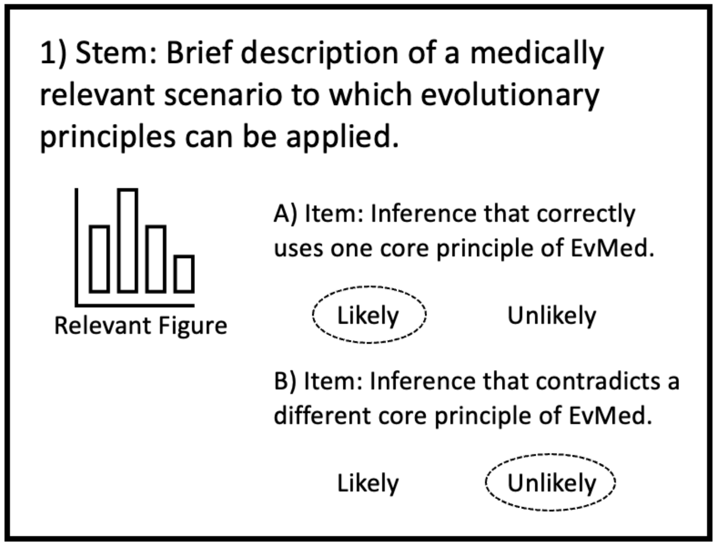


Figure 1. Question format on the EvMed Assessment.

The question stems describe a scenario in one to two paragraphs, often accompanied by a table or figure. These scenarios cover a wide range of health-related topics, including genetic conditions, cancer, infectious disease, aging, and mental health. I included a variety of health-related topics to reflect the broad scope of evolutionary medicine as a discipline (Bull & Read, 2019; Graves et al., 2016; Moltzau Anderson & Horn, 2020; Nesse et al., 2010) and to make the assessment relevant to a range of EvMed courses. For each stem, I wrote items that present a single inference or prediction based on information in the stem, and test competencies in one or more core principles. While I intended for each item to assess primarily one core principle, inherent conceptual overlap led to some items covering additional closely related principles (e.g., life history theory and trade-offs) (Grunspan et al., 2018). The entire set of items within a question covers several core principles that can be applied to the given scenario. To ensure that items assess conceptual understanding rather than knowledge or familiarity with technical terms, I limited the use of scientific and medical jargon.

To develop the first draft of the EvMed Assessment (Version 1.0), a collaborator (D.G.) and I wrote questions based on the core principles of EvMed (Grunspan et al., 2018). Between us, D.G. and I have extensive prior teaching and coursework experience in evolutionary medicine and a general familiarity with the EvMed literature. We wrote stems by condensing information from a wide range of sources and generating hypothetical scenarios based on well-studied phenomena. This is a standard approach to question development for concept inventories that require students to transfer conceptual knowledge to novel scenarios (Couch et al., 2019; Gormally et al., 2012; Price et al., 2014; Summers et al., 2018). Additional collaborators (R.N. and S.B.) provided feedback on questions. R.N. is an international expert in evolutionary medicine, having written several books on the subject, and S.B. is an expert in assessment development in biology education. Version 1.0 of the assessment included a total of 70 likely/unlikely items across 14 question stems (Figure 2). This version included items covering every principle except sexual selection, where we struggled to identify examples of evolutionary inferences that could be conveyed with a brief description, had a clear-cut answer (Morrow, 2015), and did not resort to gender essentialism.

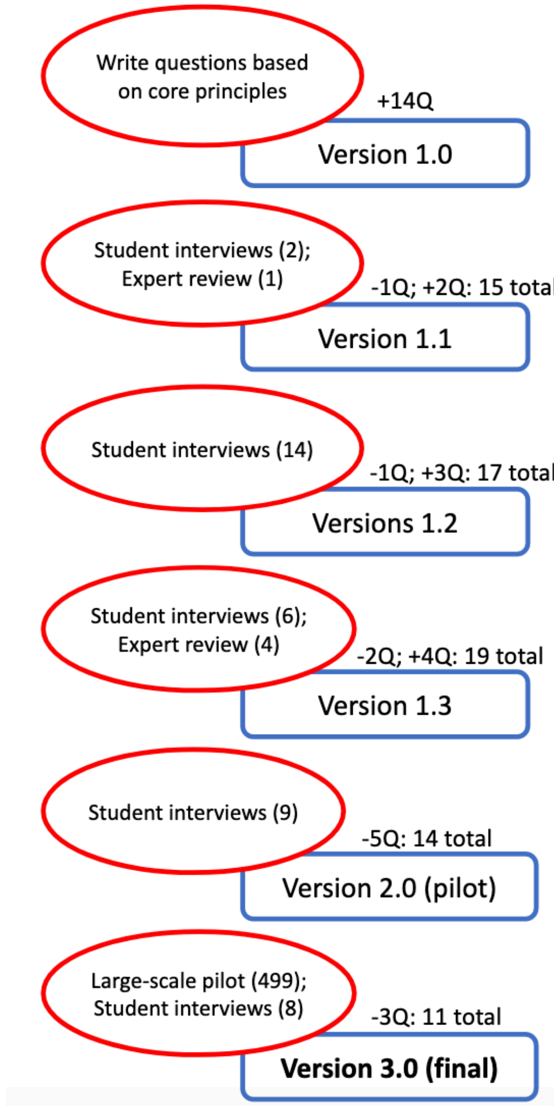


Figure 2. Development and validation process of the EvMed Assessment. Parenthetical numbers indicate sample size; Q stands for "questions."

Qualitative validation: Version 1.0 to 2.0

After Version 1.0 was written, I revised it through iterative evaluation and improvement of its content and substantive validity. The evaluation took place through multiple rounds of student interviews and expert reviews, with revisions made to address issues identified with individual questions and items. See Figure 2 for an overview of the entire assessment development process.

Content validity addresses whether the assessment offers an accurate representation of the relevant knowledge domain. It is typically evaluated using expert review (AERA et al., 2014; Artino et al., 2014; Campbell & Nehm, 2013; Mead et al., 2019). To assess content validity, seven scholars active in teaching or conducting research in EvMed reviewed the assessment for accuracy on two separate occasions. A single expert (R.N.) reviewed the initial draft of the assessment (Version 1.0). The other six experts (one PhD student, four postdoctoral scholars and one Assistant Professor) reviewed the updated Version 1.2. The reviewers were asked to (1) point out any information that is inaccurate or presented as being more scientifically settled than it really is, (2) select what they consider to be the correct answer for each item and rate their certainty of the answer, and (3) provide any other concerns they have (e.g., clarity, difficulty level, etc.). I revised or deleted items if at least one reviewer answered "incorrectly" with high certainty or deemed the item overly complex or debatable. I also addressed any noted issues with accuracy or clarity as recommended by the reviewers.

Substantive validity addresses whether students are using the skills and thought processes that the developers want them to use when answering items on an assessment (AERA et al., 2014; Artino et al., 2014; Campbell & Nehm, 2013; Mead et al., 2019). This involves checking for issues such as students answering correctly despite using incorrect reasoning, or answering incorrectly despite using correct reasoning (e.g., because they misinterpreted what the item was asking). To assess substantive validity, D.G. and I conducted a total of 31 think-aloud interviews with undergraduate students on

Versions 1.0 – 1.3 of the assessment. During an interview, the student would read the question stem out loud, read and answer each item out loud, and explain the reasoning behind their answer choices as they went. Meanwhile, the interviewer would take notes and ask clarifying questions about the student's reasoning. The interviewer would note down if the student (1) interpreted the stems and items correctly, (2) was able to select the correct answer despite using the wrong reasoning, or (3) selected the wrong answer despite using the correct reasoning. The latter two issues were of particular concern because an effective assessment needs to consistently differentiate between correct use of the concepts being tested vs. the presence of misconceptions (Couch et al., 2019; Semsar et al., 2019; Summers et al., 2018).

The 31 think-aloud interviews were distributed across four initial rounds of revision (Figure 2). In order to obtain a sample that reflects the student population for which this assessment is intended, I recruited participants who were currently enrolled in an EvMed course, and recruited late in the semester to ensure they had some level of content knowledge. All interviews were conducted by either T.M. or D.G. and audio recorded. An item or stem was revised any time two or more students displayed the same issue, such as when multiple students were able to answer an item using only common knowledge without engaging with any of the core concepts. This process resulted in Version 2.0 of the assessment, which had 14 questions containing a total of 62 items across all questions (Figure 2).

Quantitative validation of Version 2.0

Data Collection and Processing

A large-scale pilot test of Version 2.0 was conducted to quantitatively evaluate the psychometric properties of the assessment. Given the length and cognitive complexity of the assessment, the potential effect of survey fatigue on student attentiveness and completion rate was a concern (Revilla & Höhne, 2020; Summers et al., 2018). To keep testing time under the recommended 30 minute time limit for most respondents (Revilla & Höhne, 2020; Summers et al., 2018), I reduced the length of the assessment by dividing Version 2.0 into seven Core Questions and seven Supplemental Questions. When taking the assessment, all students received the seven Core Questions along with a random sample of five Supplemental Questions. The goal for the Core Questions was to identify the minimum number of questions needed to cover all EvMed core principles. I categorized items by the principles they test and mapped out which question sets would provide coverage over an appropriate range of principles while also providing a variety of health topics. Through this mapping, I identified seven Core Questions that together provide coverage across all the core principles. The remaining questions were deemed Supplemental.

Version 2.0 of the assessment was administered as an out-of-class assignment to 732 undergraduate students across seven different courses at six institutions (Table 1). The assessment was administered online, and a demographic survey was provided at the end. Of the participating courses, four administered the assessment once in a single semester, two administered it in two consecutive semesters, and one administered it in three consecutive semesters. Thus, the assessment was administered to students in 11 course offerings. Of these, ten were on EvMed and one was on evolution (Table 1). The EMA was administered mainly in EvMed courses near the end of term so that the sample population would reflect the target population for the assessment. However, the focus on EvMed courses made it difficult to obtain a large sample size because EvMed courses tend to be relatively small. To increase the sample size and evaluate how the assessment performs with students who have an evolution background but have not received targeted instruction in EvMed, I also administered the assessment in one high-enrollment Evolution course. Participants received a small amount of extra credit based on completion.

Table 1. Overview of courses that administered Version 2.0 of the EvMed Assessment. Only students who took at least 10 min. to complete the entire assessment were included in the final sample.

Course	**University**	**Number of semesters**	**Total students**	**Final student sample**
EvMed	Non-selective R1 HSI; southwest	1	21	16
Evolution	Non-selective R1 HSI; southwest	1	305	179
EvMed	Non-selective R1; mountain west	2	70	54
EvMed	Non-selective M1; southeast	2	32	25
EvMed	Very selective R1; west coast	1	45	31
EvMed	Selective R1; southeast	1	13	11
EvMed	Selective R1; west coast	3	246	183

Of the 732 students who completed the assessment, 684 (93.4%) consented to their data being used for validation. Because this assessment was not administered as a high stakes assignment, there was concern that some students may have not put in sufficient effort (e.g., quickly clicked through the questions). Thus, D.G. and I inspected how long students took to complete the entire survey and determined that a minimum of 10 minutes appeared to be an appropriate cut-off, which is consistent with the cut-offs used in similar studies (Couch et al., 2019; Summers et al., 2018). This eliminated an additional 184 students from the sample. One additional student was removed because they left most items blank, leaving 499 students (68.3% of initial population) in the final sample. See Table 2 for a summary of participant demographics.

Statistical analyses

Classical test theory (CTT) was used to examine the psychometric properties of the assessment. CTT is a theoretical framework that is widely used to assess educational and psychological measurement tools, including their overall validity and reliability, as well as the performance of individual items within an assessment (Cappelleri et al., 2014; Lord & Novick, 1968). While more sophisticated frameworks such as item response theory (IRT) have more recently been developed, CTT remains appropriate to use when the sample size does not meet the recommended minimum for IRT (Couch et al., 2019; Şahin & Anil, 2017; Semsar et al., 2019).

Table 2. Demographics of students who took part in think-aloud interviews and the pilot study. Interview sample includes 31 students who participated in the initial qualitative validation (Version 1.0 – 1.3) and 8 who participated in additional post-pilot validation (Version 2.0).

Student Characteristic	Interviews (N = 39)		Quantitative Validation (N = 499)	
	N	%	N	%
Gender				
Men	10	25.6%	108	21.6%
Women	29	74.4%	382	76.4%
Nonbinary	0	0%	10	2%
Race/Ethnicity				
Asian/Asian American	9	23.1%	114	22.8%
Black/African American	3	7.7%	18	3.6%
Hispanic/Latinx	11	28.2%	63	12.6%
Native American/Native Hawaiian	1	2.6%	5	1%
White	15	38.4%	228	45.6%
Other	0	0%	72	14.4%
Academic year				
Lower (first year/sophomore)	22	56.4%	140	28%
Upper (junior/senior)	15	38.5%	353	70.6%
Graduate	2	5.1%	5	1%
Major				
Biology-related fields	20	64.5%	n/a	n/a
Anthropology-related fields	5	16.1%	n/a	n/a
Other	6	19.4%	n/a	n/a
Missing	8	20.5%	n/a	n/a
Highest parental education level				
Completed Bachelor's degree	29	76.3%	n/a	n/a
Did not complete Bachelor's degree	9	23.6%	n/a	n/a
First Language				
English	28	71.8%	n/a	n/a
Not English	11	28.2%	n/a	n/a

Assessment-level reliability and validity

An assessment's reliability is based on the consistency of its measurements. One form of reliability is internal consistency, which is a measure of how consistent each participant's responses are across items. Cronbach's α was used to measure the internal consistency of

Version 2.0. For assessments that aim to measure a single construct, internal consistency is considered ‘excellent’ when $\alpha > 0.9$, ‘good’ when $0.8 > \alpha > 0.9$, and ‘acceptable’ when $0.7 > \alpha > 0.8$ (Crocker & Algina, 1986). Cronbach’s α was 0.71 for the entire assessment, and 0.73 for the Core Questions alone. Based on these criteria, the assessment has an acceptable level of internal consistency. However, it is worth noting that these valuations are not straight forward, especially when an assessment is designed to measure more than one construct. Because individual students’ performance may vary between different core principles, these α values (0.71 and 0.73) are likely appropriate.

An assessment’s validity is based on how well the scores reflect the construct of interest. One form of validity is criterion validity, which is based on the idea that test scores should correlate with variables theorized to be associated with the construct being measured. To evaluate criterion validity, students’ performance on the assessment was compared to their year of study and whether they were enrolled in an evolution course or an EvMed course. Test scores were higher for students who were further along in their educational career and likely had a stronger evolution background. A small sample of first year students who did particularly well on the assessment are a notable exception, although they may be exceptional students as evidenced by their enrollment in upper-level courses. Further, it was found that students enrolled in EvMed courses scored higher ($M = 22.38$, SD = 4.53) than students enrolled in the Evolution course ($M = 19.84$, SD = 4.34) ($t(389) = 6.13$, $p<0.001$). This provides a level of criterion validity that this test is measuring students’ EvMed knowledge (Table 3).

Table 3. Mean student scores on the EvMed Assessment, listed according to academic year and course enrollment. Standard errors are in parentheses. One student did not provide their academic year.

	Percent correct on Core	**Percent correct on Core + Supplement**
Year in college		
First (n=8)	65.6 (6.25)	65.2 (5.45)
Second (n = 132)	63.5 (1.19)	62.0 (1.03)
Third (n = 136)	67.3 (1.30)	66.5 (1.07)
Fourth (n = 195)	69.0 (0.99)	67.1 (0.87)
Fifth+ undergraduate (n = 22)	72.6 (2.72)	71.5 (2.34)
Graduate student (n = 5)	83.1 (3.64)	78.3 (4.13)
Course		
EvMed (n = 320)	70.2 (0.79)	68.7 (0.67)
Evolution (n = 179)	62.2 (1.01)	60.9 (0.86)

Item-level statistics

Validity and reliability are properties of the assessment as a whole. However, it is also important to analyze the performance of individual items. To do this, item difficulty and item discrimination were examined. Item difficulty (p) is the percent of students who answered the item correctly. It ranges from 0.0 (everyone answered incorrectly) to 1.0 (everyone answered correctly). While there are no standard cut-off values for item difficulty, some recommend values between 0.3 and 0.7 (Ashraf & Jaseem, 2020; De Champlain, 2010). Item discrimination (D) is a measure of how well an item distinguishes between students with high test scores versus those with low test scores. Discrimination scores range from -1.0 to 1.0; values near 1.0 indicate that mainly high-scoring students answered the item correctly, values near 0.0 indicate that high-scoring and low-scoring students performed equally, and negative values indicate that low-scoring students were more likely to answer correctly (Ashraf & Jaseem, 2020).

Discrimination scores above 0.2 are recommended (Bichi, 2016). Using this cut-off value, items with $D < 0.2$ were flagged for further review, as these items are potentially problematic due to insufficient differentiation. A total of 29 out of 62 items were flagged. Meanwhile items that had both good difficulty scores (0.3 – 0.7) and good discrimination scores (≥0.2) were deemed suitable for inclusion in Version 3.0 without further review.

Final draft of the assessment (Version 3.0)

Creating Version 3.0

Review of the 29 flagged items began with an examination of each item's discrimination score, difficulty score, think-aloud interview results, and expert feedback from the previous validation stage. Two questions were delated because a majority of items within these questions were flagged for low discrimination, which indicates that these questions were performing poorly as a whole. Additionally, a third full question and three individual items from other questions were deleted because expert reviewers had expressed mild misgivings about their overall quality. These items were initially kept in Version 2.0 to further assess their performance, but deleted after they displayed low discrimination. In total, eight flagged items were deleted in this manner.

Conversely, seven flagged items were kept as-is because (1) their discrimination scores were moderately below the recommended cut-off ($D \geq 0.13$), (2) their difficulty scores were within the recommended range, and (3) prior student interviews suggested no validity issues. Based on prior interviews, most of these items appear to display somewhat low discrimination because they test principles for which student misconceptions are particularly widespread.

The remaining 14 flagged items that could not be kept or deleted based on existing information were revised to address potential clarity issues and then re-assessed for substantive validity via additional student interviews. Eight additional interviews were conducted with students from an upper-level biology course. All students had taken at least one upper-level course on evolution, but no courses on EvMed specifically (see Table 2 for demographics). Based on these interviews, one item was deleted and the remaining 13 items were kept for the final version of the assessment.

Overview of Version 3.0

The final version of the EvMed Assessment consists of six Core Questions and five Supplemental Questions. Each question has 3 – 6 items. The six Core Questions present scenarios within six different health-related topics: genetic conditions, autoimmune disorders, infectious disease, cancer, mental health, and drug function. Items in the Core Questions cover all of the EvMed core principles at least once, with the exception of sexual selection (Table 4). In addition to the EvMed core principles, several items also test general principles of biology that are nevertheless relevant within EvMed (e.g., whether antibiotics can treat viruses).

The Supplemental Questions were designated as such because they (1) address core principles that already have good coverage in the Core Questions, and (2) cover health-related topics that are either already covered in the Core Questions (e.g., infectious disease) or pertain to human health only indirectly (e.g., zoology). All Supplemental Questions passed the validity checks and can be added into the EvMed Assessment as one sees fit. A complete copy of the EvMed Assessment is in the **Supplement.**

Table 4. Assessment coverage of the EvMed core principles. Items are listed according to the core principles they test. Some items test two or more closely related core principles. There are no items on sexual selection. Numbers denote questions and letters denote items in the question.

Core Principle	Items (Core)	Items (Supplemental)
Types of explanation	3D	
Evolutionary processes	1C, 3C, 4A, 4B, 4D, 4E, 4F	8A, 9A, 9B, 9C, 9D, 9E, 10A, 10D, 11A
Reproductive success	1A, 1B, 5A	7D, 9A, 9B
Sexual selection		
Constraints on natural selection	5B	
Trade-offs	2B, 2C, 4C	7A, 7B, 7C, 7D, 9A, 9B, 11C
Life history theory	1A, 1B	7A, 7B, 10A, 10B, 10D
Multiple levels of selection	4A, 4B, 4D, 4E, 4F	
Phylogeny	2E, 6A	10C, 10D
Coevolution	6B, 6C	
Plasticity	5D	8A, 8B, 8C
Defenses	3A, 3B, 3C, 5A, 5B, 5C, 5D, 6C	
Mismatch	5D, 6C	
Cultural Practices	2A	11C
General Biology	2D, 3C	1D, 11B

DISCUSSION

Using the EvMed Assessment

The EMA is a multi-purpose assessment tool. Instructors can use the EMA to (1) gather data on student learning to improve their instruction, (2) identify which core principles students find most difficult, and (3) draw on a pool of questions with validity evidence when designing in-class activities or course exams. When administered at the very beginning of a course, the EMA can be used to identify which core principles are least familiar to students and adjust instruction for that term accordingly. When administered at the end of a course, the EMA can be used to identify which core principles students continue to struggle with and revise future iterations of the course to address those principles more effectively. Furthermore, instructors can also use the EMA as a test bank of high Bloom's level questions with validation evidence, which can be used for formative assessments (e.g., worksheets or clicker questions) and summative assessments (e.g., exams) alike. When using the EMA as a test bank, instructors are encouraged to pick and choose content from both the Core Questions and Supplemental Questions to match the learning goals of the lesson or unit.

Limitations and Future Directions

While the EMA is supported by a range of validity evidence, it is not quite a programmatic assessment nor is it quite a concept inventory. Both types of assessments with validation evidence are typically developed within mature fields that have ample instructional resources and a broad literature base of discipline-based education research (Brownell et al., 2014; Furrow & Hsu, 2019). These resources provide a literature base on common student misconceptions, which the developers of a programmatic assessment or concept inventory can draw upon when designing their assessment. Because EvMed education is still a relatively novel area of research (Antolin et al., 2012; Graves et al., 2016; Grunspan et al., 2018, 2019; Nesse et al., 2010), I did not have an extensive literature base on student misconceptions about EvMed core principles to draw upon, and instead had to rely on general misconceptions about evolution (University of California Berkeley, 2021b) and my and my collaborators' observations during teaching and learning. Yet this state of EvMed education research is what makes the EMA a valuable early foray into developing assessment resources for EvMed. I encourage other researchers to build on our current work by developing and refining additional assessment resources for EvMed.

CHAPTER 3

"IT'S MORE OF A ME THING THAN AN EVOLUTION THING:" EXPLORING THE VALIDITY OF EVOLUTION ACCEPTANCE MEASURES

INTRODUCTION

What is evolution acceptance and why does it matter?

Evolution is "the central organizing principle of modern biology" (National Academy of Sciences, 2008), and as such has been identified as one of the five core concepts for biology education (American Association for the Advancement of Science, 2011; Brownell et al., 2014). Despite its central role in biology, evolution remains socially controversial among both the general public (Gallup Inc, 2019; Miller et al., 2022; Pew Research Center, 2019) and among college biology students in the United States (Dunk & Wiles, 2018; Ferguson & Jensen, 2021; Laidlaw et al., 2022; Siciliano-Martina & Martina, 2020). It is important for students not only to understand evolution, but to also accept it as the best scientific explanation for the unity and diversity of life on Earth (Nadelson & Hardy, 2015; Sickel & Friedrichsen, 2013; Smith, 2010). Individuals who reject one or more aspects of evolution are unlikely to apply evolutionary concepts to solve biology-related problems once they leave the classroom (Smith, 2010; Smith & Siegel, 2016). While most undergraduate biology students do not go on to become academic researchers (U.S. Bureau of Labor Statistics, 2019), the applications of evolution are not limited to evolutionary biology research. Both evolutionary processes and evolutionary history are relevant to biomedical research (e.g., selecting appropriate animal models, impact of evolutionary history on patterns of human health and disease), public health (e.g., risks of zoonotic disease transfer), agriculture (e.g., importance of genetic diversity in crops), and teaching biology to the next generation of students (Grunspan et al., 2018; Nadelson & Hardy, 2015; Preuss & Robert, 2014; Rühli & Henneberg, 2013; Sickel & Friedrichsen, 2013).

While the precise definition of what it means for a student to "accept evolution" has been debated by the evolution education research community (Ingram & Nelson, 2006; Sinatra et al., 2003; Smith, 2010; Smith & Scharmann, 1999; Smith & Siegel, 2016; Southerland et al., 2001; Wiles, 2014), some common themes and similar definitions have emerged. One common theme is that *acceptance of evolution* is distinct from *understanding of evolution* in that understanding of evolution pertains to one's awareness of factual information about evolution, while acceptance is about whether one agrees that the theory of evolution is the best available explanation for the development of life on Earth (Ingram & Nelson, 2006; Sinatra et al., 2003; Southerland et al., 2001; Wiles, 2014). Studies have found that students can accept the theory of evolution as being generally true despite displaying misconceptions about how evolution works (Kuschmierz et al., 2021; Sinatra et al., 2003). Conversely, students can score well on assessments of evolution understanding despite rejecting the veracity of evolutionary theory (Ingram & Nelson, 2006; Sinatra et al., 2003). Researchers also make a distinction between acceptance of evolution and understanding the nature of science (NOS). NOS encompasses ideas within the philosophy of science about how to distinguish between scientific vs. non-scientific questions, what constitutes scientific evidence, and the tentative nature of science (Smith, 2010; Smith & Scharmann, 1999). While a student's understanding of NOS could potentially influence their acceptance of evolution, these are two conceptually distinct constructs. Most recently, a diverse group of 19 researchers convened to discuss the measurement of evolution acceptance. One product of the meeting was a consensus definition that defines evolution acceptances as, "agreement that evolution is valid and the best explanation from science for the unity and diversity of life on Earth, which includes speciation, the common ancestry of life, and that humans evolved from non-human ancestors" (Barnes et al., under review).

One of the primary uses of measuring student acceptance of evolution is to identify if and how evolution acceptance is associated with understanding of evolution, understanding of NOS, and other potentially related constructs. Information about such correlations can be used to inform the design of instructional strategies that aim to increase students' evolution acceptance. However, *how* evolution acceptance is measured may influence results and lead to inconsistencies across studies. For example, some studies have found strong correlations between understanding and acceptance of evolution (Rutledge & Warden, 2000; Stanisavljevic et al., 2013; Trani, 2004). However, others have found weak correlations (Athanasiou & Papadopoulou, 2012; Cavallo et al., 2011; Deniz et al., 2008; Nadelson & Sinatra, 2009), or no correlation at all (Brem et al., 2003; Lawson, 1983; Sinatra et al., 2003). Similarly, some studies indicate that evolution acceptance is correlated with understanding of NOS (Dunk et al., 2017; Rutledge & Warden, 2000), whereas other studies have failed to increase acceptance by increasing understanding of NOS (Cofré et al., 2017; Coleman et al., 2015). Such inconsistencies can make it difficult for researchers to identify the most promising approaches for increasing evolution acceptance.

How can measurement of evolution acceptance influence study results?

Differences between study populations can contribute to differences in evolution acceptance research findings, but inconsistency in the measurement of evolution acceptance also plays a role. Before the 2000s, evolution education researchers had no standardized measure of evolution acceptance. Researchers developed unique survey tools that were often used in a single study. These surveys differed in item wording, number of items, and range of answer choices for each item (Koevering & Stiehl, 1989; Lawson, 1983; Zimmerman, 1987). These differences in measurement limited researchers' ability to compare findings across studies. The 1999 publication of the Measure of Acceptance of the Theory of Evolution (MATE) was an important step forward in evolution education research because it provided the education research community with a peer-reviewed, publicly available survey tool supported by multiple forms of validity evidence (Rutledge & Warden, 1999). The MATE proved to be a popular tool, as usage of the MATE in education research rapidly grew following its publication (Kuschmierz, Meneganzin, et al., 2020; Mead et al., 2019).

Soon after the MATE was published, researchers started to voice concerns about how accurately it measured students' acceptance of evolution (Barnes et al., 2019a; Nadelson & Southerland, 2012; Romine et al., 2018; Sickel & Friedrichsen, 2013; Smith, 2010). These concerns led to the development of other measures of evolution acceptance, namely the Inventory of Student Evolution Acceptance (I-SEA: Nadelson & Southerland, 2012) and the Generalized Acceptance of Evolution Evaluation (GAENE: Smith et al., 2016). Like the MATE, the I-SEA and the GAENE are multi-item survey tools that employ a 5-point Likert scale ranging from “strongly agree” to “strongly disagree;” both were designed for general use by instructors and education researchers and published in stand-alone articles with accompanying validity evidence.

The I-SEA sought to improve upon the original MATE by dividing the measurement of evolution acceptance into three subscales for microevolution, macroevolution, and human evolution (Nadelson & Southerland, 2012). The authors made these distinctions based on prior research showing that students perceive differences between micro- and macroevolution, and between human and non-human evolution, even though these concepts are not biologically distinct (Nehm & Ha, 2011; Reznick & Ricklefs, 2009). Smith and colleagues (2016) then sought to improve upon both the MATE and the I-SEA by creating the GAENE 2.1, a new measure based on an explicit definition of evolution acceptance as, “The mental act or policy of deeming, positing, or postulating that the current theory of evolution is the best current available scientific explanation of the origin of new species from preexisting species.” Four years after its publication, the GAENE 2.1 was updated into the GAENE 3.0 in an effort to make the instrument more psychometrically robust by eliciting a broader distribution of scores (Glaze et al., 2020).

Although the development of new instruments can lead to improvements in the measurement of evolution acceptance, it can also create novel challenges. Though the

MATE, I-SEA, and GAENE 3.0 share similarities, these survey tools each consist of a unique set of items that focus on slightly different aspects of evolution acceptance (Glaze et al., 2020; Nadelson & Southerland, 2012; Rutledge & Warden, 1999; Smith et al., 2016). Even though these survey tools are intended to only measure evolution acceptance, there is emerging evidence that some items conflate evolution acceptance and other constructs such as understanding of evolution and understanding of NOS (Dunk et al., 2017; Smith, 2010; Southerland et al., 2001). Other studies conducted by my colleagues and I suggest that this could be happening. In one study, I found that the MATE frequently measures not only evolution acceptance, but also other things like understanding of evolution, understanding of NOS, and perception of scientists' views on evolution (Chapter 4 of this dissertation). In another study, my collaborators found that results on the MATE, I-SEA, and GAENE 2.1 differ from each other even when the same students take all three surveys (Barnes et al., 2019). This suggests there may be systematic differences in (1) *how* the surveys conflate evolution acceptance with other constructs, (2) the *extent* to which they conflate acceptance with other constructs, and (3) *which* other constructs each survey measures. This is a cause for concern because such differences can bias results in studies that seek to examine the relationship between evolution acceptance and other potentially related constructs. These are, in short, questions of survey validity.

What is survey validity, and how can gaps in validity evidence influence the measurement of evolution acceptance?

Put simply, survey validity addresses the question of whether a survey truly measures what researchers intend it to measure. In this case, a student's own personal acceptance of evolution is the construct that surveys such as the I-SEA and GAENE aim to measure. Evidence for the validity of a survey is gathered by assessing content validity, internal structure validity, external structure validity, and response process validity (AERA et al., 2014; Artino et al., 2014; Campbell & Nehm, 2013; Mead et al., 2019). Content validity addresses whether the survey presents a complete and accurate representation of the relevant knowledge domain, internal structure validity addresses whether individual items on the survey are all measuring the same thing, and external structure validity addresses whether the survey as a whole statistically correlates with other similar measures (Campbell & Nehm, 2013; Mead et al., 2019). These three forms of validity evidence have been gathered for the I-SEA, GAENE 2.1, and GAENE 3.0 both during their initial development (Glaze et al., 2020; Nadelson & Southerland, 2012; Smith et al., 2016), and in later validation studies (Barnes et al., 2019; Romine et al., 2018; Sbeglia & Nehm, 2018, 2019).

However, the I-SEA and GAENE have not been assessed for response process validity. Response process validity addresses whether prospective participants interpret survey items the way that researchers intended. Process validity is violated when a participant selects answers based on reasons other than what the researchers intended, and can indicate that an item is measuring extraneous information other than the targeted construct (AERA et al., 2014; Artino et al., 2014). Response process validity is assessed through cognitive interviews, in which participants "think aloud" as they reason through why they answered a survey item in a particular way (García, 2011; Willis, 2004). My concurrent study (Chapter 4) on the original MATE demonstrates that cognitive interviews with students can reveal substantial issues related to measuring extraneous constructs that other forms of validation have been unable to clearly detect. In it, I find that even though the MATE was designed to only measure acceptance of evolution, it measures understanding of evolution, understanding of NOS, and students' perceptions of scientist views on evolution in addition to their levels of personal acceptance of evolution. I also find that interpretations of the word "evolution" vary between students whenever items use the term without specifying species or context (Chapter 4). Educational research standards and my own findings with the MATE thus indicate that assessing response process validity is an essential component in gathering validity evidence for the I-SEA and the GAENE if researchers want to continue to use these instruments to measure evolution acceptance.

Current Study

Given that cognitive interviews can identify response process errors that other forms of validation testing are less suited to detect, we contend that process validity testing via cognitive interviews with students is an essential step in validating any measure of evolution acceptance. Thus, the goal of the current study was to explore the process validity of two recent instruments that have been developed to measure evolution acceptance - the I-SEA and the GAENE (2.1 and 3.0) - using student interviews in order to identify what process validity issues -if any- characterize these instruments.

METHODS

The goal of this study was to explore the response process validity of the I-SEA and the GAENE (2.1 and 3.0) using cognitive interviews with undergraduate students. Each interview consisted of a one-on-one “think aloud” interview using one of the survey tools, followed by several open-ended questions about the student’s views on evolution and a brief demographic survey. I conducted three separate rounds of interviews, with different students in each round (see Table 5 for summary). The first round occurred in Fall 2020. All participants were recruited from an upper-level biology course at Arizona State University and took either the I-SEA or the GAENE 2.1. The second round occurred in Spring 2021. Participants were recruited from several public universities across the U.S. and took either the I-SEA or the GAENE 2.1. The third round occurred in Fall 2021. Participants were recruited from an upper-level biology course at one institution and an introductory-level biology course at another institution; these students were interviewed on the new items added to the GAENE 3.0. Below I provide more detailed descriptions of the measures used, recruitment methods, interview protocols, and data analysis.

Table 5. Summary of data collection, including number and timing of student interviews

Instrument	Number of survey questions	Semester collected	Interviews
I-SEA	24	Fall 2020, Spring 2021	22
GAENE 2.1	13	Fall 2020, Spring 2021	17
GAENE 3.0 (new items)	10	Fall 2021	21

Survey Tools

The I-SEA is a 24-item survey that is partitioned into three subscales for macroevolution, microevolution, and human evolution (an "item" is a question or statement on a survey). Each subscale consists of eight forward- and reverse-coded items scored on a 5-point Likert scale ranging from "strongly agree" to strongly disagree" (Nadelson & Southerland, 2012). The subscale categories were not labeled on the student version of the survey.

The GAENE 2.1 is a 13-item survey intended to measure evolution acceptance as a single construct. The GAENE 2.1 is also scored on a 5-point Likert scale; however, all the items are forward-coded, meaning that agreeing with an item indicates acceptance of evolution (Smith et al., 2016). The GAENE 3.0 is a 22-item survey. It contains 12 items from the GAENE 2.1, one item that was removed during the initial development of the GAENE 2.1 but added back in for the GAENE 3.0, and nine entirely new items (Glaze et al., 2020). The first two rounds of interviews used the GAENE 2.1 because the GAENE 3.0 was not yet published when data collection began. The third round of data collection focused exclusively on the items that were on the GAENE 3.0 but not the GAENE 2.1.

Recruitment

In the first round of data collection, I recruited 17 students from an upper-level biology course for majors at a research-intensive public university in the southwest during the Fall 2020 semester. These participants received extra credit worth one daily assignment grade in the course as an incentive for participation. Though religiously diverse, the majority of students recruited from this course exhibited high levels of evolution acceptance and had taken 5 or more college-level biology courses, so they had fairly strong biology and evolution backgrounds

In the second round of data collection, I sought to expand the diversity of my sample by sending individual emails to students who received low scores on other measures of evolution acceptance as part of a separate study. This second set of 22 additional participants came from a nationwide sample of students at public universities. Students recruited in this manner received a $15 Amazon gift card for participation.

I conducted the third round of data collection to assess the response process validity of items that were present on the GAENE 3.0 but not on the GAENE 2.1. During the Fall 2021 semester, I recruited 21 students from two courses: an upper-level biology course at the first-round institution, and an introductory-level biology course at a public R2 university in the southeast. Students who participated were offered either extra credit worth one daily assignment grade or a $10 Amazon gift card.

Cognitive Interviews

I conducted 22 interviews with the I-SEA, 17 with the GAENE 2.1, and 21 with the new items on the GAENE 3.0 (see Table 5). During the Fall 2020 and Spring 2021 rounds of data collection, interviews alternated between the GAENE 2.1 and the I-SEA, such that each instrument was tested in both rounds. During the cognitive interviews, participants read each item from the given instrument out loud, selected an answer out loud, and explained why they selected the answer that they chose as opposed to the other answers available to them. At the end of each interview, the interviewer also asked a set of free-response questions that addressed the student's acceptance of various aspects of evolution, including macroevolution and human evolution. The purpose of these free-response questions was to give participants the opportunity to describe their views on evolution in their own words and potentially clarify any inconsistencies across their interview. I conducted and recorded all interviews virtually via Zoom.

Students were asked to fill out a brief demographic survey after the interview.
Given the qualitative nature of this study, the purpose of collecting demographic information was not to use it for data analysis, but to track the diversity of my sample. The survey contained questions on religiosity and religious affiliation to help me include students with a variety of religious perspectives. To check whether the sample contained students with different levels of evolution education, the survey also asked how many college-level biology courses they had taken and whether any of these courses had been primarily about evolution. This was not intended to be a direct measure of students' knowledge about evolution, but a proxy of their prior exposure.

Data Analysis

To identify any response process errors in how students answer items on the GAENE and the I-SEA, I qualitatively analyzed the cognitive interviews using a combination of deductive and inductive coding (Cho & Lee, 2014; Krippendorff, 2018) To enable direct comparison between evolution acceptance instruments, student responses were initially coded using a deductively developed, relatively broad codebook with codes based on prior critiques of evolution acceptance instruments (Barnes et al., 2019; Nadelson & Southerland, 2012; Romine et al., 2018; Sickel & Friedrichsen, 2013; Smith, 2010) and process errors that I had previously identified on the MATE (Chapter 4). The deductive codebook included codes to be applied whenever a student answered an item based on either (1) their factual understanding of evolution, (2) their understanding of NOS, (3) defining "evolution" in a way that explicitly excludes human or macroevolution when the item itself does NOT specify microevolution alone, (4) their perception of scientists' views on evolution (rather than their own views), or (5) misalignment between an item that assumes Biblical creationism and their own religion's creation account. I included these five codes in the initial codebook because these were the response process errors that I had previously found when conducting cognitive interviews on the original MATE. However, I understood that some of these response process errors may not arise in the present study, and included an "other" code for any novel process errors. This codebook was used as a tool for structuring notes during the interview process.

Because I wanted to identify all potential process errors, I proceeded to inductively code the interview data after the interviews were complete. I developed an inductively derived codebook by listening to each interview recording, assigning a detailed new code whenever a student made a novel process error, and conducting a constant comparison analysis in which each student's process errors were compared to existing codes to determine if an existing code is applicable or if a new code is warranted (Cho & Lee, 2014). During this process, student responses were analyzed and further broken down into inductively derived subcodes of the existing deductive codes, and entirely new codes were developed for novel process errors initially labeled as “other.” For example, I inductively analyzed student responses coded as “understanding of evolution” to identify the subcodes “low certainty about own understanding of evolution” and “misconceptions about evolution” (See **Supplemental Material** for the final full codebook).

After I coded all interviews, a second researcher used the codebook to independently code 10% of the interviews. A comparison of the codes yielded an acceptable level of interrater agreement (Cohen’s kappa = 0.77).

The following results include quotes from students in the study. Names have been changed to protect identity and some quotes have been lightly edited for clarity. The Institutional Review Board of Arizona State University approved the procedures for this study (ASU IRB #00010903).

RESULTS

Participants

I interviewed a total of 60 students for this study. Table 6 displays a summary of the participants in terms of gender, race/ethnicity, religious affiliation, academic year, major, and prior evolution exposure. Prior evolution exposure was classified as “high” for those who have taken a course focused primarily on evolution, “medium” for those who have taken ≥ 3 biology courses but no course on evolution, and “low” for those who had taken ≤ 2 biology courses and no course on evolution. For gender, nonbinary and fill-in-the-blank options were provided but not selected by any of the participants. The "GAENE" column shows a combined sample of students who interviewed on the GAENE 2.1 and students who interviewed on the new items from the GAENE 3.0. Three participants did not fill out the demographic form.

Table 6. Participant demographics, evolution education, and evolution acceptance scores (n=60).

Demographic Variable	GAENE 2.1 + 3.0	I-SEA	Demographic Variable	GAENE 2.1 + 3.0	I-SEA
Gender Identity			**Religious Affiliation**		
Man	26.3%	36.3%	None	31.6%	22.7%
Woman	68.4%	59.1%	Christian	47.4%	50%
No answer	5.3%	4.5%	Hindu	2.6%	9.1%
Race/Ethnicity			Muslim	5.3%	9.1%
Asian/Asian American	15.8%	22.7%	Other	5.3%	4.5%
Black/African American	15.8%	13.6%	No answer	7.9%	4.5%
Hispanic/Latinx	7.9%	18.2%	**Major**		
Native American	2.6%	0%	Biology	65.8%	45.4%
White	42.1%	22.7%	Other STEM	21.1%	40.9%
More than one	10.5%	9.1%	Other non-STEM	7.9%	9.1%
No answer	5.3%	13.6%	No answer	5.3%	4.5%
Academic Year			**Prior Evolution Exposure**		
Lower-level	52.6%	36.3%	High	28.9%	36.3%
Upper-level	39.5%	76.5%	Medium	34.2%	22.7%
No answer	5.3%	4.5%	Low	31.6%	36.3%
			No answer	5.3%	4.5%

I-SEA Finding 1: Students struggle to answer items on the I-SEA when they lack knowledge about evolution.

Students struggling to answer items due to their limited understanding of evolution was the single most common process validity issue on the I-SEA. While this process error arose on at least one item for most (91%) participants, it was largely clustered with particular students and on particular items.

I found that four students – Anemone, Marinus, Rio, and Tethys – struggled with understanding of evolution across the entire I-SEA instrument. Students were categorized as "struggling with understanding of evolution" when they had knowledge-related process errors on six or more items, while all other students had this process error on three items or less. On average, these four students had knowledge-related process errors on over a third (39%) of their answers on the I-SEA. It is worth noting that three of the four had low prior exposure to college-level evolution instruction, while the fourth did not fill out the demographic questionnaire.

Marinus' responses were a good example of how a student's limited understanding of evolution can impact survey validity. He described fully accepting evolution in the open-ended response and received average composite scores of 4.1 (macro), 4.9 (micro), and 4.6 (human). However, Marinus' responses were at times affected by misconceptions about evolution. For example, his answer for *Item 18: Although humans may adapt, humans have not/do not evolve* revealed the misconception that evolution occurs at the level of individuals rather than populations, and that individuals that fail to reproduce play no part in the evolution of a species. This misconception directly influenced his answer choice:

> Marinus (disagree): “I would disagree with this simply because our ancestors were used to living in caves, which was vastly different compared to now. [I didn’t pick strongly disagree] because if the ones that do not evolve eventually die out, then technically those individuals don’t get to evolve and reproduce.”

Another student with many knowledge-related process errors was Tethys, who likewise expressed full acceptance of evolution during the open-ended interview and received average composite scores of 3.9 (macro), 4.0 (micro), and 4.1 (human). They chose “undecided” rather than “strongly agree” for *Item 4: I think all complex organisms evolved from single celled organisms*:

> Tethys (undecided): “I’m [going to] say undecided because I’m not too informed on single-celled organisms to know if [complex organisms] evolved from them or not. **Just [without] knowing I would think so, but then again, I’m not sure**.”

While a majority of the knowledge-related process errors were concentrated in the four students discussed above, other students’ one or two knowledge-related process errors were concentrated in a subset of items. Namely, of the remaining 18 students who were confident in their knowledge about evolution, 4 students (22%) had knowledge-related process errors on Item 14 and 6 students (33%) had these process errors on Item 24.

<u>Item 14</u>*: I think there is an abundance of observable evidence to support the theory describing how variations within a species can happen*.

This item revealed uncertainty about the extent of the “abundance of observable evidence” even for students who were otherwise confident in their knowledge. For example, Azure described accepting microevolution and speciation within closely related taxa and provided answers that were largely consistent with this view (composite score of 4.9 for microevolution). The one exception was Item 14, for which she said:

> Azure (agree): "[I don't strongly agree because] we haven't gotten too far into this topic in my biology class. I feel like I'd need to see a bit more observable evidence to strongly agree. I need a little bit more in depth research about it. I'd need to learn more evidence in my class."

Item 24: *Physical variations in humans (i.e., eye color, skin color) were derived from the same processes that produced variation in other groups of organisms.*

Item 24 revealed low confidence in personal knowledge, misconceptions, and conceptual mix-ups among participants. One example of a mix-up comes from Ariel, who had an average composite score of 3.0 on the human evolution subscale despite saying that she believes God created humans in their present form. Part of the reason for Ariel's unexpectedly high score comes from her misinterpretation of what the item refers to:

> Ariel (strongly agree): "Yes, I'd strongly agree. I'm thinking how eye color, skin color [are] all genetic. So, I'm thinking it's within the DNA and I don't think it has anything to do with evolution per se, because it's not like we're evolving blue eyes. That stuff we inherit from our parents."

Ariel interpreted Item 24 as genetic processes related to inheritance and the central dogma, rather than to evolutionary processes such as natural selection or genetic drift. This indicates that the overall scientific complexity of this item can leave students confused about what "processes" the item refers to, which leads to an answer that reflects the student's views on a different topic.

I-SEA Finding 2: Several items on the I-SEA consistently produce process errors due to unclear wording of the items.

Validity issues with individual items arose when multiple students with varying views and social identities answered items based on difficulties related to item wording, rather than on their own acceptance or rejection of evolution. Below, we describe items that exhibited process validity errors for at least 15% of participants. Given the typically small sample sizes in cognitive interview studies, 15% has been used as a standard cut-off for differentiating between potential validity issues vs. random errors (Nápoles-Springer et al., 2006).

<u>Item 6</u>*: There is little or no observable evidence to support the theory that describes how one species of organism evolves from a different ancestral form*.

In addition to the several students who expressed low certainty about their knowledge of the observable evidence, an additional 4 students (18%) struggled to interpret the phrase "different ancestral form." Two examples of this come from Moana, who said that she fully accepts evolution, and Ocean, who said that she accepts human macroevolution but believes that higher taxa such as mammals and insects do not share a common ancestor.

> Moana (undecided): "This question doesn't make much sense. What do you mean by 'different ancestral form'? Different from what?"
>
> Ocean (undecided): "I don't really understand this question. … Maybe you mean an ancestral form that looks different from [the species in question]? Like a fish and a tiger? Or it could mean like, the connection between humans and apes? I don't really know what is being asked."

<u>Item 7</u>*: The forms and diversity of organisms have changed dramatically over time*, and <u>Item 16</u>*: There is overwhelming evidence supporting the theory of evolution to explain how variations in a species develop over time.*

For Items 7 and 16, 4 students (18%) on each item avoided the "agree" or "strongly agree" options due to these items' use of strong adjectives such as "dramatically" and "overwhelming."

Two examples come from Triton, who said that he's undecided about his views on human evolution and the shared ancestry of all life, and Anemone, who said that she fully accepts evolution.

> Triton (Item 7, undecided): "I'm not sure. Like, are we talking about when the earth started, or a thousand years? Is it asking if an organism has the ability to change dramatically? I'm not too sure what is supposed to be changing dramatically."
>
> Anemone (Item 16, agree): "I would say only 'agree' and not 'strongly agree' just because of the word 'overwhelming.' Like, I don't know, just certain words… I'm not really sure what is meant by the word 'overwhelming' in this case."

Overall, these findings suggest that while process validity issues are neither especially numerous nor especially frequent on the I-SEA, there is nevertheless room for improvement in terms of scientific complexity and overall item clarity.

GAENE Finding 1: The GAENE can overestimate evolution acceptance for students who use an incomplete definition of evolution.

While the I-SEA delineates between micro, macro, and human evolution, most items on the GAENE use the terms "evolution" or "evolutionary theory" without the same specification. This leaves room for survey-takers to potentially use an incomplete definition of evolution that excludes more controversial concepts such as our shared ancestry with primates (human macroevolution) and the shared ancestry of distantly related higher taxa (such as mammals and cephalopods). An examination of student explanations in response to both the individual items and the open-ended questions reveals that 71% of students who described *not* fully accepting evolution used an incomplete definition of evolution at least once. For students who used an incomplete definition of evolution at least once, doing so influenced their answer choices for *approximately one third* (34%) of the items, on average.

One example of this trend comes from Ariadne, a Baha'i student who described accepting speciation but rejecting both human macroevolution and the shared ancestry of higher taxa:

> Ariadne (open-ended interview questions): "I think there were multiple starting species, which had babies and evolved. We definitely didn't start out with one [ancestral species]. I'm thinking there were probably hundreds of starting species. I'm thinking maybe like, the big cats evolved from a similar ancestor and maybe a lot of rodents [share a common ancestor]."

Yet despite these self-described views, Ariadne selected *agree* or *strongly agree* for eight items on the GAENE 2.1, resulting in a composite score of 3.5. We can see an example of how this occurred in Ariadne's reasoning for the item, "Nothing in biology makes sense without evolution," in which she described how excluding most forms of macroevolution from her definition of "evolution" led her to agree with this item:

> Ariadne (agree): "[When the item refers to 'evolution'] I'm thinking of something on the border between micro and macro. I think the majority of it just stems back to natural selection. I guess that's more of a microevolution idea."

Another example comes from Cassandra, a Southern Baptist student who likewise described rejecting both human macroevolution and the shared ancestry of higher taxa. Cassandra selected *strongly agree* for the new GAENE 3.0 item, "All evidence supports the claim that evolution is true," with the following explanation:

> Cassandra (strongly agree): “I do believe that we all have to come from something or someone. [Interviewer: What concepts are you associating with the word “evolution?”] I do believe that **some** species of animals have evolved from other species of animals, and I do believe that humans have evolved **since the Stone Age**. **Species themselves** evolve, but **I don’t believe that we all evolved from one species**.”

The quote above demonstrates that when answering this item, Cassandra strongly agreed because she was using a definition of evolution that includes microevolution and some macroevolution for a subset of species, yet excludes the shared ancestry of life. Yet this is one of the new GAENE items that was designed to differentiate between those with a “high” vs. a “very high” level of acceptance, which suggests that limiting “evolution” to microevolution and limited macroevolution was not the survey authors’ intent.

As these examples show, students who do not fully accept evolution tend to define “evolution” in a way that *includes* the ideas that they agree with and *excludes* the ideas that they disagree with. Doing so results in a score that may be higher than we would expect for students who do not accept human evolution and/or macroevolution.

GAENE Finding 2: Several items on the GAENE consistently produced process errors because students frequently used constructs other than their own acceptance of evolution.

Validity issues with individual items arose when multiple students with varying views and social identities answered particular items based on factors other than their own acceptance or rejection of evolution. These factors can be categorized as emotions and behaviors pertaining to evolution, such as willingness to advocate for evolution and emotional attachment to evolution. We describe how these factors influenced students' answers for particular items below.

Item 1: Everyone should understand evolution.

Half (50%) of the students who said that they fully accept evolution stated that while they personally accept evolution, they do not think that it is essential for evolution to be taught to those who do not wish to learn it. Two students who expressed this view were Helen and Jocasta. Both said that they fully accept evolution and had composite scores of 4.7 and 4.3, respectively, yet both selected *disagree* for this item.

> Helen (Item 1, disagree): "Although I personally think it's an important topic, I do not think it should be required of everyone and do not think that people who do not hold scientific views, but instead hold the religious views, I don't think that they should need to understand evolution."
>
> Jocasta (Item 1, disagree): "In my personal opinion, yeah, I strongly agree. But in a more realistic fashion, [given] that some people are more religious, I'm gonna say disagree because that [would be like] somebody who's majoring in physics telling me that I have to understand physics."

As these quotes demonstrate, a low score on this item does not necessarily indicate that a student personally rejects evolution. Rather, some students are simply less insistent than others when it comes to teaching evolution to reluctant individuals. This finding supports the results from a recent study that also found this item to have poor item fit based on psychometric analyses; the authors of said study hypothesized that, "the misfit of [Item 1] is likely caused by responses from students who accept evolution, but nonetheless do not view it as a necessity for engaging in other courses of study or for advancing one's quality of life" (Romine et al., 2018, pg. 16).

Item 6: *I would be willing to argue in favor of evolutionary theory in a public forum such as a school club, church group, or meeting of public-school parents.*

A similar conflation between evolution acceptance and a person's behaviors and emotions can be seen for Item 6. About 59% of all students (and 75% of those who fully accept evolution) took into account the potential for social stress in the described scenario when selecting their answer, which tended to reduce agreement. This stands in contrast to Item 10 (*I would be willing to argue in favor of evolution in a small group of friends*), for which no student indicated that social stress would impact their answer. Penelope's responses to these two items illustrate the unique features of Item 6:

> Penelope (Item 6, strongly disagree): "The word 'argue' kind of gave me a negative connotation. **I wouldn't necessarily argue in favor of evolutionary theory, but I would slightly debate, [though] not in a violent or harsh way.** It seems a little controversial to talk about in a church or meeting of public-school parents."
>
> Penelope (Item 10, agree): "[This scenario is] a small group of friends so it makes it more realistic. Also, I would argue in favor of evolution in front of my friends because **it seems like a safe space** to talk about evolution and kind of just learn from them, **as opposed to the church group**."

As can be seen from the contrast between Penelope's two answers, her choice to strongly disagree with Item 6 was not based on the extent of her personal acceptance of evolution, or even on her willingness to present an argument in favor of evolution under certain circumstances. Instead, her answer choice reflected the perceived hostility of the public forum scenario. These findings indicate that Item 6 may consistently underestimate students' acceptance of evolution, particularly for students who prefer to avoid confrontation or feel unsafe in certain spaces. These findings likewise support the results from Romine et al. (2018), who found Item 6 to have poor item fit and hypothesized that, “fear of public speaking is quite common, and it is straightforward to argue that one can display acceptance of evolution without extraversion” (Romine et al., 2018, pg. 17).

<u>Item 17</u>: *Evolution is the most important theory devised by man.*

One hundred percent of students who fully accept evolution stated that while the theory of evolution is both important and true, there are other major scientific theories that are equally or more important for science and/or society. In fact, only two students from the entire sample selected *agree* for this item, and not a single person selected *strongly agree*. The responses of Pandora and Eurydice, two agnostic students who fully accept evolution, show why:

> Eurydice (strongly disagree): “I can’t think of any other theories that are extremely important, but personally I don’t think that learning how we came to be is the most important thing we’ve ever done. I’m sure there have been other theories in science, particularly in medicine, that have helped us more.”
>
> Pandora (disagree): “I think that it’s important to understand our origins, but I don’t think that’s the *most* important theory. I would say bigger overarching theories, like the theory of relativity, that help us understand [the universe] as a whole [are more important].”

These quotes show that our participants’ near-universal lack of agreement with this item does not reflect a lack of agreement with any evolutionary concept; rather, it indicates that these students do not think that the unifying theory of biology is inherently more important than theories in other scientific disciplines or more important than other ideas in biology.

Item 18: I would bet my life on the claim that evolution is true.

Just over half (56%) of students who fully accept evolution said that it is simply not in their personality to bet their life on any theory, no matter how well supported. The responses of Callisto and Antigone illustrate this mindset:

> Callisto (undecided): "I'll have to go with undecided because I'm not going to bet my life on anything. But I do believe that evolution is true."
>
> Antigone (disagree): "I'm going to disagree with that one only because I would bet my life on very little. **It's probably more of a me thing than an evolution thing**."

Both Callisto and Antigone explicitly stated that their answer choices do not reflect their views on evolution but instead reflect their risk-averse personalities. The notion of betting their life on *any* idea gave them pause.

Item 19: Understanding evolution has changed my life.

Just under half (44%) of students who reported that they fully accept evolution said that their understanding of it has not changed their life as a whole. This trend took two main forms. One subset of students said that while they recognize the importance of evolution within science, this theory has had little impact on their personal, day-to-day life outside of science. One such student was Callisto, a non-religious student who described fully accepting evolution:

> Callisto (disagree): "I believe that [evolution] is important and true, but it hasn't impacted my life in any type of way."

Another subset of students said that they have never deliberately rejected evolution, so learning about it did not produce a change in their acceptance of evolution – they simply moved from a state of ignorance to a state of knowledge about the topic. One such student was Alcyone, a spiritual student who also fully accepts evolution:

> Alcyone (disagree): “I grew up in a religious household, so evolution wasn’t something that I thought about much. So, when I learned about it [in] school, I just kind of took it for what it was instead of [experiencing] cognitive dissonance.”

GAENE Finding 3: Several items on the GAENE were consistently impacted by students’ knowledge about evolution and the nature of science.

I found that several items on the GAENE consistently test students’ factual knowledge about evolution and NOS, such that their answers may not reflect their general acceptance of the evolutionary concepts in question.

<u>*Item 12*</u>*: Evolution is a scientific fact* and <u>*Item 16*</u>*: Evolution is a fact.*

For students who described fully accepting evolution, 63% of responses on Item 12 and 19% of responses on Item 16 were affected by students’ understanding of NOS, namely the knowledge that scientific theories are categorically different from scientific facts. Evolution includes both fact and theory. The existence of evolutionary change is an observable fact, while the theory of evolution explains the mechanisms that drive this change (Branch & Mead, 2008; Lenski, 2020). However, students frequently either did not know or forgot about evolutionary change as a scientific fact, and focused exclusively on the theory of evolution when answering these items. Examples of this come from Jocasta and Circe:

> Jocasta (Item 12, strongly disagree): “From what I’ve learned so far in every biology class I’ve taken, there is no such thing as a scientific fact.”

> Circe (Item 16, undecided): "Evolution is a theory, so I feel like 'fact' is an inappropriate word for that. So maybe undecided? [Evolution] is factual and supported by evidence, but it's more of a theory because a fact is something that you can test in a lab, and you can't necessarily do that with evolution because it's a long-term process."

As these quotes show, Circe and Jocasta both interpreted these items as being only about the *theory* of evolution, and answered based on the understanding that a scientific theory is, by definition, different from a scientific fact. In particular, Circe's explanation shows that she is aware that evolutionary theory is based on factual evidence, but failed to label the factual evidence itself as "evolution." Yet both of these students described accepting the shared ancestry of all life in the free-response portion of the interview, and did not dismiss evolution as "just" a theory in the colloquial sense when answering these items. Further, Jocasta's response highlighted a potential misconception about there not being any facts in science; although science can change, there are some concepts that are so well established that they become factual (Gregory, 2008).

<u>*Item 14*</u>*: All evidence supports the claim that evolution is true* and <u>*Item 15*</u>*: All species can be traced back to a single ancestor.*

For both Items 14 and 15, just under half (44%) of students who fully accept evolution said that they do not have enough knowledge about evolution to be sure of the scientific accuracy of these statements. Two examples of this trend come from Antigone and Dido, who both described fully accepting evolution despite themselves recognizing that they have some knowledge gaps on the topic:

> Antigone (Item 14, undecided): "I'm not super educated on the in-depth details of evolution. As far as the basics go, I'd say that the concept as a whole is true, but I

would want to do a little more research before I wholeheartedly agree with [this statement].”

Dido (Item 15, undecided): “I’d probably say undecided **just because of my lack of knowledge**. Reading this question without any evolution or genetics background, I would think that there could be multiple ancestors. … I don’t know if they all share one ancestor.”

As we can see here, it is possible for students accept the shared ancestry of life in a broad sense without being factually aware of whether all of life traces back to a single ancestor vs. multiple ancestors. It is likewise possible for students to recognize that evolution is a very well-supported theory without being sure of what “all evidence” encompasses and whether all that evidence directly supports evolution. These knowledge gaps pose a validity issue for these items because students’ answers reflect uncertainty about factual details, rather than uncertainty about the basic concepts of large-scale shared ancestry or the existence of abundant supporting evidence for evolution.

<u>*Item 20*</u>*: Evolution explains how bacteria that are resistant to an antibiotic can arise in a population exposed to that antibiotic,* and <u>*Item 21*</u>*: Evolution explains how careful breeding can produce members of a species that look different from their ancestors.*

For both Items 20 and 21, one third (33%) of all students were uncertain about what scenario the item refers to or how that scenario connects to evolution. Students who struggled with these items include Clytemnestra, an agnostic student who fully accepts evolution, and Hippolyta, a Protestant student who rejects human evolution and believes that non-human species were created at higher taxonomic ranks such as class or kingdom:

Clytemnestra (Item 20, undecided): “I don’t know if it’s the question that I don’t understand, or if it’s tying evolution into bacterial resistance. I would just say the

> question's unclear to me. [It's unclear how] evolution ties into antibiotic resistance."
>
> Hippolyta (Item 21, undecided): "I'm not sure what to put for this one because I'm not quite clear on 'careful breeding.' … when I think of 'careful breeding,' I think about a person breeding poodles for more desirable traits, and I'm not clear on how this connects to species looking different from their ancestors. [But] maybe 'careful breeding' means like in nature, where [colorful] male peacocks are more desirable to females."

These quotes show that about one third of students in our study were confused by Item 20 because they are unacquainted with antibiotic resistance as an example of natural selection, while another third of students were confused by Item 21 because they were uncertain about whether "careful breeding" refers to artificial selection or sexual selection, and/or how artificial selection relates to natural evolution.

DISCUSSION

This study aimed to explore process validity of two evolution acceptance instruments- the ISEA and the GAENE. Overall, I found problems with the validity of these instruments that need to be addressed if we want to continue to use these instruments to accurately measure student acceptance of evolution.

For the I-SEA, I found that the main response process validity issue was students struggling with the survey due to their limited understanding of evolution. There were several students who described fully accepting evolution when asked to explain their views in their own words, yet repeatedly selected more neutral answers primarily because they were uncertain about whether certain items aligned with the scientific consensus, and not because they expressed personal doubt about the truthfulness of familiar scientific information. There were also two items that proved challenging for other students as well because they were either uncertain about what qualifies as “an abundance of observable evidence” (Item 14), or were liable to be confused about what concept the item is referring to (Item 24). In addition, there were several other items that may benefit from the use of fewer superlatives or simpler language. One recent study found that the human evolution subscale is not unidimensional, with Items 17, 20, 23, and 24 clustering separately from Items 18, 19, 21, and 22; the authors suggested that these clusters represent macroevolution and microevolution respectively (Sbeglia & Nehm, 2019). While the item-set that seemingly represents microevolution did elicit more response process errors in which students excluded human macroevolution from their interpretation (9 process errors across 4 “micro” items, vs. 3 process errors across 4 “macro” items), no individual item exceeded the 15% threshold for students exhibiting the same process error on an item.

For the full GAENE 3.0, I found response process validity issues on half of the items. These validity issues involve the measurement of several constructs other than evolution acceptance, including understanding of evolution (Items 14, 15, 20, & 21), understanding of NOS (Items 12 & 16), and several constructs pertaining to participants’ priorities and personality traits, such as the perceived importance of evolution relative to other scientific theories (Item 17) and their willingness to engage in public debate (Item 6). Some of the items we found to have validity issues have been flagged as potentially problematic by other studies (Items 2 & 6 in (Romine et al., 2018) and Items 6 & 12 in (Sbeglia & Nehm, 2018)) while validity issues with other items are identified here for the first time. The lack of specificity in many items' use of the term "evolution" also allows students who do not fully accept evolution to simply exclude the concepts that they reject from their interpretation of "evolution." This same process validity issue was also present on the original MATE (Chapter 4).

The GAENE 2.1 item "some parts of evolutionary theory could be true" was deleted from the GAENE 3.0 due to poor fit in the Rasch model (Glaze et al., 2020). I found that two students described the phrasing of this item as a "trick question" because the statement that only *some* parts of evolution (vs. all or most) merely *could* be true (vs. are probably true) suggests that neither the "agree" nor "disagree" options reflect full acceptance of evolution (likewise for broad rejection of evolution). Meanwhile six students either agreed or strongly agreed with the item based on the stated reasoning that they personally accept some parts of evolution but reject other parts, which makes this item an accurate reflection of their views. This finding supports the decision of Glaze and colleagues to not include this item in the GAENE 3.0.

Limitations

Though I tried to recruit participants with a diverse range of views about evolution, roughly half of the participants who completed the I-SEA and GAENE 2.1, and about three-quarters of the participants who completed the new GAENE 3.0 items expressed full acceptance of evolution when describing their views in their own words. As such, there may be some less-common process errors that are particular to students with a low acceptance of evolution that I failed to detect. Future research with populations with very low evolution acceptance could extend this work and illuminate additional process errors.

Recommendations for Instrument Use and Future Research

The prevalence of process validity issues with the GAENE 3.0 in this study and original MATE in my other study (Chapter 4) illustrate the importance for measures of evolution acceptance to clearly define "evolution" for survey-takers. When used without context, this term can be interpreted to include all aspects of evolution, everything except for human evolution, or microevolutionary processes alone. This allows students who do not fully accept evolution to exclude any concepts they reject from their personal definition of evolution, which causes the instrument to overestimate their acceptance. To potentially circumvent this issue, I recommend that instructors and researchers who wish to use the GAENE 3.0 provide students with the definition of evolution acceptance provided in the original publication, "Evolution acceptance is the mental act or policy of deeming, positing, or postulating that the current theory of evolution is the best current available scientific explanation of the origin of new species from preexisting species" (Smith et al., 2016, pg. 1296). The I-SEA largely avoids this validity issue by describing specific evolutionary concepts (e.g., Item 20: "I think that humans and apes share an ancient ancestor.") instead of referring to "evolution" as a whole. The main drawback of describing specific evolutionary concepts is that items can easily become prone to scientific misinterpretation by students (e.g., confusing natural selection for trait heritability) or require the use of specific knowledge about evolutionary processes or history. While this appears to be less of a concern for students with prior exposure to college-level evolution instruction, I find that it can cause issues of conflation between understanding and acceptance for students who may not have learned about evolution recently. As such, I recommend that instructors and researchers consider the general level of evolutionary knowledge of their study population when deciding whether the I-SEA is the best instrument for their purposes.

This study further demonstrates that establishing process validity evidence via cognitive interviews is an essential step in developing a reliable survey instrument. While asking several students to review an instrument for clarity is a step in that direction, issues with measuring constructs other than the intended construct are liable to slip through review unless systematic interviews are conducted with a diverse population of students. Establishing process validity is all the more important when adding items that are intended to be very easy or very difficult for survey-takers to agree with, as these items may be particularly at risk for being easy or difficult for reasons other than a student's level of evolution acceptance. Furthermore, I advise researchers who seek to further improve the measurement of student evolution acceptance to closely examine the alignment between survey items and an instrument's measurement goals. This includes addressing questions such as, "What is our definition of evolution acceptance?" "Does each item align with this definition of evolution acceptance?" and "What views on evolution do we want the minimum and maximum scores on our instrument to reflect?" Only then will we be able to more accurately measure student acceptance of evolution.

CHAPTER 4

INTRODUCING THE MATE 2.0: A REVISED MEASURE OF ACCEPTANCE OF THE THEORY OF EVOLUTION

INTRODUCTION

Researchers have conducted hundreds of studies over the past thirty years to document low levels of evolution acceptance among students and the public, determine what causes low acceptance, and identify what can be done to increase evolution acceptance. However, researchers have used many different surveys to measure evolution acceptance (Gallup, 2019; Glaze et al., 2020; Nadelson & Southerland, 2012; Pew, 2013; Short & Hawley, 2012; Smith et al., 2016), which makes it difficult to compare findings from studies that can disagree with one another. Additionally, researchers have increasingly recognized that the most common instrument used to measure evolution acceptance, the Measure of Acceptance of the Theory of Evolution (MATE) (Rutledge & Warden, 1999), may have limitations that could be causing confusion about how to increase evolution acceptance (Barnes et al., 2019; Glaze & Goldston, 2015; Lloyd-Strovas & Bernal, 2012; Sickel & Friedrichsen, 2013). However, after 20 years of being the most used survey tool for measuring evolution acceptance, researchers have not yet published an updated and improved version of the MATE based on researchers' critiques. The goals of this study were to articulate current weaknesses of the MATE, revise the MATE based on these identified weaknesses, and then test the new instrument using a population of undergraduate biology students so that we could present a revised instrument with validity evidence.

Acceptance of Evolution Survey Tools

Prior to the publication of peer-reviewed evolution acceptance survey tools, evolution education researchers used many different unique survey tools to measure acceptance of evolution. These unique instruments were typically constructed for use in a single study and led to a lack of consistency in measurement across studies. For example, in one survey of Wisconsin biology teachers, researchers measured evolution acceptance using a unique 14-item instrument measured on a 5-point Likert scale (Koevering & Stiehl, 1989). Meanwhile, a different survey of Ohio high school biology teachers gauged evolution acceptance using two yes-or-no questions that simply asked participants whether they accept evolution and whether scientists accept evolution (Zimmerman, 1987). Such differences in item wording, number of items, and range of answer choices for each item hindered researchers' ability to compare findings across studies. This, in turn, may have been a barrier for evolution education researchers in developing a consistent literature base in which new studies build on prior work.

A major step forward in evolution education research occurred in 1999 with the publication of the Measure of Acceptance of the Theory of Evolution (MATE). The MATE consists of 20 items with which a respondent is asked to agree or disagree on a 5-point Likert scale. It was the first measure of evolution acceptance that had substantial validation evidence (Rutledge & Warden, 1999). The MATE remained the only measure of evolution acceptance with such validation evidence for the next 12 years. During this time, use of the MATE grew, and instruments from sociological public polls such as the

Gallup and the Pew also made an appearance in the evolution education literature. Other measures of evolution acceptance with validation evidence have been introduced within the past decade, namely the Inventory of Student Evolution Acceptance (I-SEA: Nadelson & Southerland, 2012), the Generalized Acceptance of Evolution Evaluation (GAENE: Smith et al., 2016), and a recent revised version of the GAENE (Glaze et al., 2020). Nevertheless, the proportion of evolution education studies that use the MATE has only continued to grow. The MATE is currently the most popular instrument to measure evolution acceptance in college level evolution education studies broadly (Mead et al., 2019) and the most used evolution acceptance instrument in international evolution education studies (Kuschmierz, Meneganzin, et al., 2020). Figure 3 illustrates these trends.

While the development of the MATE was an essential first step in standardizing the measurement of evolution acceptance, the authors of the MATE never intended for this to be the final version of the measure. Rutledge and Warden wanted future researchers to update and strengthen the MATE (Romine et al., 2016; Rutledge & Warden, 1999). Further, multiple researchers in the field have voiced concerns about limitations of this instrument (Barnes et al., 2019; Nadelson & Southerland, 2012; Romine et al., 2018; Sickel & Friedrichsen, 2013; Smith, 2009; Wagler & Wagler, 2013). Many of these critiques highlight ways in which the MATE may conflate acceptance of evolution with other related constructs, such as understanding of evolution, understanding of the nature of science, and perceptions of scientists' views on evolution. Further, Rutledge and Warden (1999) developed the MATE for high school biology teachers, a group with a significant background in the science of biology and its central tenets. Many of the criticisms of the MATE may be a consequence of its usage with populations for which it was not initially developed and thus may need to be revised for use among populations of undergraduate biology students.

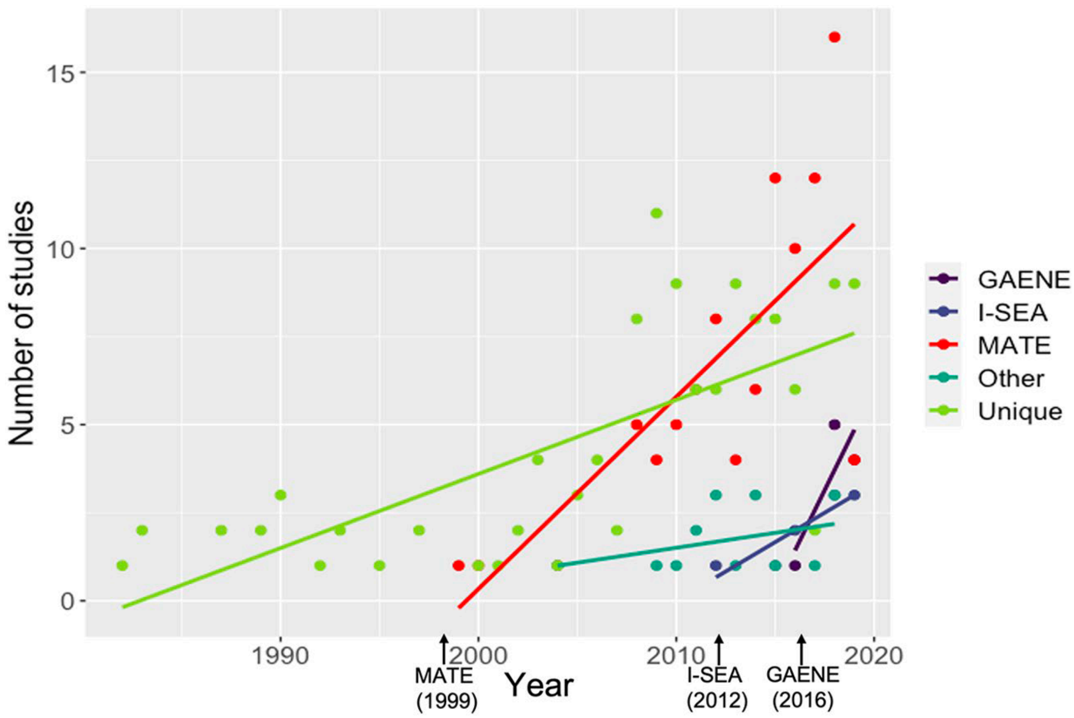


Figure 3. The use of instruments for measuring acceptance of evolution in peer-reviewed studies. Marked time points are the publications of the MATE (Rutledge and Warden, 1999), the I-SEA (Nadelson and Southerland, 2012), and the GAENE (Smith *et al.*, 2016). Measures labeled "other" include the Evolutionary Attitudes and Literacy Survey (EALS) and sociological polls such as the Gallup and the Pew.

Researchers have often questioned whether the MATE measures student conceptions that are not their actual acceptance of evolution. One concern has been that the MATE conflates understanding of the nature of science (NOS) with acceptance of evolution. For example, the MATE item "Evolutionary theory generates testable predictions with respect to the characteristics of life" may measure the respondent's understanding of what constitutes a testable scientific prediction, in addition to – or instead of – their acceptance of the idea that evolution occurs (Smith, 2009). A second concern has been that the MATE conflates understanding of evolution with acceptance of evolution. For example, the MATE items “The age of the earth is at least 4 billion years” and “The age of the earth is less than 20,000 years” appear to measure a respondent’s factual knowledge about the age of the earth in addition to their acceptance of the idea that evolution has occurred over a long period of time on an old earth (Smith et al., 2016). A third concern has been that the MATE conflates acceptance of evolution with a respondent’s perception of scientists’ views on evolution. For example, the item “Most scientists accept evolution to be a scientifically valid theory” may prompt students to answer about what they think the current scientific consensus about evolution is, rather than about their own personal acceptance of evolution (Sickel & Friedrichsen, 2013; Rissler et al., 2014). Finally the term “evolution” is not clearly defined in the survey tool, and items do not specify whether they refer to microevolution, macroevolution, or human evolution (Nadelson & Southerland, 2012). For example, the MATE item “The theory of evolution is incapable of being scientifically tested” requires the survey respondent to picture their own definition of “the theory of evolution,” which may or may not include macroevolutionary concepts such as the shared ancestry of all life on earth or human evolution, which are known to be particularly contentious aspects of evolution for students (Barnes, Dunlop, et al., 2020; Barnes et al., 2019; Barnes, Supriya, et al., 2020).

A recent study from my colleagues revealed that using different surveys to measure evolution acceptance with the same population of students can lead to different research results (e.g., the I-SEA (Nadelson & Southerland, 2012), the GAENE (Smith et al., 2016), the MATE (Rutledge & Sadler, 2007; Rutledge & Warden, 1999)) and others) (Barnes et al., 2019a). While prior studies have taken quantitative approaches to examine the dimensionality of the MATE (Romine et al., 2016, 2018; Wagler & Wagler, 2013) and how results from the MATE compare to other evolution acceptance instruments (Barnes et al., 2019; Metzger et al., 2018), no studies thus far have examined the validity and accuracy of the MATE through the actual voices of students who are taking the survey. In this study, I explored potential weaknesses of the MATE through student cognitive interviews to illuminate response process errors (AERA 2014; Mead et al., 2019).

Prior validity evidence and missing validity evidence for the current MATE

The MATE has validity evidence, but some studies indicate that the MATE could be improved, and some forms of validity evidence are missing. When the MATE was first published in 1999, the authors assessed the content validity of the MATE by presenting the items to a panel of experts and including only items that the panel agreed contributed to the construct of acceptance of evolution. To establish the construct validity of the MATE the authors used Principle Components Analysis (PCA) to illustrate that the MATE was a single factor and that each item on the MATE contributed significantly to the assessment of the one factor (Rutledge & Warden, 1999). The authors also showed that the measure was internally consistent (Rutledge & Warden, 1999) and then later showed evidence of test-retest reliability among university students (Rutledge & Sadler, 2007). In future studies, researchers found evidence that the MATE might be better analyzed as a multidimensional instrument (Romine et al., 2016, 2018; Wagler & Wagler, 2013), but that treating the MATE as a bidimensional measure did not add insight into analyses on acceptance of evolution (Metzger et al., 2018). Further, researchers have shown that students' scores from the MATE are correlated with scores from other measures of evolution acceptance (Barnes et al., 2019; Metzger et al., 2018; Romine et al., 2018; Sbeglia & Nehm, 2018), indicating evidence for concurrent validity of the MATE with other measures of evolution acceptance (AERA, 2014; Mead et al., 2019).

One source of validity evidence that is currently missing for the MATE is process validity evidence. Process validity is violated when a participant responds to an item for reasons other than intended by the researchers and can indicate that an item is measuring extraneous information other than the targeted construct (AERA, 2014; Mead et al., 2019). We refer to specific violations of response process validity as "response process errors." Response process validity is often established through cognitive interviews with participants in which they "think aloud" as they reason why they answered in a particular way to a survey item (Fonteyn et al., 1993; García, 2011; Willis, 2004). Students taking the survey may be the most direct source of information as to whether they are answering based on their acceptance of evolution or something else, but this form of validity evidence for the MATE is currently lacking.

The current study and definition of evolution acceptance

One aim for this study was to identify what process validity issues exist with the current MATE based on cognitive interviews with students. Prior quantitative analyses of MATE scores have already provided evidence that the MATE may be multidimensional (Romine et al., 2016, 2018), but qualitative cognitive interviews can uncover validity issues with the MATE that quantitative analyses leave undetected. Based on prior critiques, I expected that students would describe answering certain items on the MATE using reasoning that is not strictly based on their own acceptance of evolution (e.g., using extraneous constructs like their understanding of evolution). I designed the interviews to be semi-structured so that I could also uncover potentially novel ways in which students answer questions on the MATE.

The second aim of the study was to update the MATE based on any weaknesses discovered in cognitive interviews with students and prior published critiques. A common criticism of the MATE is that the original authors did not provide an adequate operational definition of acceptance of evolution (Romine et al., 2016; Smith et al., 2016). So, when revising items on the MATE based on the cognitive interviews, I also believed it important to provide an adequate definition of acceptance of evolution that aligns with these items. The original authors of the MATE used the terminology "acceptance of evolution" to distinguish between scientific and unscientific ways of adopting information and warned against the use of describing acceptance as "believing in evolution." According to the original authors, to say that one "believes in evolution" implies a similar underlying process for adopting scientific information as religious information (for example, "I believe in God"), thus it was important that the definition not include the word belief to avoid this misunderstanding. Since the publication of the MATE, researchers have extensively discussed the nuances of meaning between the words "accepting," "understanding," "believing," and "knowing" evolution and I took into account these discussions while constructing the definition presented below (Smith, 2009a, 2009b; Smith & Siegel, 2004, 2016). Also, the original authors of the MATE did not specify whether their definition of evolution was that of microevolution, macroevolution, or human evolution acceptance. These ideas have since been shown to be separate constructs with regard to how students assess their plausibility (Nadelson & Southerland, 2012), so I incorporated these critiques when constructing my definition of evolution acceptance. My collaborators and I iteratively reviewed and revised working definitions of evolution acceptance for the MATE 2.0 until we agreed on the following definition:

> ***Acceptance of Evolution***: The agreement that it is scientifically valid that all species have evolved from prior species.

We chose to focus on macroevolution (which includes human macroevolution) for the definition of acceptance of evolution since microevolution acceptance is relatively high among students and thus may not be the most impactful target for evolution acceptance studies in the future (Barnes, Dunlop, et al., 2020; Barnes et al., 2019; Barnes, Supriya, et al., 2020). This definition includes the term "species;" though multiple species concepts exist, we intend to use the Biological Species Concept given its utility in discussing sexually reproducing, multicellular organisms and its widespread use in the biology community (Gao & Rieseberg, 2020, Wu et al., 2020).

METHODS

Cognitive Interviews with original MATE survey

The first step of the study was to explore the process validity of the original MATE so that items could be revised based on any weaknesses found. I conducted 62 cognitive interviews with students across different religious affiliations, levels of acceptance, and levels of knowledge about evolution. To acquire this diversity in participants, I recruited from an upper-level biology course for biology majors, four introductory-level STEM courses, and four introductory-level non-STEM courses. All students were recruited from the same large, research-intensive public university in the southwest in the Fall 2019 and Spring 2020 semesters (henceforth the "primary institution"). Study participants received either extra credit worth one daily assignment grade in the course or a $10 cash payment to incentivize them to participate.

During the cognitive interviews, the participants read each item from the original MATE out loud, chose an answer, and explained why they selected the answer that they chose as opposed to the other answers available to them. At the end of each interview, the interviewer asked a set of free-response questions that addressed the student's acceptance of various aspects of evolution including macroevolution and human evolution (See ***Supplemental Materials*** for questions). The purpose of these free-response questions was to give students the opportunity to describe their views on evolution in their own words and potentially clarify any inconsistencies across their interview.

Students were asked to fill out a brief demographic survey at the end of the interview (See ***Supplemental Materials*** for a copy of the survey). Given the qualitative nature of this study, the purpose of collecting demographic information was not to use it for data analysis, but to track the diversity of our sample. The survey contained questions on religiosity and religious affiliation to help us include students with a variety of religious perspectives. To check whether the sample contained students with different levels of evolution education, the survey also asked in which college courses, if any, the student had learned about evolution. This was not intended to be a direct measure of students' knowledge about evolution, but a measure of their prior exposure.

To find any process errors in how students respond to MATE items, I qualitatively analyzed the cognitive interviews using a combination of deductive and inductive coding (Cho & Lee, 2014; Krippendorff, 2018). Student responses were initially coded using a deductively developed, relatively broad codebook with codes I expected to emerge from the data based on prior critiques of the MATE. Based on prior published critiques of the MATE (Nadelson & Southerland, 2012; Rissler et al., 2014; Sickel & Friedrichsen, 2013; Smith, 2009a, 2009b; Smith et al., 2016), I began the interviews with a code book that included codes to be applied if students answered questions based on their understanding of the nature of science (NOS), understanding of evolution, varying definitions of evolution (microevolution, macroevolution, human evolution), their perceptions of scientists' views on evolution, or whether the item assumed the student was Christian.

To establish the general interview protocol, Elizabeth Barnes and I conducted the first three interviews together. After each interview, we compared our individual notes and came to consensus on the appropriate use of the deductively derived codes. I then conducted the next three interviews alone, while Barnes listened to the interview recordings after. Again, we each took notes and came to agreement on how to apply the deductively derived codes. I then conducted all the remaining interviews.

Because I also wanted to identify any weakness in the original MATE that were not hypothesized before data collection, after the interviews were complete, I proceeded with inductive coding to analyze the interview data. I developed a more detailed codebook by listening to each interview recording, assigning a new code whenever a student made novel use of extraneous reasoning, and conducting a constant comparison analysis in which each student's use of extraneous reasoning is compared to existing codes to determine if an existing code is applicable or if a new code is warranted (Cho & Lee, 2014). During this process, the deductively derived codes from the initial interview codebook were broken down into inductively derived subcodes. For example, the deductively derived code "misconceptions about the nature of science" was divided into inductively derived subcodes such as "misconceptions about what counts as scientific testing" and "misconceptions about the term 'theory' in science" (See ***Supplemental Materials*** for the full codebook).

To determine interrater reliability, I coded all interviews with the codebook and Barnes used the codebook to independently code 10% of the interviews, which did not include any of the initial six interviews which she initially helped conduct. A comparison of the codes assigned by us yielded an acceptable level of interrater agreement (Cohen's kappa = 0.81).

Creating the MATE 2.0

Based on the findings from the cognitive interviews, the prior literature on evolution acceptance measurement, and our new definition of evolution acceptance, my colleagues and I revised items on the MATE to create the MATE 2.0. We removed items that consistently measured extraneous constructs other than evolution acceptance and could not be meaningfully reworded for improvement. For example, "The age of the Earth is at least 4 billion years" consistently measured knowledge of the age of the Earth in addition to evolution acceptance according to the students taking the survey, so we removed this item. In the cases in which a mis-performing item could be improved, we reworded the item to consider the critiques of the item by participants or by prior literature. For example, the item "Current evolutionary theory is the result of sound scientific research and methodology" was reworded as "The idea that new species evolve from earlier species is the result of scientific research" because students (1) used inconsistent definitions of evolution, with some using microevolution while others used macroevolution and/or human evolution (code: definition of evolution), and (2) were unaware of what counted as "sound scientific research and methodology" (code: NOS understanding). We also added a new item, "All of life on earth evolved from previous species," because acceptance of the shared ancestry of all life was not included in the original MATE items, and yet shared ancestry is a foundational assumption of evolutionary theory (Dobzhansky, 1973) that is often rejected by those who do not accept evolution. We opted to retain a mixture of forward and reverse item types to maintain structural consistency across iterations of the MATE. Revised items were phrased to retain their original coding type, and deleted items were not evenly split between forward vs. reverse coding. See Supplemental Table 5 for the full list of how and why each item on the MATE was deleted or revised.

Using the revised items from the MATE, I conducted cognitive interviews with 29 undergraduate students to confirm we had sufficiently improved the process validity of the items. I interviewed students across different religious affiliations, levels of acceptance, and levels of knowledge about evolution. I recruited the first eight students from an upper-level biology course for majors. While religiously diverse, the majority of students recruited in this way exhibited relatively high levels of evolution acceptance and high prior exposure to evolution. To include more students with a lower acceptance of evolution, I sent individual emails to recruit an additional 10 students who received low scores on measures of evolution acceptance in our pilot of the MATE 2.0 and in a separate research study. To include more students with less prior exposure to evolution, I also recruited 17 students from an introductory biology course for non-majors and from a summer program for incoming biology first-years. Students from the upper-level biology course and the summer program were incentivized with the equivalent of one assignment in extra-credit points. The rest were offered a $15–$25 Amazon gift card (gift card incentives rose over the course of the year per standard participant increases). The interview process and data analysis methods for the MATE 2.0 remained largely identical to the methods for assessing the original MATE. The main difference was that the initial interviews were conducted in person, while the MATE 2.0 interviews were conducted remotely via Zoom.

To explore the structural and concurrent validity of the MATE 2.0, I administered the new MATE 2.0 as well as another published measure of evolution acceptance, the I-SEA (Nadelson and Southerland, 2012), to 2881 students in 22 introductory biology classes across seven U.S. states (AZ, FL, MI, NC, TX, AL, MN). The I-SEA includes three constructs of evolution acceptance: acceptance of microevolution, macroevolution, and human evolution. Concurrent validity evidence is gathered when one measure correlates significantly with another measure aimed at the same construct (AERA, 2014). I expected the new MATE 2.0 to have higher bivariate correlations with measures of macro and human evolution and a lower correlation with the measure of microevolution acceptance because our definition of evolution acceptance is focused on macroevolution (which includes human evolution).

To provide structural validity evidence for the MATE 2.0, a colleague and I performed a dimensionality analysis using Rasch modeling to confirm that the MATE 2.0 is a single dimension. We fit a polytomous partial credit model (*irtmodel= PCM* in the R package *TAM* (Robitzsch et al., 2018)) and conducted a Principal Component Analysis (PCA) of the residuals of this model. Low eigenvalue of the first contrast, i.e., a value less than 2, indicates that the residuals are small and without structure and therefore, data fits a unidimensional model (Boone, 2016; Sbeglia & Nehm, 2018).

The following results include quotes from students in the study; names have been changed to protect identity and some quotes have been lightly edited for clarity. The Institutional Review Board of Arizona State University approved the procedures for this study (ASU IRB #00010903). I present the results and discussion together to emphasize how this work builds on prior research.

RESULTS AND DISCUSSION

Cognitive Interviews with the original MATE

To find ways to improve the MATE among a diverse sample of college students, I interviewed a total of 62 students for this portion of the study. Table 7 outlines the diversity of the sample in terms of student religious affiliation, gender, race/ethnicity, academic year, and evolution education exposure. I report average composite scores because they reflect an individual's average answer choice on the Likert scale, i.e. a 4.0 out of 5 indicates that a participant on average "agreed" with evolution. The average total score on the MATE was 87.0 (± 11.1) and the average composite score was 4.4 (± 0.6).

I classified students' exposure to college-level evolution instruction as high, medium, low, or none; as intended, the sample was fairly evenly distributed across these categories (Table 7). The 'high' category consisted of students who had taken an upper-level Evolution course and contained 20 (32%) students. The 'medium' category consisted of 16 (26%) students who had learned about evolution as part of introductory-level and upper-level biology courses, but had never taken an upper-level evolution-specific course. The ‘low’ category included 18 (29%) students who had learned about evolution as part of a single introductory-level biology or non-biology course. The ‘none’ category consisted of eight (13%) students who had never learned about evolution in a college course.

Table 7. Characteristics participants of cognitive interviews with the original MATE. Percentages may not add up to 100% due to rounding.

Demographic	N (%)	Demographic	N (%)
Gender Identity		**Religious Affiliation**	
Man	19 (31%)	Non-religious (atheist, agnostic, nothing)	32 (52%)
Woman	43 (69%)	Buddhist	2 (3%)
Race/Ethnicity		Christian	19 (31%)
Asian/Asian American	15 (24%)	Hindu	3 (5%)
Black/African American	3 (5%)	Jewish	1 (2%)
Hispanic/Latinx	10 (16%)	Muslim	1 (2%)
Middle Eastern	1 (2%)	Other religion	3 (5%)
Native American	1 (2%)	Did not answer question	1 (2%)
White	28 (45%)	**Interview-Based Acceptance**	
Multi-racial	4 (6%)	Full acceptance	47 (75%)
Evolution Education Exposure		Human exception	3 (5%)
High	20 (32%)	Creation of higher taxa	4 (6%)
Medium	16 (26%)	Rejection	5 (8%)
Low	18 (29%)	**Academic Year**	
None	8 (13%)	Lower-level (First-year, Sophomore)	27 (44%)
		Upper-level (Junior, Senior)	35 (56%)

Below are the findings from the cognitive interviews with the original MATE. I first present response errors present across items on the MATE (Findings 1 and 2), and then present response process errors specific to particular items on the MATE (Finding 3).

Finding 1: The MATE can overestimate evolution acceptance for students who use an incomplete definition of the theory of evolution.

An examination of the item-level responses showed that students used a definition of evolution that is either limited to microevolution or excludes humans. Most commonly, this consisted of defining "evolution" as "microevolution," which does not include macroevolution or human evolution. That is to say, some students' definition of evolution included only evolutionary *processes* such as natural selection, not the macroevolutionary outcomes of these processes. The use of this definition led students to answer items in a manner that reflects their acceptance of microevolution, but not macroevolution. Previous studies have found that acceptance of microevolution is generally high, even in populations that exhibit significantly lower acceptance of macroevolution and human evolution (Barnes et al. 2019; Sbeglia & Nehm, 2019).

For example, Rowan was a Catholic participant who said that he believes that God created most species in more or less their present form. Yet his average score on items on the MATE was a 3.6 out of 5, indicating that his average answer was between Undecided and Agree in favor of evolution. The cause of this higher-than-expected score is demonstrated in his reasoning for *Item 1: Organisms existing today are the result of evolutionary processes that have occurred over millions of years*, with which he agreed.

> Rowan (agree): "I think to some extent there has been evolution. I'm deeply religious, so I believe that organisms were created [by God]. But I do believe that they've adapted to better suit the change in their environment over time since they were created. I'm not a firm believer in everything stemming from single-celled organisms. But I would say the animals on the different Galapagos Islands, and how different they are from island to island [is an example of evolutionary change]. "

As this quote demonstrates, Rowan agreed with Item 1 because he accepts that evolutionary processes can produce variation that leads to visibly different populations or closely related species, yet he did not perceive that he has to accept that the evolution of all life from single-celled organisms is an essential component of evolution. He displayed this pattern of reasoning across a dozen items in the 20-item survey.

Similarly, Iris was a Protestant student who described her views as rejecting much of macroevolution and the common ancestry of life. Nevertheless, her average score on MATE items was 4.4 out of 5, indicating that her average answer was between Agree and Strongly Agree in favor of evolution. This is in part because Iris also defined "evolution" as "microevolution." This can be seen in her reasoning for *Item 3: Modern humans are the product of evolutionary processes that have occurred over millions of years*, with which she agreed.

> Iris (agree): "I know people say that humans have come from apes, and I don't think that is necessarily true. So, I do think that humans have evolved, but not necessarily from another species."

In a related trend, some students included non-human macroevolution in their definition of evolution but did not apply the theory to humans. For example, Ginger was a Protestant student who described accepting all of evolution except for the evolution of humans. Though her composite score of 3.8 reflected these views, her answers for several items were influenced by whether evolution was assumed to apply to humans. In her answer for Item 1, Ginger said "I choose 'agree' but not 'strongly agree' because [I don't know if] 'organisms' also includes humans or not." This was an answer-selection process that accurately reflected her views. In contrast, she selected "strongly disagree" for *Item 2: The theory of evolution is incapable of being scientifically tested,* using the following reasoning:

> Ginger (strongly disagree): "I think there has been lots of science that has tested it."
> Interviewer: "Given your previous answer, were you thinking of human evolution when you answered this question?"
> Ginger: "No"

Together, these responses support the validity concerns raised by Nadelson and Southerland (2012) in that they demonstrate the risk of using survey items that use the term "evolution" without specifying micro- or macroevolution, or without explicitly stating whether evolution is being applied to humans. Students who accept some aspects of evolutionary theory, but not others, are likely to include only the aspects with which they agree in their definition of "evolution." Doing so can cause such students to receive an overly high MATE score that suggests that their acceptance of evolution is greater than it actually is.

Finding 2: The MATE can underestimate evolution acceptance for students who have misconceptions about the nature of science.

In addition to overestimating the evolution acceptance of some students, the MATE can underestimate the acceptance levels of others. Fifteen students who described accepting that life largely arose from a common ancestor received lower than their expected MATE scores and had item answers that were influenced by nature of science (NOS) misconceptions. The presence of these misconceptions typically resulted in students selecting answers that indicate a lower acceptance of evolution than their actual view.

One example of this comes from Sage, an atheist student who described accepting all of evolution and had an average score of 4.4. Sage's explanations for several of her answers revealed misconceptions about what qualifies as scientific testing; namely, she perceived that scientific testing requires the scientist to directly observe a natural event as it is happening. This can be seen in her reasoning for *Item 4: The theory of evolution is based on speculation and not valid scientific observation and testing*, with which she disagreed, but not strongly disagreed.

> Sage (disagree): "I don't think we have scientific observation and testing, but we have evidence from the past. ... The evidence that I'm thinking of is, like, the human skeletons that were dug up. I guess that counts as observation. But I wouldn't say it's testing since you can't really test the theory of evolution on something in the past because no one was there to watch it."

Another example comes from Dale, a Catholic student who had an average composite score of 4.4 and said he fully accepted evolution and believed it to be a mechanism of God's creation; this view is typically referred to as theistic evolution (Yasri & Mancy, 2016). His explanations revealed a misconception about the difference between a scientific theory and a scientific fact. This can be seen in his response to *Item 10: Evolution is not a scientifically valid theory*, with which he disagreed, but not strongly disagreed.

> Dale (disagree): "I feel like it is pretty scientifically valid. They do have evidence to prove that evolution has occurred. [I don't strongly disagree] only because, like, since it is a theory, by the definition of a theory, it technically hasn't been proven true yet."

As these quotes illustrate, having one or more misconceptions about the NOS can lead students to select answers that indicate a partial acceptance of evolution even when their self-described views are fully consistent with the scientific consensus on evolution. Not only does this trend have the capacity to artificially reduce students' MATE scores, but it also poses validity issues for studies that use the MATE to examine the relationship between evolution acceptance and understanding of NOS. Given that NOS misconceptions can influence students' MATE scores, use of this measure will likely inflate any correlations between these two constructs. Multiple studies have found greater understanding of NOS to be positively correlated with acceptance of evolution, as measured by the MATE (Dunk et al., 2017; Rutledge & Mitchell, 2002; Rutledge & Warden, 2000); the current interview findings suggest that the strength of these correlations may be inaccurately high due to construct conflation on the MATE.

Finding 3: Specific items on the MATE consistently produce process errors, which result from the use of extraneous constructs and unclear wording of the items.

Validity issues with individual items arose when multiple students with varying views and social identities answered items based on factors other than their own acceptance or rejection of evolution. This trend contained two main sub-trends: (1) items which appeared to elicit the use of extraneous constructs and other reasoning unrelated to evolution acceptance and (2) items with unclear wording that students struggled to interpret. Below I describe individual items from the MATE that appear to measure constructs other than a student's personal acceptance of evolution.

MATE Item 2: The theory of evolution is incapable of being scientifically tested. Approximately 20% of students cited an inaccurate understanding of what counts as scientific testing when answering this item. Students selected either "disagree" or "undecided" rather than "strongly disagree" even when they expressed full acceptance of evolution. Two examples of this trend come from Lilac and Sage, both of whom had said that they fully accept evolution.

> Lilac (undecided): "You can build phylogenies and analyze how things are related to each other... but there's no set of experiments you could run to test this theory."
>
> Sage (disagree): "I'd say disagree. But it would have to be one of those studies that goes over, like, several lifetimes. So, I think it's capable of being scientifically tested; I just think we haven't actually done it yet."

MATE Items 5 & 17: "Most scientists accept evolutionary theory to be a scientifically valid theory" and "Much of the scientific community doubts if evolution occurs." More than 80% of students answered each of these two parallel items based on their impression of the extent to which evolution is accepted among scientists, which is consistent with concerns previously voiced by other researchers (Sickel & Friedrichsen, 2013; Rissler et al., 2014). This is a problem because these students either did not reference their personal views on evolution in explaining their answer or went so far as to explicitly point out how their answer to this item is not a reflection of their own views.

In one pattern that I identified, students with a self-described high acceptance of evolution claimed that while many scientists do accept evolution, some scientists *do not* accept it. This pattern arose in approximately one third (Item 17) and two thirds (Item 5) of students who claim to fully accept evolution, and typically resulted in responses that underestimate students' own level of acceptance. This pattern can be seen in Ivy's response to Item 5, and Petunia's response to Item 17.

> Ivy (Item 5, agree): "**I know where I stand, but I don't know where everybody else stands.** ... I'm not religious, but a lot of people who are religious kind of dismiss evolution. Most scientists probably do agree with it, but I think that scientists who don't agree with it would be those who are super religious."

> Petunia (Item 17, disagree): "I would say 'disagree,' because for this question, I feel like I would need to see a poll or actual statistics for how many people. Because it's not really an opinion thing. ... **I would LIKE to say strongly disagree, but then again, I feel like there probably are some scientists in the community that do doubt it."**

Interestingly, I also found the opposite pattern in students with lower levels of evolution acceptance. Forty percent of students for Item 5 and 60% of students for Item 17 who described some rejection of evolution emphasized the broad acceptance of evolution within the scientific community. The most striking example for both items comes from Herb, a Protestant student who expressed Biblically literal, Young Earth Creationist views in which species were created separately from one another by God within the last 20,000 years. His answer choices for items 5 and 17 imply a high level of evolution acceptance; yet based on Herb's answers to all of the other items on the survey, he would have received the lowest possible score on the MATE.

> Herb (Item 5, strongly agree): "From what I've read online, 90% of scientists agree with this, or something like that."

> Herb (Item 17, strongly disagree): "I would strongly disagree with that. The scientific community does not doubt evolution; they accept it. That's based on what I've seen online and in the news."

Together, these two patterns suggest that for a large portion of students, Items 5 and 17 operate in the opposite way from how they were intended. Instead of claiming that most scientists share their own views, as the items assume, many students instead emphasize the existence of scientists who do *not* share their own views; this tendency is found across the spectrum of evolution acceptance.

MATE Items 6 & 8: "The available data are ambiguous (unclear) as to whether evolution actually occurs" and "There is a significant body of data that supports evolutionary theory."

Items 6 and 8 both ask about whether there are data to support the theory of evolution. For both items, students (15% each) stated that they were not sufficiently familiar with the data to strongly agree or strongly disagree with these items. Yet out of the 13 students who cited an insufficient familiarity with the data on one or both of these items, 11 of them fully accepted evolution based on their self-described views. This trend can be seen in the responses of Daisy and Azalea:

> Daisy (Item 6, disagree): "I feel like I don't have enough knowledge to tell if [the item] is actually true. … But I feel like from what I know, [the theory of evolution] can be validated."

> Azalea (Item 8, agree): "I would say agree. This goes back to number 6, where it's like, **I would say 'strongly agree' if I knew the exact amount of data that supports evolutionary theory.** But with everything that I've been taught, I feel like there are data that supports evolutionary theory."

As these quotes demonstrate, this trend does not reflect actual uncertainty about whether evidence for evolution exists and is known by scientists. Instead, it appears to reflect students’ perception that they personally are not very familiar with the supporting evidence. This is a problem because these items do not evaluate students’ acceptance of evolution when interpreted in this way. A student’s confidence in their knowledge about evolutionary data would likely increase as they learn more about evolution, even if their level of acceptance remains the same, so this item could present a problem for comparing understanding of evolution with acceptance of evolution. For instance, in a pre-post instruction study design, a researcher may conclude that they have increased acceptance of evolution by increasing understanding of evolution when in reality they have only increased evolution understanding and not acceptance.

MATE Item 11: The age of the earth is at least 4 billion years.

As previous researchers have argued (Smith et al., 2016), Item 11 assesses not only whether a student accepts the idea that the earth is very old, but also whether the student is factually aware that it is more than 4 billion years old. In fact, approximately half of all participants stated that they know the earth to be far older than 20,000 years, but that they do not know whether it is more than 4 billion years old. This prompted students to avoid selecting “strongly agree” despite fully accepting the general idea that the earth is ancient. For example, Azalea and Savannah said the following:

> Azalea (agree): "I definitely know that it's more than a million. I definitely know that it's more than... see, that's what I mean. Four billion? I just don't know the exact number."

> Savannah (disagree): "I have no idea. Four billion seems like a lot. Yeah, I'd say it's less than 4 billion. Maybe it’s like 1 billion years [old]."

These responses demonstrate that students who do not “strongly agree” with Item 11 are not necessarily Young Earth Creationists who believe species were created in their current form within the last 20,000 years. This item instead reflects the fact that many students are simply unaware that the earth is about 4.54 billion years old, despite accepting that it is millions or billions of years old.

MATE Item 13: Evolutionary theory generates testable predictions with respect to the characteristics of life.

Two main patterns arose in students’ responses to Item 13: answers based on misconceptions about NOS, and confusion about the wording of the item. The NOS misconceptions referenced for this item all pertained to what counts as scientific testing; approximately one third of the participants revealed misconceptions about scientific testing when they explained their answer choice. One common misconception was that the only way to test evolutionary hypotheses is through live observation of the event or process in question. An example of this comes from Rosemary, who agreed – but not strongly agreed – with Item 13.

> Rosemary (agree): "Not every prediction is testable. I guess, like, how the first parts of evolution came about [are not testable]. There's no way to go back millions of years and test if that was true or not."

Rosemary is an agnostic student who later stated that she accepts that all plants and animals evolved from single-celled ancestors, but is skeptical about whether humans evolved from primates and thinks that birds, mammals, and reptiles evolved from unrelated single-celled organisms. In light of these self-described views, her answer for Item 13 appears to primarily reflect a limited understanding of how scientists construct and test hypotheses about early evolutionary history.

A similar misconception that repeatedly arose was that scientific predictions are limited to predictions of future events, and do not include predicted observations about past events. This largely took the form of students interpreting “testable predictions” to mean predictions about how current species will evolve in the future, which would make scientific testing an impractically slow process. This misunderstanding can be seen in the response process of Oliver, whose self-described views are consistent with a full acceptance of evolution.

> Oliver (agree): "Evolution is something that's really hard to predict because it does take years for something to evolve and adapt. So, we can continue to generate those hypotheses, but we would need several centuries to even prove those evolutionary theories."

In addition to the students whose answers were impacted by misconceptions about scientific testing, another one third of students struggled to select an answer simply because they were confused by the phrase “characteristics of life.” They did not know what the phrase meant, and thus had difficulty interpreting the item. The responses of Lily and Marigold demonstrate this pattern:

> Lily (agree): "Agree? The wording on this one is a little funny. 'The characteristics of life?' I think that evolutionary theory does generate testable predictions. Maybe it's that last part that's a little odd."
>
> Marigold (undecided): “I don’t even know what that means. The end of it, ‘with respect to the characteristics of life,’ I just like, don’t understand.”

MATE Item 14: The theory of evolution cannot be correct since it disagrees with the Biblical account of creation.

While Item 14 appeared to operate largely as intended for Christian and non-religious students, approximately half of the non-Christian religious students struggled with this item. We initially hypothesized that the phrase “Biblical account of creation” may be unsuitable for students who follow a religion other than Christianity. To address this hypothesis, the interviewer asked every student if they would answer the item differently if it read “my religion’s account of creation” instead of “the Biblical account of creation.” Five out of nine students of non-Christian faiths said yes. For instance, Basil, a Jain (ancient Indian religion) student who selected “strongly disagree” but said that he would switch to “undecided” if the item was not specific to Christianity.

> Basil (strongly disagree): "I would answer differently. I would probably put 'undecided.' I definitely do believe that my religion [played a role in the origins of life].”

While many religious students in the United States are Christian, the interviews reveal the validity issues that can arise when an item explicitly excludes the creation stories of other religions. When a survey is intended to measure the evolution acceptance of students of any or no religion, items that are specific to Christianity can systematically bias the scores of students who follow religions other than Christianity (which on average is about 13% of introductory biology students (Barnes et al., 2021).

MATE Item 15: Humans exist today in essentially the same form in which they always have.

For Item 15, I found that participants' answers were influenced by how they interpreted the word "humans." A number of students (15%) interpreted "humans" to mean only our current species, *Homo sapiens*, and not any of the earlier species from which we evolved. The fact that this is a reverse-scored item, however, makes clear that "humans" was intended to include both modern humans and all of our hominin ancestors. This posed a problem for students who generally accept human macroevolution, because defining "humans" as "*Homo sapiens*" makes Item 15 a scientifically true statement – humans have existed in essentially the same form as long as they have been deemed modern humans. This trend is apparent in the response of Forrest, who described fully accepting evolution, including our shared ancestry with other primates.

> Forrest (undecided): "I'm just wondering what the scope of 'human' is. Like, are we talking about *Homo sapiens*, or like…? Maybe it's referring to how humans have existed between now and a few thousand years ago, **or whenever we started to become human**."

As these quotes demonstrate, defining "humans" as *Homo sapiens* alone can lead students to select an answer choice that is inconsistent with their actual views.

MATE Item 18: The theory of evolution brings meaning to the diverse characteristics and behaviors observed in living forms.

Many students struggled with the wording of Item 18. Approximately one third of the students were uncertain about how to interpret the phrase "brings meaning." Students pointed out that "brings meaning" can be interpreted in multiple ways, and that their answers would depend on which interpretation they choose to use.

> Liana (agree): "I think 'meaning' can have multiple meanings. I think [the theory of evolution] does help to explain why our physical characteristics are [the way

> they are]. But I think in terms of 'meaning' as in a more existential meaning, I think that kind of depends on the person."

As this student observed, in this context the term “meaning” can be interpreted as “scientific explanation” or as “philosophical or spiritual purpose.”

MATE Item 19: With few exceptions, organisms on earth came into existence at about the same time.

Item 19 was designed to represent a Young Earth Creationist view on the origins of life, with agreement indicating that the student believes that current species were divinely created in more or less their present form over a brief timespan. Yet 18% of students interpreted this item to have the entirely opposite meaning. These students interpreted Item 19 as saying all of the species we see today descended from one common ancestor, which was alive at a single point in time. To disagree with this statement would be to say that present-day species evolved from many different “first” ancestors, which were not related to each other, and which lived at different points in time. Students who accept the shared ancestry of all life and used the opposite interpretation of Item 19 selected answers on the agreement side of the scale, while a correct interpretation of the item would have led them to select answers of “disagree” or “strongly disagree.” One example of this opposite interpretation comes from Briar, who strongly agreed with this item despite appearing to fully accept evolution.

> Briar (strongly agree): "Does this mean to say that organisms started at once, or that humans and dinosaurs existed simultaneously? It seems obvious to me that the answer would have to be strongly agree, because the first organism is at the same time as the first organism.”

The interview responses demonstrate that students who interpret Item 19 to have the opposite meaning as intended also provide answers that are opposite to their views.

Revising the MATE to create the MATE 2.0

My colleagues and I revised the MATE based on the issues revealed in the cognitive interviews as well as the prior critiques of the MATE from researchers. We removed items from the MATE that students indicated did not measure their evolution acceptance. We revised items that partially measured evolution acceptance to remove references that caused errors in the students' response processes. We also made sure that each new item was in line with the definition of acceptance of evolution chosen for this measure. We added a prompt to the survey to clarify the definition of a species, which included humans. This initial revised version of the MATE consisted of 11 items but after further cognitive interviews and Rasch dimensionality analyses we removed two items due to response process errors and marginal acceptable fit statistics. The final version of the MATE 2.0 consists of nine items (Table 8).

Table 8. The MATE 2.0*

Prompt: A species is a group of similar organisms. For example, dogs, cats, and humans are all different species. Given this definition of a species, please indicate whether you agree or disagree with the following statements, based on your personal opinion.	
1	All species that exist today have evolved from previous species.
2	Modern humans have evolved from earlier non-human species.
3	**The idea that new species evolve from earlier species is NOT supported by scientific evidence.**
4	Current scientific evidence suggests that new species can evolve from earlier species.
5	**The idea that new species evolve from earlier species is NOT a scientifically valid theory.**
6	The idea that new species evolve from earlier species is the result of scientific research.

7	The idea that species can evolve into new species explains the diversity of life on Earth.
8	The idea that new species evolve from earlier species is a scientifically valid theory.
9	All of life on earth evolved from previous species.
10	Organisms exist today in largely the same form in which they always have. *DELETED
11	Humans exist today in largely the same form in which they always have. *DELETED

*Items are answered on a scale of: 1) strongly disagree, 2) somewhat disagree, 3) neutral, 4) somewhat agree, and 5) strongly agree. Bolded items should be reverse-coded using a scale from (1) strongly agree to (5) strongly disagree. The final draft of the MATE 2.0 consists of items 1–9; items 10 and 11 were deleted during the validation process.

Validity Evidence for the MATE 2.0

Process validity, structural validity, and concurrent validity evidence for the MATE 2.0 are reported below.

Finding 4: Cognitive interviews with students provide process validity evidence for the MATE 2.0.

To see if we had sufficiently revised items on the MATE to resolve process errors, I conducted cognitive interviews with a total of 29 students using the new items from the MATE 2.0. Of these students, five identified as non-religious, 18 as Christian (including Catholic, Protestant, LDS, and other denominations), two as Hindu, and one each as Buddhist, Muslim, and spiritual. For race/ethnicity, four identified as Asian/Asian American, six as Black, four as Hispanic/Latinx, one as Native American, nine as white, three as multiracial, and one declined to state. Seven participants identified as men and 21 as women. One student declined to provide any demographic information.

In response to the open-ended questions at the end of the interview, 15 students described themselves as fully accepting evolution (including the shared ancestry of all life and humans' shared ancestry with other primates) and 14 described views that involve rejecting at least one major aspect of evolutionary theory. Of the latter group, all 14 questioned or rejected human macroevolution, nine stated that non-human species evolved following divine creation at intermediate taxonomic ranks, with examples ranging from classes such as mammals to families such as felids and canids, and two denied the existence of any macroevolution beyond limited speciation within a genus. For students who fully accepted evolution, the average total score on the MATE 2.0 was 41.4 (± 3.1) out of 45, and the average composite score was 4.6 (± 0.3) out of 5. For students who accepted some but not all macroevolution, the average total score was 32.9 (± 7.4) and the average composite score was 3.7 (± 0.8). And for students who denied all macroevolution, the average total score was 22.5 (± 0.7) and the average composite score was 2.5 (± 0.08).

Cognitive interviews on the MATE 2.0 occurred in two rounds. The initial set of interviews for Draft 1 of the MATE 2.0 occurred in the winter of 2020/2021 and included 12 students recruited from an upper-level biology course at the primary institution and from a nation-wide set of students who had previously participated in related research. Items 10 and 11 arose as potentially problematic items during the first round. For Item 10, "Species exist today in essentially the same form in which they always have," four students (33%) selected Undecided or Disagree despite fully accepting evolution because they interpreted "essentially the same form" to include basic biological features shared across the tree of life. For Item 11, “Humans exist today in essentially the same form in which they always have,” six students (50%) who claimed to accept human macroevolution selected Agree or Disagree (but not Strongly Disagree) because they were either comparing present-day humans to “cave men” such as Neanderthals or were unclear as to what comparison they should be making. Conversely, three students who claimed to reject human macroevolution did not select Agree or Strongly Agree for Item 11 because they thought that the item was referring to microevolutionary and/or developmentally plastic changes in present-day humans vs. prehistoric *Homo sapiens*. I flagged these items for potential deletion but kept them in the pilot to further assess their performance in the quantitative analyses.

The second round of cognitive interviews for Draft 2 of the MATE 2.0 occurred in the summer of 2021 and included 17 students recruited from two introductory-level biology courses at the primary institution (ASU) and as well as from a nation-wide sample of students who had participated in the quantitative piloting of this survey. Draft 2 was created by removing Items 10 and 11 following the Rasch analyses. Together, the two rounds of interviews demonstrate that the remaining nine items of the MATE 2.0 produce far fewer process errors than the original MATE instrument. The revision process had consisted of deleting items that consistently produced process errors and revising the remaining items to specify macroevolution and deemphasize factual knowledge about evolution and NOS.

Though a few students' limited understanding of NOS did influence their answers for Items 5 and 8, "The idea that new species evolve from earlier species is/is not a scientifically valid theory," this occurred for only 3 - 4 students (10 - 14%) out of 29. This was the most common process error on any individual item; all others occurred three times (10%) or less. The other occurrences of a 10% process error rate on a MATE 2.0 item were low confidence in one's own knowledge for Item 2, a misunderstanding of the tentative nature of science for Item 3, and a tendency to interpret "explains the diversity of life" as "explains *some* of the diversity of life" for Item 7. To contrast, Item 19 on the original MATE was the least-problematic item to be deleted solely due to item-specific process errors, which arose for 18% of participants. Similarly, an average of five out of 20 answers were influenced by NOS misconceptions for students who struggled NOS on the original MATE, yet students who had an NOS-related process error on either Item 5 or 8 of the MATE 2.0 received this code for only one to two items in total.

Additionally, the cognitive interviews revealed that most items primarily measure acceptance of macroevolution rather than acceptance of the shared ancestry of all life. This was apparent primarily for students who expressed the belief that non-human species evolved following divine creation at intermediate taxonomic ranks. For example, Jelena had a mean composite score of 4.3 and selected Strongly Agree for items such as 7, "The idea that species can evolve into new species explains the diversity of life on Earth." But in describing her own views on macroevolution, Jelena stated that lions and tigers do share a common ancestor, birds and fish might share a common ancestor, but mammals and fish do not. These responses indicate that Item 7 and similarly phrased items are capturing her views on macroevolution (speciation) but not necessarily on the shared ancestry of higher taxa at the rank of phylum or above. The exception to this trend is Item 9, “All of life on earth evolved from previous species.” Students who did not fully accept evolution had an average composite score of 3.5 across the entire survey, but only 2.6 on Item 9. I advise researchers to keep this in mind when interpreting score results from the MATE 2.0.

Finding 5: Rasch analyses of responses to the MATE 2.0 provide structural validity evidence

The eigenvalue of the first contrast was 1.05 for the unidimensional model suggesting that the unidimensional model is a good fit to the data. Weighted mean squares item fit statistics (WMNSQ, equal to infit MNSQ) for the Rasch models ranged from 0.81-1.37 which is largely within the acceptable range (i.e., 0.7-1.3 logits). However, Items 5 and 8 fell slightly outside of the range for acceptable fit statistics, which was unsurprising given the process errors reported by students in the cognitive interviews. Reliability measures for the model were greater than the acceptable cutoff of 0.7. Expected a posteriori/plausible value reliability index (EAP/PV), a measure of item reliability, was 0.91. Person reliabilities as estimated by WLE person separation index, which estimates if a similar order of person abilities would be generated by items of similar difficulty was 0.88. Since Items 5 and 8 were marginally outside of acceptable fit statistics and these items also showed some response process errors during cognitive interviews, we decided to remove these two items from the final version of the instrument. The eigenvalue of the first contrast for the Rasch model without Items 5 and 8 was 0.87 and weighted mean squares item fit statistics were all within the acceptable range of 0.7-1.3 logits. EAP/PV reliability index was 0.91 and WLE person separation index was 0.87. See ***Supplemental Figures 1 and 2*** for the Wright maps and ***Supplemental Tables 3 and 4*** for fit statistics for the Rasch model.

Finding 6: Correlations of the MATE 2.0 with other evolution acceptance measures provide concurrent validity evidence

Using the sample of 2881 students, I found evidence for concurrent validity of the MATE 2.0. Bivariate correlations between MATE 2.0 scores and the macroevolution acceptance and human evolution acceptance scales of the Inventory of Student Evolution Acceptance (I-SEA) (Nadelson & Southerland, 2012) were high (macroevolution $r = .81$, $p < .001$; human evolution $r = .82$, $p < .001$). These high correlations show that the new MATE 2.0 has concurrent validity with the I-SEA macro and human evolution acceptance scales. The correlation between MATE 2.0 scores and the microevolution acceptance scale of the I-SEA was a moderate correlation and lower than with the macroevolution and human acceptance scales of the I-SEA ($r = .67$, $p < .001$). This lower correlation provides evidence that we created items that were in line with our definition of evolution acceptance, which included macroevolution of humans and non-humans and not microevolution.

Other Considerations

Scoring of the MATE 2.0

Researchers can score the new MATE 2.0 in a variety of ways depending on the use of the instrument. The MATE 2.0 uses a 5-point Likert scale ranging from Strongly Agree to Strongly Disagree. Though some research suggests that removing a neutral option preserves variability in the data (Bishop, 1987; Johns, 2005), I did not remove the neutral option from the MATE 2.0 because the interviews revealed no apparent issues with students' use of the neutral option. The original MATE instrument was scored by aggregating items and assigning a somewhat arbitrary cut off for low, medium, high, and very high scores (Rutledge & Sadler, 2007). To make the scores on the MATE 2.0 less arbitrary, researchers can calculate a student's average composite score across all items, which will indicate that student's average agreement rating with the nine items on the scale (i.e., an average score of "4" across items would indicate a participant, on average, "agreed" with each item on a scale from 1 (strongly disagree) – 5 (strongly agree)). For instance, among our population of students, the average Likert agreement across items was 3.99, which indicates this population on average, was between neutral and agree on their acceptance of evolution as determined by the MATE 2.0. Furthermore, using average composite scores allows for easy direct comparison with other measures of evolution acceptance that use a 5-point Likert scale but contain different numbers of items.

Some researchers have argued for using analyses for Likert scale data through the lens of Rasch modeling in which the different "difficulty" of each item to agree with is taken into account when creating scores (Boone, 2016). Rasch analyses also account for differences in psychological distances between any two adjacent responses on the Likert scale. This is important because the psychological distance between "agree" to "strongly agree" might be smaller than that between "neutral" to "agree" (Boone, 2016). Lastly, Rasch models yield equal interval logit scale measures, which are more suitable for parametric analyses such as regression analyses (Barnes, Supriya, et al., 2020; Boone, 2016; Sbeglia & Nehm, 2019). For these reasons, researchers can covert MATE scores using Rasch analysis to "ability" scores and use those scores for input in analyses. However, an evolution instructor who wants to measure their evolution acceptance of students in their course will likely not want to use Rasch and instead can use what is previously described.

Naming the MATE 2.0

The creation of the MATE 2.0 involved making significant changes to the original survey, which brings up the question of whether to retain the "MATE" name or to create an entirely new name for the revised survey. I opted to retain the name "MATE 2.0" because unlike other studies that addressed concerns about the validity of the MATE by creating entirely new measures (I-SEA: Nadelson & Southerland, 2012; GAENE: Smith et al., 2016), I addressed these concerns by identifying specific response process errors for each survey item, and then either deleting or rephrasing each item with the express purpose of addressing the validity issues that had been found. The name "MATE 2.0" is thus meant to reflect how the revised survey was developed directly from the original MATE.

Survey Content

The MATE 1.0 and the MATE 2.0 both seek to measure acceptance of evolution as a single construct (Rutledge & Warden, 1999). However, the interviews conducted in this study indicate that student interpretations are inconsistent within items on the MATE 1.0, wherein one student may interpret an item as being about microevolution while another may interpret the same item as being about macroevolution. Given that my colleagues and I define evolution acceptance as "the agreement that it is scientifically valid that all species have evolved from prior species," I revised the MATE 2.0 to measure acceptance of macroevolution consistently across students.

CONCLUSIONS

In this study I explored the process validity of the Measure of Acceptance of the Theory of Evolution (MATE) and created a new updated MATE 2.0. I found that the original MATE can overestimate or underestimate students' evolution acceptance. Students reported answering questions based on (1) their understanding of evolution, (2) their understanding of the nature of science, (3) their perceptions of scientists' views of evolution, (4) varying definitions of evolution including microevolution, macroevolution, and human evolution, and (5) confusing wording of items. I revised the original MATE based on the interviews and prior published critiques to create the "MATE 2.0" and provided new process validity evidence, structural validity evidence, and concurrent validity evidence for the new measure. Considering that the original MATE is the most used instrument in evolution acceptance literature, I hope that researchers will instead use this modified instrument to negate some of the limitations of the original MATE.

CHAPTER 5

CONCLUSION

The first study in this dissertation, "The *EvMed Assessment*: A test for measuring student understanding of core concepts in evolutionary medicine," addresses the growing demand for pedagogical resources specific to EvMed by developing an assessment that measures students' ability to apply the core principles of EvMed to various health-related scenarios. The assessment consists of 11 question sets containing a short description of a health-related scenario followed by several likely/unlikely items. I evaluated the EvMed Assessment for validity and reliability using expert reviews (content validity), student interviews (response process validity/substantive validity), Cronbach's α (reliability), and Classical Test Theory (item difficulty and item discrimination). EvMed instructors can use this assessment as a pre/post measure of student learning in an EvMed course to inform curriculum revision, or as a test bank to draw upon when developing in-class assignments or exams.

The second study in this dissertation, "*It's More Of A Me Thing Than An Evolution Thing*: Exploring The Validity Of Evolution Acceptance Measures Using Student Interviews," examined the response process validity of the Inventory of Student Evolution Acceptance (I-SEA) and the Generalized Acceptance of Evolution Evaluation (GAENE). The I-SEA and GAENE were developed by other researchers to improve on the limitations of earlier measures of evolution acceptance, yet neither survey had been tested for response process validity, which can assess the extent to which students use constructs other than their acceptance of evolution to answer survey items. I filled this gap in the validity evidence for the I-SEA and GAENE by conducting cognitive interviews with undergraduate students; this gap needed to be filled because it is important for these surveys to measure evolution acceptance accurately and in isolation from other constructs so that researchers can accurately determine what factors are associated with low evolution acceptance. From the interviews, I found that the I-SEA measures understanding of evolution for about 18% of students, while the GAENE elicits inconsistent interpretations of the term "evolution" and at times measures students' emotional attachment to and willingness to advocate for evolution. Researchers can use these findings to better inform their choice of survey when designing future studies, and to further improve the measurement of evolution acceptance.

The third study in this dissertation, "Introducing the MATE 2.0: A Revised Measure of Acceptance of the Theory of Evolution," examined the response process validity of the Measure of Acceptance of the Theory of Evolution (MATE) and addressed the validity issues found on the original MATE by developing a revised MATE 2.0 survey. I chose to explore the validity of the MATE because while the I-SEA and GAENE present more recent efforts to develop a measure of evolution acceptance, the MATE has remained by far the most widely-used measure of evolution acceptance even after the publication of the other two surveys. To assess the response process validity of the MATE, I conducted cognitive interviews with 62 undergraduate students. These interviews revealed that students answer items on the MATE based on constructs other than their acceptance of evolution, which lead to answer choices that do not fully align with students' self-described evolution acceptance. To address the issue of conflation between evolution acceptance and other constructs, my collaborators and I revised the original MATE into the MATE 2.0 by the deleting survey items that present the most severe conflation, and revising items that present mild to moderate conflation. To assess the validity of the MATE 2.0, I conducted 29 cognitive interviews with undergraduate students, administered the revised survey to 2,881 students in 22 classes, and assessed the survey's structural validity through a Rasch dimensionality analysis and its concurrent validity through correlations with the I-SEA. Researchers and instructors can use the MATE 2.0 when they need a relatively short measure of evolution acceptance that exhibits minimal conflation with other potentially related constructs and is appropriate for religiously diverse and low-evolution-knowledge undergraduate populations.

REFERENCES

American Association for the Advancement of Science. (2011). *Vision and Change in Undergraduate Biology Education*. https://visionandchange.org/finalreport/

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). The standards for educational and psychological testing. *American Educational Research Association*.

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching*, *39*(10), 952–978. https://doi.org/10.1002/tea.10053

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectvies*. Longman.

Antolin, M. F., Jenkins, K. P., Bergstrom, C. T., Crespi, B. J., De, S., Hancock, A., Hanley, K. A., Meagher, T. R., Moreno-Estrada, A., Nesse, R. M., Omenn, G. S., & Stearns, S. C. (2012). Evolution and medicine in undergraduate education: A prescription for all biology students. *Evolution; International Journal of Organic Evolution*, *66*(6), 1991–2006. https://doi.org/10.1111/j.1558-5646.2011.01552.x

Artino, A. R., La Rochelle, J. S., Dezee, K. J., & Gehlbach, H. (2014). Developing questionnaires for educational research: AMEE Guide No. 87. *Medical Teacher*, *36*(6), 463–474. https://doi.org/10.3109/0142159X.2014.889814

Ashraf, Z. A., & Jaseem, K. (2020). Classical and Modern Methods in Item Analysis of Test Tools. *International Journal of Research and Review*, *7*(5).

Athanasiou, K., & Papadopoulou, P. (2012). Conceptual Ecology of the Evolution Acceptance among Greek Education Students: Knowledge, religious practices and social influences. *International Journal of Science Education*, *34*(6), 903–924. https://doi.org/10.1080/09500693.2011.586072

Barnes, M. E., Dunlop, H. M., Holt, E. A., Zheng, Y., & Brownell, S. E. (2019). Different evolution acceptance instruments lead to different research findings. *Evolution: Education and Outreach*, *12*(1), 4. https://doi.org/10.1186/s12052-019-0096-z

Barnes, M. E., Dunlop, H. M., Sinatra, G. M., Hendrix, T. M., Zheng, Y., & Brownell, S. E. (2020). "Accepting Evolution Means You Can't Believe in God": Atheistic Perceptions of Evolution among College Biology Students. *CBE—Life Sciences Education*, *19*(2), ar21. https://doi.org/10.1187/cbe.19-05-0106

Barnes, M. E., Misheva, T., Supriya, K., Rutledge, M., & Brownell, S. E. (2022). A Revised Measure of Acceptance of the Theory of Evolution: Introducing the MATE 2.0. *CBE—Life Sciences Education*, *21*(1), ar10. https://doi.org/10.1187/cbe.21-05-0127

Barnes, M. E., Roberts, J. A., Maas, S. A., & Brownell, S. E. (2021). Muslim undergraduate biology students' evolution acceptance in the United States. *PLOS ONE*, *16*(8), e0255588. https://doi.org/10.1371/journal.pone.0255588

Barnes, M. E., Supriya, K., Dunlop, H. M., Hendrix, T. M., Sinatra, G. M., & Brownell, S. E. (2020). Relationships between the Religious Backgrounds and Evolution Acceptance of Black and Hispanic Biology Students. *CBE—Life Sciences Education*, *19*(4), ar59. https://doi.org/10.1187/cbe.19-10-0197

Basile, A. J., Schwartz, D. B., Rigdon, J., & Stapell, H. (2018). Status of evolutionary medicine within the field of nutrition and dietetics: A survey of professionals and students. *Evolution, Medicine, and Public Health*, *2018*(1), 201–210. https://doi.org/10.1093/emph/eoy022

Berkman, M. B., & Plutzer, E. (2011). Defeating Creationism in the Courtroom, But Not in the Classroom. *Science*, *331*(6016), 404–405. https://doi.org/10.1126/science.1198902

Bichi, A. A. (2016). *CLASSICAL TEST THEORY: An Introduction to Linear Modeling Approach to Test and Item Analysis*. *02*(09).

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, *32*(3), 347–364. https://doi.org/10.1007/BF00138871

Bishop, G. F. (1987). Experiments with the Middle Response Alternative in Survey Questions. *Public Opinion Quarterly*, *51*(2), 220–232. https://doi.org/10.1086/269030

Boone, W. J. (2016). Rasch Analysis for Instrument Development: Why, When, and How? *CBE—Life Sciences Education*, *15*(4), rm4. https://doi.org/10.1187/cbe.16-04-0148

Branch, G., & Mead, L. S. (2008). "Theory" in Theory and Practice. *Evolution: Education and Outreach*, *1*(3), Article 3. https://doi.org/10.1007/s12052-008-0056-5

Branchaw, J. L., Pape-Lindstrom, P. A., Tanner, K. D., Bissonnette, S. A., Cary, T. L., Couch, B. A., Crowe, A. J., Knight, J. K., Semsar, K., Smith, J. I., Smith, M. K., Summers, M. M., Wienhold, C. J., Wright, C. D., & Brownell, S. E. (2020).

Resources for Teaching and Assessing the Vision and Change Biology Core Concepts. *CBE Life Sciences Education*, *19*(2), es1. https://doi.org/10.1187/cbe.19-11-0243

Brassil, C. E., & Couch, B. A. (2019). Multiple-true-false questions reveal more thoroughly the complexity of student thinking than multiple-choice questions: A Bayesian item response model comparison. *International Journal of STEM Education*, *6*(1), 16. https://doi.org/10.1186/s40594-019-0169-0

Brem, S. K., Ranney, M., & Schindel, J. (2003). Perceived consequences of evolution: College students perceive negative personal and social impact in evolutionary theory. *Science Education*, *87*(2), 181–206. https://doi.org/10.1002/sce.10105

Brownell, S. E., Freeman, S., Wenderoth, M. P., & Crowe, A. J. (2014). BioCore Guide: A Tool for Interpreting the Core Concepts of Vision and Change for Biology Majors. *CBE—Life Sciences Education*, *13*(2), 200–211. https://doi.org/10.1187/cbe.13-12-0233

Brüne, M., & Schiefenhövel, W. (Eds.). (2019). *The Oxford handbook of evolutionary medicine*. Oxford University Press.

Bull, J. J., & Read, A. F. (2019). Evolution, medicine and public health: Embracing the future. *Evolution, Medicine, and Public Health*, *2019*(1), 199–200. https://doi.org/10.1093/emph/eoz029

Campbell, C. E., & Nehm, R. H. (2013). A Critical Analysis of Assessment Quality in Genomics and Bioinformatics Education Research. *CBE—Life Sciences Education*, *12*(3), 530–541. https://doi.org/10.1187/cbe.12-06-0073

Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). Overview of Classical Test Theory and Item Response Theory for Quantitative Assessment of Items in Developing Patient-Reported Outcome Measures. *Clinical Therapeutics*, *36*(5), 648–662. https://doi.org/10.1016/j.clinthera.2014.04.006

Cavallo, A. M. L., White, K. J., & McCall, D. (2011). The Mismatch Among Students' Views About Nature of Science, Acceptance of Evolution, and Evolutionary Science Understandings. *Science Education Review*, *10*(2), 6.

Cho, J., & Lee, E.-H. (2014). Reducing Confusion about Grounded Theory and Qualitative Content Analysis: Similarities and Differences. *The Qualitative Report*, *19*(32), 1–20. https://doi.org/10.46743/2160-3715/2014.1028

Cofré, H., Cuevas, E., & Becerra, B. (2017). The relationship between biology teachers' understanding of the nature of science and the understanding and acceptance of the theory of evolution. *International Journal of Science Education*, *39*(16), 2243–2260. https://doi.org/10.1080/09500693.2017.1373410

Coleman, J., Stears, M., & Dempster, E. (2015). Student teachers' understanding and acceptance of evolution and the nature of science. *South African Journal of Education*, *35*(2), Article 2. https://doi.org/10.4314/saje.v35i2

Couch, B. A., Wright, C. D., Freeman, S., Knight, J. K., Semsar, K., Smith, M. K., Summers, M. M., Zheng, Y., Crowe, A. J., & Brownell, S. E. (2019). GenBio-MAPS: A Programmatic Assessment to Measure Student Understanding of Vision and Change Core Concepts across General Biology Programs. *CBE—Life Sciences Education*, *18*(1), ar1. https://doi.org/10.1187/cbe.18-07-0117

Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart, and Winston. http://books.google.com/books?id=tfgkAQAAMAAJ

De Champlain, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical Education*, *44*(1), 109–117. https://doi.org/10.1111/j.1365-2923.2009.03425.x

Deane, T., Nomme, K., Jeffery, E., Pollock, C., & Birol, G. (2014). Development of the Biological Experimental Design Concept Inventory (BEDCI). *CBE—Life Sciences Education*, *13*(3), 540–551. https://doi.org/10.1187/cbe.13-11-0218

Deniz, H., Donnelly, L. A., & Yilmaz, I. (2008). Exploring the factors related to acceptance of evolutionary theory among Turkish preservice biology teachers: Toward a more informative conceptual ecology for biological evolution. *Journal of Research in Science Teaching*, *45*(4), 420–443. https://doi.org/10.1002/tea.20223

Dobzhansky, T. (1973). Nothing in Biology Makes Sense except in the Light of Evolution. *The American Biology Teacher*, *35*(3), 125–129. https://doi.org/10.2307/4444260

Dunk, R. D. P., Petto, A. J., Wiles, J. R., & Campbell, B. C. (2017). A multifactorial analysis of acceptance of evolution. *Evolution: Education and Outreach*, *10*(1), 4. https://doi.org/10.1186/s12052-017-0068-0

Dunk, R. D. P., & Wiles, J. R. (2018). *Changes in Acceptance of Evolution and Associated Factors during a Year of Introductory Biology: The Shifting Impacts of Biology Knowledge, Politics, Religion, Demographics, and Understandings of the Nature of Science* (p. 280479). bioRxiv. https://doi.org/10.1101/280479

Ferguson, D. G., & Jensen, J. L. (2021). Role models, compatibility, and knowledge lead to increased evolution acceptance. *Evolution: Education and Outreach*, *14*(1), 16. https://doi.org/10.1186/s12052-021-00155-x

Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A Description of Think Aloud Method and Protocol Analysis. *Qualitative Health Research*, *3*(4), 430–441. https://doi.org/10.1177/104973239300300403

Frisbie, D. A. (1992). The multiple true-false item format: A status review. *Educational Measurement: Issues and Practice*, *11*, 21–26. https://doi.org/10.1111/j.1745-3992.1992.tb00259.x

Furrow, R. E., & Hsu, J. L. (2019). Concept inventories as a resource for teaching evolution. *Evolution: Education and Outreach*, *12*(1), 2. https://doi.org/10.1186/s12052-018-0092-8

Gallup. (2019). *40% of Americans Believe in Creationism*. Gallup.Com. https://news.gallup.com/poll/261680/americans-believe-creationism.aspx

Gallup Inc. (2019). *Evolution, Creationism, Intelligent Design*. Gallup.Com. https://news.gallup.com/poll/21814/Evolution-Creationism-Intelligent-Design.aspx

García, A. A. (2011). Cognitive Interviews to Test and Refine Questionnaires. *Public Health Nursing*, *28*(5), 444–450. https://doi.org/10.1111/j.1525-1446.2010.00938.x

Glaze, A. L., & Goldston, M. J. (2015). U.S. Science Teaching and Learning of Evolution: A Critical Review of the Literature 2000–2014. *Science Education*, *99*(3), 500–518. https://doi.org/10.1002/sce.21158

Glaze, A. L., Snyder, S., Devereaux, R. A., & Smith, M. U. (2020). The Generalized Acceptance of EvolutioN Evaluation (GAENE) 3.0: Enhancement and Validation of a New Measure of Acceptance. *Journal of Higher Education Theory and Practice*, *20*(7), Article 7. https://doi.org/10.33423/jhetp.v20i7.3157

Gluckman, P., Beedle, A., Buklijas, T., Low, F., & Hanson, and M. (2016). *Principles of Evolutionary Medicine* (Second Edition, Second Edition). Oxford University Press.

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 Sciences Education*, *11*(4), 364–377. https://doi.org/10.1187/cbe.12-03-0026

Graves, J. L., Reiber, C., Thanukos, A., Hurtado, M., & Wolpaw, T. (2016). Evolutionary science as a method to facilitate higher level thinking and reasoning in medical training. *Evolution, Medicine, and Public Health*, *2016*(1), 358–368. https://doi.org/10.1093/emph/eow029

Gregory, T. R. (2008). Evolution as Fact, Theory, and Path. *Evolution: Education and Outreach*, *1*(1), Article 1. https://doi.org/10.1007/s12052-007-0001-z

Grunspan, D. Z., Dunk, R. D. P., Barnes, M. E., Wiles, J. R., & Brownell, S. E. (2021). A comparison study of human examples vs. Non-human examples in an evolution lesson leads to differential impacts on student learning experiences in an introductory biology course. *Evolution: Education and Outreach*, *14*(1), 9. https://doi.org/10.1186/s12052-021-00148-w

Grunspan, D. Z., Moeller, K. T., Nesse, R. M., & Brownell, S. E. (2019). The state of evolutionary medicine in undergraduate education. *Evolution, Medicine, and Public Health*, *2019*(1), 82–92. https://doi.org/10.1093/emph/eoz012

Grunspan, D. Z., Nesse, R. M., Barnes, M. E., & Brownell, S. E. (2018). Core principles of evolutionary medicine: A Delphi study. *Evolution, Medicine, and Public Health*, *2018*(1), 13–23. https://doi.org/10.1093/emph/eox025

Hidaka, B. H., Asghar, A., Aktipis, C. A., Nesse, R. M., Wolpaw, T. M., Skursky, N. K., Bennett, K. J., Beyrouty, M. W., & Schwartz, M. D. (2015). The status of evolutionary medicine education in North American medical schools. *BMC Medical Education*, *15*(1), 38. https://doi.org/10.1186/s12909-015-0322-5

Ingram, E. L., & Nelson, C. E. (2006). Relationship between achievement and students' acceptance of evolution or creation in an upper-level evolution course. *Journal of Research in Science Teaching*, *43*(1), 7–24. https://doi.org/10.1002/tea.20093

International Society for Evolution, Medicine, and Public Health. (n.d.). *ISEMPH - EvMedEd*. EvMedEd. Retrieved February 6, 2023, from https://isemph.org/EvMedEd

Johns, R. (2005). One Size Doesn't Fit All: Selecting Response Scales For Attitude Items. *Journal of Elections, Public Opinion and Parties*, *15*(2), 237–264. https://doi.org/10.1080/13689880500178849

Kalas, P., O'Neill, A., Pollock, C., & Birol, G. (2013). Development of a Meiosis Concept Inventory. *CBE—Life Sciences Education*, *12*(4), 655–664. https://doi.org/10.1187/cbe.12-10-0174

Knight, J. (2010). Biology concept assessment tools: Design and use. *Microbiology Australia*, *31*(1), 5–8. https://doi.org/10.1071/ma10005

Koevering, T. E., & Stiehl, R. (1989). *Evolution, Creation & Wisconsin Biology Teachers*. https://doi.org/10.2307/4448903

Krippendorff, K. (2018). *Content Analysis: An Introduction to Its Methodology*. SAGE Publications.

Kummer, T. A., Whipple, C. J., Bybee, S. M., Adams, B. J., & Jensen, J. L. (2019). Development of an Evolutionary Tree Concept Inventory. *Journal of Microbiology & Biology Education*, *20*(2), 20.2.42. https://doi.org/10.1128/jmbe.v20i2.1700

Kuschmierz, P., Beniermann, A., Bergmann, A., Pinxten, R., Aivelo, T., Berniak-Woźny, J., Bohlin, G., Bugallo-Rodriguez, A., Cardia, P., Cavadas, B. F. B. P., Cebesoy, U. B., Cvetković, D. D., Demarsy, E., Đorđević, M. S., Drobniak, S. M., Dubchak, L., Dvořáková, R. M., Fančovičová, J., Fortin, C., … Graf, D. (2021). European first-year university students accept evolution but lack substantial knowledge about it: A standardized European cross-country assessment. *Evolution: Education and Outreach*, *14*(1), 17. https://doi.org/10.1186/s12052-021-00158-8

Kuschmierz, P., Beniermann, A., & Graf, D. (2020). Development and evaluation of the knowledge about evolution 2.0 instrument (KAEVO 2.0). *International Journal of Science Education*, *42*(15), 2601–2629. https://doi.org/10.1080/09500693.2020.1822561

Kuschmierz, P., Meneganzin, A., Pinxten, R., Pievani, T., Cvetković, D., Mavrikaki, E., Graf, D., & Beniermann, A. (2020). Towards common ground in measuring acceptance of evolution and knowledge about evolution across Europe: A systematic review of the state of research. *Evolution: Education and Outreach*, *13*(1), 18. https://doi.org/10.1186/s12052-020-00132-w

Laidlaw, C. T., Bybee, S. M., Shumway, S., Ogden, T. H., Peck, S., & Jensen, J. L. (2022). An Examination of Constructivism, Active Learning, and Reflexive Journaling and Their Independent and Combined Effects on Student Acceptance of Biological Evolution. *Journal of College Science Teaching*, *51*(3), 88–98.

Lawson, A. E. (1983). Predicting science achievement: The role of developmental level, disembedding ability, mental capacity, prior knowledge, and beliefs. *Journal of Research in Science Teaching*, *20*(2), 117–129. https://doi.org/10.1002/tea.3660200204

Lenski, R. E. (2020). *Evolution: Fact and Theory*. National Center for Science Education. https://ncse.ngo/evolution-fact-and-theory

Lloyd-Strovas, J. D., & Bernal, X. E. (2012). A Review of Undergraduate Evolution Education in U.S. Universities: Building a Unifying Framework. *Evolution: Education and Outreach*, *5*(3), 453–465. https://doi.org/10.1007/s12052-012-0435-9

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley.

Marbach-Ad, G., Briken, V., El-Sayed, N. M., Frauwirth, K., Fredericksen, B., Hutcheson, S., Gao, L.-Y., Joseph, S., Lee, V. T., McIver, K. S., Mosser, D., Quimby, B. B., Shields, P., Song, W., Stein, D. C., Yuan, R. T., & Smith, A. C. (2009). Assessing student understanding of host pathogen interactions using a concept inventory. *Journal of Microbiology & Biology Education*, *10*(1), 43–50. https://doi.org/10.1128/jmbe.v10.98

Mead, L. S., Kohn, C., Warwick, A., & Schwartz, K. (2019). Applying measurement standards to evolution education assessment instruments. *Evolution: Education and Outreach*, *12*(1), 5. https://doi.org/10.1186/s12052-019-0097-y

Metzger, K. J., Montplaisir, D., Haines, D., & Nickodem, K. (2018). Investigating undergraduate health sciences students' acceptance of evolution using MATE and GAENE. *Evolution: Education and Outreach*, *11*(1), 10. https://doi.org/10.1186/s12052-018-0084-8

Michigan State University. (n.d.). *Evo-Ed: Making Evolution Relevant and Accessible*. EvoEd. Retrieved April 19, 2023, from https://evo-ed.org/

Miller, J. D., Scott, E. C., Ackerman, M. S., Laspra, B., Branch, G., Polino, C., & Huffaker, J. S. (2022). Public acceptance of evolution in the United States, 1985–2020. *Public Understanding of Science*, *31*(2), 223–238. https://doi.org/10.1177/09636625211035919

Moltzau Anderson, J., & Horn, F. (2020). (Re-) Defining evolutionary medicine. *Ecology and Evolution*, *10*(20), 10930–10936. https://doi.org/10.1002/ece3.6825

Morrow, E. H. (2015). The evolution of sex differences in disease. *Biology of Sex Differences*, *6*, 5. https://doi.org/10.1186/s13293-015-0023-0

Mueller, S., & Reiners, C. S. (2022). Pre-service Chemistry Teachers' Views about the Tentative and Durable Nature of Scientific Knowledge. *Science & Education*. https://doi.org/10.1007/s11191-022-00374-8

Nadelson, L. S., & Hardy, K. K. (2015). Trust in science and scientists and the acceptance of evolution. *Evolution: Education and Outreach*, *8*(1), 9. https://doi.org/10.1186/s12052-015-0037-4

Nadelson, L. S., & Sinatra, G. M. (2009). Educational Professionals' Knowledge and Acceptance of Evolution. *Evolutionary Psychology*, *7*(4), 147470490900700400. https://doi.org/10.1177/147470490900700401

Nadelson, L. S., & Southerland, S. (2012). A More Fine-Grained Measure of Students' Acceptance of Evolution: Development of the Inventory of Student Evolution Acceptance—I-SEA. *International Journal of Science Education*, *34*(11), 1637–1666. https://doi.org/10.1080/09500693.2012.702235

Nadelson, L. S., & Southerland, S. A. (2009). Development and Preliminary Evaluation of the Measure of Understanding of Macroevolution: Introducing the MUM. *The Journal of Experimental Education*, *78*(2), 151–190. https://doi.org/10.1080/00220970903292983

Nápoles-Springer, A. M., Santoyo-Olsson, J., O'Brien, H., & Stewart, A. L. (2006). Using cognitive interviews to develop surveys in diverse populations. *Medical Care*, *44*(11 Suppl 3), S21-30. https://doi.org/10.1097/01.mlr.0000245425.65905.1d

National Academy of Sciences. (2008). *Science, Evolution, and Creationism*. https://doi.org/10.17226/11876

Natterson-Horowitz, B., Aktipis, A., Fox, M., Gluckman, P. D., Low, F., Mace, R., Read, A. F., Turner, P. E., & Blumstein, D. T. (2023). The future of evolutionary medicine: Sparking innovation in biomedicine and public health. *Frontiers in Science*. https://doi.org/10.3389/fsci.2023.997136

Nehm, R. H., Beggrow, E. P., Opfer, J. E., & Ha, M. (2012). Reasoning About Natural Selection: Diagnosing Contextual Competency Using the ACORNS Instrument. *The American Biology Teacher*, *74*(2), 92–98. https://doi.org/10.1525/abt.2012.74.2.6

Nehm, R. H., & Ha, M. (2011). Item feature effects in evolution assessment. *Journal of Research in Science Teaching*, *48*(3), 237–256. https://doi.org/10.1002/tea.20400

Nehm, R. H., & Mead, L. S. (2019). Evolution assessment: Introduction to the special issue. *Evolution: Education and Outreach*, *12*(1), 7. https://doi.org/10.1186/s12052-019-0098-x

Neiles, K. Y., & Arnett, K. (2021). Backward Design of Chemistry Laboratories: A Primer. *Journal of Chemical Education*, *98*(9), 2829–2839. https://doi.org/10.1021/acs.jchemed.1c00443

Nesse, R. M., Bergstrom, C. T., Ellison, P. T., Flier, J. S., Gluckman, P., Govindaraju, D. R., Niethammer, D., Omenn, G. S., Perlman, R. L., Schwartz, M. D., Thomas, M.

G., Stearns, S. C., & Valle, D. (2010). Making evolutionary biology a basic science for medicine. *Proceedings of the National Academy of Sciences*, *107*(suppl_1), 1800–1807. https://doi.org/10.1073/pnas.0906224106

Nettle, D. (2010). Understanding of Evolution May Be Improved by Thinking about People. *Evolutionary Psychology*, *8*(2), 147470491000800200. https://doi.org/10.1177/147470491000800206

Orr, R. B., Csikari, M. M., Freeman, S., & Rodriguez, M. C. (2022). Writing and Using Learning Objectives. *CBE—Life Sciences Education*, *21*(3), fe3. https://doi.org/10.1187/cbe.22-04-0073

Perlman, R. (2013). *Evolution and Medicine*. Oxford University Press.

Pew. (2013, December 30). Public's Views on Human Evolution. *Pew Research Center's Religion & Public Life Project*. http://www.pewforum.org/2013/12/30/publics-views-on-human-evolution/

Pew Research Center. (2019, February 6). The Evolution of Pew Research Center's Survey Questions About the Origins and Development of Life on Earth. *Pew Research Center's Religion & Public Life Project*. https://www.pewresearch.org/religion/2019/02/06/the-evolution-of-pew-research-centers-survey-questions-about-the-origins-and-development-of-life-on-earth/

Plutzer, E., Branch, G., & Reid, A. (2020). Teaching evolution in U.S. public schools: A continuing challenge. *Evolution: Education and Outreach*, *13*(1), 14. https://doi.org/10.1186/s12052-020-00126-8

Popper, K. (1959). *The Logic of Scientific Discovery*. Routledge. https://www.routledge.com/The-Logic-of-Scientific-Discovery/Popper/p/book/9780203994627

Power, M. L., Snead, C., Reed, E. G., & Schulkin, J. (2020). Integrating evolution into medical education for women's health care practitioners. *Evolution, Medicine, and Public Health*, *2020*(1), 60–67. https://doi.org/10.1093/emph/eoaa009

Preuss, T. M., & Robert, J. S. (2014). Animal Models of the Human Brain: Repairing the Paradigm. In M. S. Gazzaniga & G. R. Mangun (Eds.), *The Cognitive Neurosciences* (5th ed., pp. 59–66). MIT Press.

Price, R. M., Andrews, T. C., McElhinny, T. L., Mead, L. S., Abraham, J. K., Thanukos, A., & Perez, K. E. (2014). The Genetic Drift Inventory: A Tool for Measuring What Advanced Undergraduates Have Mastered about Genetic Drift. *CBE—Life Sciences Education*, *13*(1), 65–75. https://doi.org/10.1187/cbe.13-08-0159

Revilla, M., & Höhne, J. K. (2020). How long do respondents think online surveys should be? New evidence from two online panels in Germany. *International Journal of Market Research*, *62*(5), 538–545. https://doi.org/10.1177/1470785320943049

Reynolds, H. L., & Kearns, K. D. (2017). A Planning Tool for Incorporating Backward Design, Active Learning, and Authentic Assessment in the College Classroom. *College Teaching*, *65*(1), 17–27. https://doi.org/10.1080/87567555.2016.1222575

Reznick, D. N., & Ricklefs, R. E. (2009). Darwin's bridge between microevolution and macroevolution. *Nature*, *457*(7231), Article 7231. https://doi.org/10.1038/nature07894

Rissler, L. J., Duncan, S. I., & Caruso, N. M. (2014). The relative importance of religion and education on university students' views of evolution in the Deep South and state science standards across the United States. *Evolution: Education and Outreach*, *7*(1), 24. https://doi.org/10.1186/s12052-014-0024-1

Romine, W. L., Todd, A. N., & Walter, E. M. (2018). A closer look at the items within three measures of evolution acceptance: Analysis of the MATE, I-SEA, and GAENE as a single corpus of items. *Evolution: Education and Outreach*, *11*(1), 17. https://doi.org/10.1186/s12052-018-0093-7

Romine, W. L., Walter, E. M., Bosse, E., & Todd, A. N. (2016). Understanding patterns of evolution acceptance—A new implementation of the Measure of Acceptance of the Theory of Evolution (MATE) with Midwestern university students. *Journal of Research in Science Teaching*, n/a-n/a. https://doi.org/10.1002/tea.21380

Romine, W. L., Walter, E. M., Bosse, E., & Todd, A. N. (2017). Understanding patterns of evolution acceptance—A new implementation of the Measure of Acceptance of the Theory of Evolution (MATE) with Midwestern university students. *Journal of Research in Science Teaching*, *54*(5), 642–671. https://doi.org/10.1002/tea.21380

Rühli, F. J., & Henneberg, M. (2013). New perspectives on evolutionary medicine: The relevance of microevolution for human health and disease. *BMC Medicine*, *11*(1), 115. https://doi.org/10.1186/1741-7015-11-115

Rutledge, M. L., & Sadler, K. C. (2007). Reliability of the Measure of Acceptance of the Theory of Evolution (MATE) instrument with university students. *The American Biology Teacher*, *69*(6), 332–335.

Rutledge, M. L., & Warden, M. A. (1999). The Development and Validation of the Measure of Acceptance of the Theory of Evolution Instrument. *School Science and Mathematics*, *99*(1), 13–18. https://doi.org/10.1111/j.1949-8594.1999.tb17441.x

Rutledge, M. L., & Warden, M. A. (2000). Evolutionary Theory, the Nature of Science & High School Biology Teachers: Critical Relationships. *The American Biology Teacher*, *62*(1), 23–31. https://doi.org/10.2307/4450822

Şahin, A., & Anil, D. (2017). The Effects of Test Length and Sample Size on Item Parameters in Item Response Theory. *Educational Sciences: Theory & Practice*. https://doi.org/10.12738/estp.2017.1.0270

Sbeglia, G. C., & Nehm, R. H. (2018). Measuring evolution acceptance using the GAENE: Influences of gender, race, degree-plan, and instruction. *Evolution: Education and Outreach*, *11*(1), 18. https://doi.org/10.1186/s12052-018-0091-9

Sbeglia, G. C., & Nehm, R. H. (2019). Do you see what I-SEA? A Rasch analysis of the psychometric properties of the Inventory of Student Evolution Acceptance. *Science Education*, *103*(2), 287–316. https://doi.org/10.1002/sce.21494

Semsar, K., Brownell, S., Couch, B. A., Crowe, A. J., Smith, M. K., Summers, M. M., Wright, C. D., & Knight, J. K. (2019). Phys-MAPS: A programmatic physiology assessment for introductory and advanced undergraduates. *Advances in Physiology Education*, *43*(1), 15–27. https://doi.org/10.1152/advan.00128.2018

Short, S. D., & Hawley, P. H. (2012). Evolutionary attitudes and literacy survey (EALS): Development and validation of a short form. *Evolution: Education and Outreach*, *5*(3), 419–428.

Shtulman, A., & Schulz, L. (2008). The Relation Between Essentialist Beliefs and Evolutionary Reasoning. *Cognitive Science*, *32*(6), 1049–1062. https://doi.org/10.1080/03640210801897864

Siciliano-Martina, L., & Martina, J. P. (2020). Shifting barriers to the acceptance of evolution in an underrepresented student group. *International Journal of Science Education*, *42*(13), 2205–2223. https://doi.org/10.1080/09500693.2020.1815247

Sickel, A. J., & Friedrichsen, P. (2013). Examining the evolution education literature with a focus on teachers: Major findings, goals for teacher preparation, and directions for future research. *Evolution: Education and Outreach*, *6*(1), 23. https://doi.org/10.1186/1936-6434-6-23

Sinatra, G. M., Southerland, S. A., McConaughy, F., & Demastes, J. W. (2003). Intentions and beliefs in students' understanding and acceptance of biological evolution. *Journal of Research in Science Teaching*, *40*(5), 510–528. https://doi.org/10.1002/tea.10087

Smith, M. K., Brownell, S. E., Crowe, A. J., Holmes, N. G., Knight, J. K., Semsar, K., Summers, M. M., Walsh, C., Wright, C. D., & Couch, B. A. (2019). Tools for

Change: Measuring Student Conceptual Understanding Across Undergraduate Biology Programs Using Bio-MAPS Assessments. *Journal of Microbiology & Biology Education*, *20*(2), 20. https://doi.org/10.1128/jmbe.v20i2.1787

Smith, M. U. (2009). Current Status of Research in Teaching and Learning Evolution: I. Philosophical/Epistemological Issues. *Science & Education*, *19*(6–8), 523–538. https://doi.org/10.1007/s11191-009-9215-5

Smith, M. U., & Scharmann, L. C. (1999). Defining versus describing the nature of science: A pragmatic analysis for classroom teachers and science educators. *Science Education*, *83*(4), 493–509. https://doi.org/10.1002/(SICI)1098-237X(199907)83:4<493::AID-SCE6>3.0.CO;2-U

Smith, M. U., & Siegel, H. (2004). Knowing, Believing, and Understanding: What Goals for Science Education? *Science & Education*, *13*(6), 553–582. https://doi.org/10.1023/B:SCED.0000042848.14208.bf

Smith, M. U., & Siegel, H. (2016). On the Relationship Between Belief and Acceptance of Evolution as Goals of Evolution Education. *Science & Education*, *25*(5), 473–496. https://doi.org/10.1007/s11191-016-9836-4

Smith, M. U., Snyder, S. W., & Devereaux, R. S. (2016). The GAENE—Generalized Acceptance of EvolutioN Evaluation: Development of a new measure of evolution acceptance. *Journal of Research in Science Teaching*, *53*(9), 1289–1315. https://doi.org/10.1002/tea.21328

Smith, V. H., Rubinstein, R. J., Park, S., Kelly, L., & Klepac-Ceraj, V. (2015). Microbiology and ecology are vitally important to premedical curricula. *Evolution, Medicine, and Public Health*, *2015*(1), 179–192. https://doi.org/10.1093/emph/eov014

Southerland, S. A., Sinatra, G. M., & Matthews, M. R. (2001). Belief, Knowledge, and Science Education. *Educational Psychology Review*, *13*(4), 325–351. https://doi.org/10.1023/A:1011913813847

Stanisavljevic, J., Papadopoulou, P., & Djuric, D. (2013). Relationship between Acceptance and Understanding of the Evolution Theory by Various Groups of Teachers. *Croatian Journal of Education : Hrvatski Časopis Za Odgoj i Obrazovanje*, *15*(3), 693–714.

Summers, M. M., Couch, B. A., Knight, J. K., Brownell, S. E., Crowe, A. J., Semsar, K., Wright, C. D., & Smith, M. K. (2018). EcoEvo-MAPS: An Ecology and Evolution Assessment for Introductory through Advanced Undergraduates. *CBE—Life Sciences Education*, *17*(2), ar18. https://doi.org/10.1187/cbe.17-02-0037

Teasdale, R., & Aird, H. (2023). Aligning multiple choice assessments with active learning instruction: More accurate and equitable ways to measure student learning. *Journal of Geoscience Education*, *71*(1), 87–106. https://doi.org/10.1080/10899995.2022.2081462

Thrall, P. H., Oakeshott, J. G., Fitt, G., Southerton, S., Burdon, J. J., Sheppard, A., Russell, R. J., Zalucki, M., Heino, M., & Ford Denison, R. (2011). Evolution in agriculture: The application of evolutionary approaches to the management of biotic interactions in agro-ecosystems. *Evolutionary Applications*, *4*(2), 200–215. https://doi.org/10.1111/j.1752-4571.2010.00179.x

Trani, R. (2004). I Won't Teach Evolution; It's against My Religion. And Now for the Rest of the Story... *The American Biology Teacher*, *66*(6), 419–427. https://doi.org/10.2307/4451708

University of California Berkeley. (2021a, September 3). Characteristics of science. *Understanding Evolution*. https://evolution.berkeley.edu/nature-of-science/characteristics-of-science/

University of California Berkeley. (2021b, September 19). Misconceptions about evolution. *Understanding Evolution*. https://evolution.berkeley.edu/teach-evolution/misconceptions-about-evolution/

U.S. Bureau of Labor Statistics. (2019). *Field of degree: Biology : Occupational Outlook Handbook*. https://www.bls.gov/ooh/field-of-degree/biology/biology-field-of-degree.htm

Wagler, A., & Wagler, R. (2013). Addressing the Lack of Measurement Invariance for the Measure of Acceptance of the Theory of Evolution. *International Journal of Science Education*, *35*(13), 2278–2298. https://doi.org/10.1080/09500693.2013.808779

Wiggins, G., & McTighe, J. (2005). *Understanding by Design*. ASCD.

Wiles, J. R. (2014). Gifted students' perceptions of their acceptance of evolution, changes in acceptance, and factors involved therein. *Evolution: Education and Outreach*, *7*(1), 4. https://doi.org/10.1186/s12052-014-0004-5

Willis, G. B. (2004). *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Sage. https://us.sagepub.com/en-us/nam/cognitive-interviewing/book225856

Wilson, R., Roff, A., Dai, P., Fortugno, T., Douds, J., Chen, G., Grove, G., Nikiforova, S., Barnholtz-Sloan, J., Frudakis, T., Chinchilli, V., Hartman, T., Demers, L., Shriver, M., Canfield, V., & Cheng, K. (2011). Genetic Ancestry, Skin Reflectance and Pigmentation Genotypes in Association with Serum Vitamin D

Metabolite Balance. *Hormone Molecular Biology and Clinical Investigation*, *7*, 279–293. https://doi.org/10.1515/HMBCI.2011.021

Wright, C., Huang, A., Cooper, K., & Brownell, S. (2018). Exploring Differences In Decisions About Exams Among Instructors Of The Same Introductory Biology Course. *International Journal for the Scholarship of Teaching and Learning*, *12*(2). https://doi.org/10.20429/ijsotl.2018.120214

Zimmer, C. (2013). *The Tangled Bank: An Introduction to Evolution* (Second edition). Roberts and Company.

Zimmerman, M. (1987). The Evolution-Creation Controversy: Opinions of Ohio High School Biology Teachers. *The Ohio Journal of Science*, *87*(4), 115–125.

APPENDIX A

EVMED ASSESSMENT

1. Consider the following two genetic disorders:

HUNTINGTON'S CHOREA

Huntington's chorea is a rare genetic disorder that is characterized by depression, forgetfulness, involuntary movements, and slurred speech. Symptoms of Huntington's start appearing between the ages of 30 and 50, and typically worsen over time. Huntington's leads to death about 15 - 20 years after the onset of symptoms. The disorder is caused by mutations found in a gene called the Huntington gene. Huntington's afflicts 0.1% of individuals in Europe, with rates tending to be even lower among populations of non-European ancestry. However, one place where rates of Huntington's chorea are higher than in populations of European descent is the village of Barranquitas in Venezuela, where the percent of individuals with Huntington's is closer to 10%. Huntington's chorea is caused by novel mutations only 10% of the time.

TAY-SACHS DISEASE

Tay-Sachs disease is a rare genetic disorder. Approximately 0.4% of individuals are carriers of Tay-Sachs, but this rate is closer to 4% in Ashkenazi Jewish, French Canadian, and Louisiana Cajun populations. Most individuals born with Tay Sachs die within the first 4 years of their life.

Using the information above, evaluate the following statements as either likely or unlikely.

1A.) Because individuals die due to the disease, the prevalence of Huntington's chorea will inevitably reach zero in Barranquitas.

Likely (0)   Unlikely (1)

1B.) Natural selection is a stronger evolutionary force on Tay-Sachs than on Huntington's chorea.

Likely (1)   Unlikely (0)

1C.) Natural selection is the best explanation for the high rate of Huntington's chorea in the Venezuelan population.

Likely (0)   Unlikely (1)

1D.) All humans have a copy of the Huntington gene in their genome.

Likely (1)   Unlikely (0)

2. Virus A is a seasonal airborne virus that is more common in the fall than at other times of the year. While this virus is responsible for many deaths each year, most healthy individuals are usually able to clear the virus after a couple of weeks. These individuals are then immune to the specific strain they were infected with for the rest of their lives. However, humans are exposed to new strains of this virus every year.

2A.) A single vaccination during childhood would provide effective lifelong protection against Virus A.

Likely (0)   Unlikely (1)

2B.) Imagine a strain of Virus A emerges that kills its host 100% of the time, but is otherwise similar to past strains. This new strain would be more likely to spread if it killed its host quickly after initial infection.

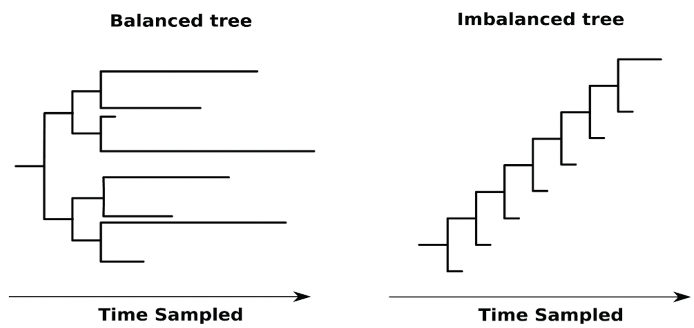
Likely (0)   Unlikely (1)

2C.) Imagine a strain of Virus A emerges that kills its host 100% of the time, but is otherwise similar to past strains. This new strain would be more likely to persist in a city with 6 million people compared to a small rural population.

Likely (1)   Unlikely (0)

2D.) An antibiotic would be an effective treatment against Virus A.

Likely (0)   Unlikely (1)



2E.) Researchers took samples of Virus A each year over the past 30 years from sick individuals. When they reconstruct a phylogeny from these viral samples, it would more closely resemble an imbalanced tree than a balanced tree

Likely (1)   Unlikely (0)

3. Many infectious illnesses are accompanied by a fever. Some drugs, such as ibuprofen, are designed to reduce fever. Having a fever contributes to the sensation of being sick, so many people take fever-reducing drugs in order to feel better. A woman decides to take ibuprofen because she currently has a fever due to having the flu.

Using the information above, evaluate the following statements as either likely or unlikely.

3A.) The ibuprofen will slow down the rate at which the flu virus replicates.

Likely (0)   Unlikely (1)

3B.) The ibuprofen will help fight off the flu virus.

Likely (0)   Unlikely (1)

3C.) The ibuprofen will lead to drug resistant flu viruses.
Likely (0) Unlikely (1)

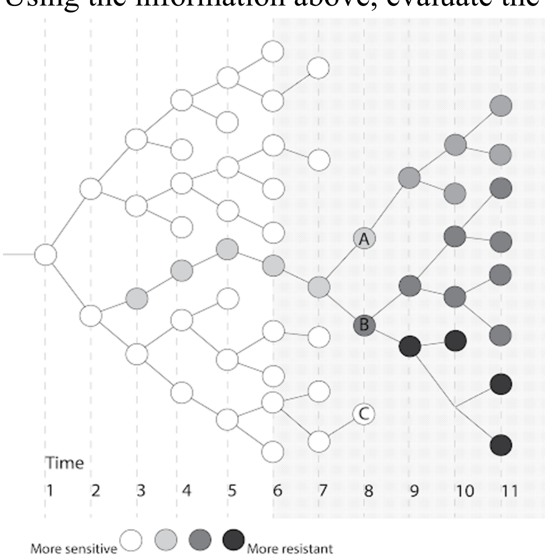
3D.) Discovering drugs like ibuprofen relies on research that uncovers evolutionary (ultimate) explanations rather than research that uncovers mechanistic (proximate) explanations.
Likely (0) Unlikely (1)

4. Chemotherapy is used to kill cancerous tumor cells. Historically, chemotherapy is administered in high doses aimed to eliminate all of the cells within a tumor. Tumors shrink in response to chemotherapy, but most often they eventually grow back because the cancer cells become resistant to the chemotherapy drug.

A researcher proposes a new way to administer chemotherapy, which they call adaptive therapy. They suggest using occasional doses of chemotherapy with the goal of killing most, but not all, of the tumor cells. They say that an issue with constant high dosage chemotherapy is that it leaves chemotherapy-resistant tumor cells without chemotherapy-sensitive cells to compete with for growth.

The figure below shows a model of a tumor over time, where traditional high dosage chemotherapy is started at time 6. Each circle represents a cancer cell. White cells are sensitive to chemotherapy, while darker cells show greater resistance to chemotherapy.

Using the information above, evaluate the following statements as likely or unlikely.



4A.) Tumors become chemotherapy resistant because each cancer cell individually develops resistance to chemotherapy over time.
Likely (0) Unlikely (1)

4B.) Assume chemotherapy was stopped between times 7 and 8 as part of an adaptive therapy treatment. For this to successfully prevent the development of a chemotherapy resistant tumor, cell C would have to outcompete cells A and B.

Likely (1) Unlikely (0)

4C.) In the absence of chemotherapy, chemotherapy-resistant cancer cells replicate faster than chemotherapy-sensitive cells.
Likely (0) Unlikely (1)

4D.) The high dosage chemotherapy would have been more effective at completely eliminating the tumor if it started at time 4.

Likely (0) Unlikely (1)

4E.) Adaptive therapy would be effective if the first round of chemotherapy that starts at time 6 was stopped at time 10.
Likely (0) Unlikely (1)

4F) The chemotherapy-resistant cancer cells had no selective advantage over the non-resistant cancer cells before chemotherapy was started.
Likely (1) Unlikely (0)

5. Humans vary in the frequency and severity of anxiety they experience. Some people have generalized anxiety disorders (GAD), which are characterized as excessive anxiety and worry about various events or activities. Most people experience a moderate amount of anxiety, while some experience no anxiety at all. Over the past several decades, the number of individuals with GAD has increased.

5A.) Individuals who never experience anxiety have higher fitness than those who experience anxiety in moderate amounts.
Likely (0) Unlikely (1)

5B.) Generalized anxiety disorder is a useful adaptation.
Likely (0) Unlikely (1)

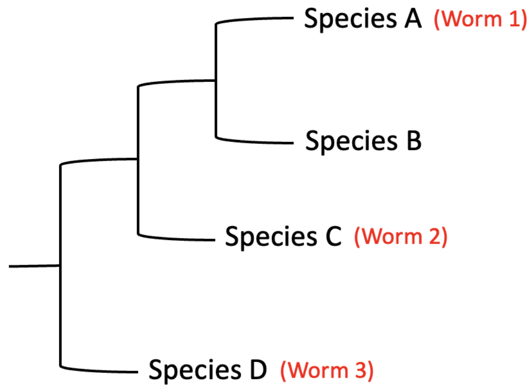
5C.) Theoretically, natural selection will result in disorders of too much anxiety more often than disorders of too little anxiety.
Likely (1) Unlikely (0)

5D.) The increase in the number of individuals with generalized anxiety disorder over the past several decades is being driven by natural selection.
Likely (0) Unlikely (1)

6.) The phylogeny below shows the relationship between four mammal species – A, B, C, and D. Species A is commonly infected with Worm 1, Species C is commonly infected with Worm 2, and Species D is commonly infected with Worm 3. In these three mammal species, individuals have chronic worm infections that typically start shortly after birth. Species B does not experience worm infections in the wild; it lives in a region that is not known to have any worms.

A zoo established captive breeding populations of Species A and B. All animals in the initial captive population of Species A were infected with Worm 1. The initial captive population of Species B entered the zoo without any infected individuals.

Shortly after the captive populations were established, all of the Species B individuals became infected with Worm 1. In response, the zoo de-wormed all captive animals of Species A and B. The animals were all adults when they were de-wormed. All future generations never experienced worm infections.



6A.) The last common ancestor of the four species experienced infections with Worm 3.
Likely (0) Unlikely (1)

6B) The symptoms of infection with Worm 1 are more harmful for Species B than Species A.
Likely (1) Unlikely (0)

6C) Species B is more likely than Species A to experience allergies in the zoo.
Likely (0) Unlikely (1)

7. The Daf-2 gene in nematode worms is thought to play some kind of role in energy allocation. Researchers designed an experiment that more closely examines the function of the Daf-2 gene in *C. elegans*, a species of nematode worm. First, the researchers created a strain of worms that had a mutation in their Daf-2 gene. They found that the mutant worms had a longer lifespan than wild type worms without the mutation.

The researchers then tracked two populations of worms over time; one population started out with 50 Daf-2 mutant worms, and the other population started out with 50 wild type worms. After 8 generations, the researchers counted the number of worms in each population. The individual worms at the end of the experiment were not the same ones as at the start; all of them had been born during the course of the experiment. The results of this research are below.

	Number of worms: start of experiment	Number of worms: end of experiment	Average lifespan
Wild Type	50	84	14 days
Mutant	50	3	27 days

7A.) Nematodes with the *Daf-2* mutation probably allocate less energy to reproduction than nematodes without the mutation.
Likely (1) Unlikely (0)

7B.) Nematodes with the *Daf-2* mutation probably allocate less energy to somatic repair than nematodes without the mutation.
Likely (0) Unlikely (1)

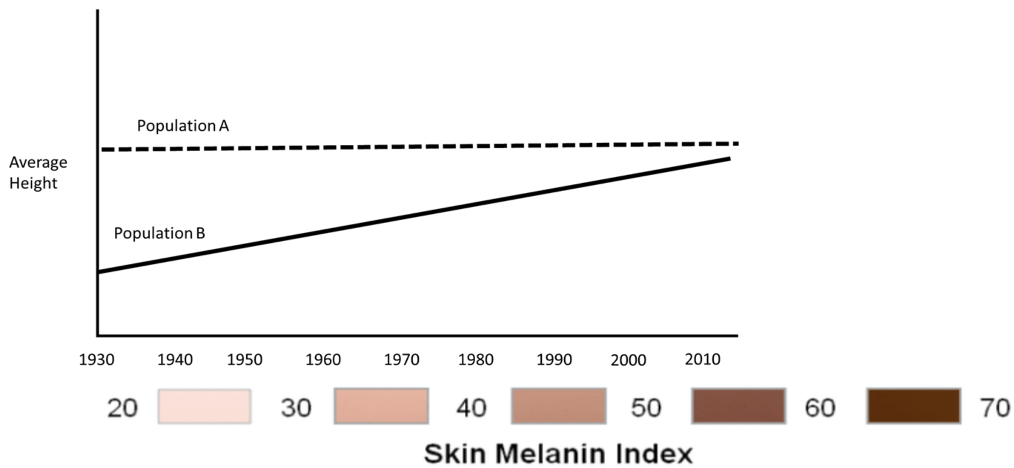
7C.) *Daf-2* influences two or more phenotypic traits.
Likely (1) Unlikely (0)

7D.) The *Daf-2* mutation that results in longer lifespan increases the fitness of nematodes.
Likely (0) Unlikely (1)

8. The average height of individuals from two different human populations between 1930 and 2010 is shown below. Population A is an equatorial population living in a relatively industrialized country. People in population A have access to good sanitation and a wide variety of nutritious food. Very little migration has occurred into or out of Population B over the past 80 years, although many health initiatives have taken place along with increased industrialization.

Individuals from Population A also have much darker skin color than those from Population B. Skin color can be measured via melanin index (MI), where a higher MI indicates darker skin, while a lower MI indicates less melanin and lighter skin. Most people in Population A have melanin indices of 50 - 60; while most people in Population B have melanin indices of 20 - 30.

From 1935-1940, thousands of families immigrated from Population B to Population A. Many immigrant couples had children after arriving in Population A.



Reference: (Wilson et al., 2011)

8A.) The graph above suggests that height is undergoing natural selection in population B.

Likely (0) Unlikely (1)

8B.) Children born in Population A to Population B immigrant parents during the 1940s will grow up to be taller, over average, than their same-age peers in Population B.

Likely (1) Unlikely (0)

8C.) Children born in Population A to Population B immigrant parents will have melanin indices closer to 55 than 25.

Likely (0) Unlikely (1)

9. A parasite endemic to Island A causes a deadly illness in humans. This parasite is transmitted between human hosts by a species of fly, and is among the most common causes of death for individuals aged 10-45. The fly that carries the parasite is not found on Island B.

Humans have the Z gene. People who have at least one copy of the H1 allele of gene Z are resistant to the disease-causing parasite. However, those who are homozygous for the H1 allele are at high risk of developing Z-associated kidney disease, which leads to an early death. The "wild type" (WT) allele of gene Z does not protect the person from the disease-causing parasite; however, the WT allele also is not associated with kidney disease.

An allele that is not found on either island is H5*. Individuals homozygous for H5* are resistant to the parasite, while carriers of H5* do not experience increased risks of kidney disease.

	Island A	Island B

Parasite	Common	Absent
WT allele frequency	0.78	0.85
H1 allele frequency	0.22	0.15
H5* allele frequency	0.00	0.00

	Risk of Gene-Associated Chronic Kidney Disease	Susceptibility to Parasite
WT/WT	None	Susceptible
WT/H1	None	Resistant
H1/H1	High	Resistant
WT/H5*	None	Susceptible
H5*/H5*	None	Resistant

9A.) If the parasites from Island A were eliminated, the frequency of the H1 allele would decrease over time.

Likely (1)   Unlikely (0)

9B.) If a drug were administered to the population on Island A that cured chronic kidney disease with no side-effects, the frequency of the H1 allele would decrease over time.

Likely (0)   Unlikely (1)

9C.) Assuming the islands have the same population size, a novel H5* allele is more likely to arise via mutation on Island A than on Island B.

Likely (0)   Unlikely (1)

9D.) If two teenage brothers, each with WT/H5* genotypes, immigrated to Island A, there would immediately be strong selection for the H5* allele leading to a higher H5* allele frequency over time.

Likely (0)   Unlikely (1)

9E.) If the allelic frequency of H5* on Island A was 0.2, it would face stronger positive selection if its effect on parasite resistance was dominant as opposed to recessive.

Likely (1)   Unlikely (0)

10. Imagine two species, Species A and species B. These species diverged from one another 20 million years ago. Species A is larger than Species B. Individuals of Species

A die of old age at around 40 years, while individuals of Species B die of old age at around 20 years. One of the two species faces higher levels of predation than the other.

10A.) Earlier natural death in Species B evolved as an adaptation that enables offspring to have access to resources and survive.

Likely (0) Unlikely (1)

10B.) Species A experiences lower levels of predation compared to Species B.

Likely (1) Unlikely (0)

10C.) Based on the information above, we can determine that the age of death in the common ancestor of Species A and Species B is around 30 years of age.

Likely (0) Unlikely (1)

10D.) Species A would be predicted to have more novel defense mechanisms against cancer compared to Species B.

Likely (1) Unlikely (0)

11. The Human Immunodeficiency Virus (HIV) infects and kills immune cells, which can lead to AIDS. AIDS is a very deadly condition that can develop in HIV-infected people of any age. The first recorded case of AIDS was in 1981, and researchers suspect that HIV first transferred to humans from other primates in the early 20th century.

Gene A influences the structure of immune cells. This gene has two alleles: A1 and A2. People who have a least one copy of the A1 allele have an average level of resistance against the flu virus, and are susceptible to HIV infection. People who have two copies of the A2 allele are highly resistant to HIV infection, but have weaker defenses against the flu virus. People who are homozygous for the A2 allele are 20% more likely to die from the flu after the age of 40.

In northern Europe, around 12% of people have at least one copy of the A2 allele, and a smaller proportion of people have two copies. The A2 allele is much more rare in the rest of the world, including southern Africa.

Table 1

Genotype	HIV Defense	Flu Defense
A1/A1	Susceptible	Normal
A1/A2	Susceptible	Normal
A2/A2	Immune	Lower, especially after age 40

Table 2			
Region	A2 Allele Frequency	Prevalence of HIV/AIDS	Prevalence of Flu
Southern Africa	3%	15% of people are diagnosed	Common (most encounter at least once in lifetime)
Northern Europe	12%	0.25% of people are diagnosed	Common (most encounter at least once in lifetime)

11A.) The relatively high frequency of the A2 allele in northern Europe is the result of selective pressure created by HIV/AIDS.

Likely (0) Unlikely (1)

11B.) The A2 allele was more adaptive in northern Europe than in southern Africa at some point before the 20th century, but is now probably more adaptive in southern Africa than in northern Europe.

Likely (1) Unlikely (0)

11C.) With genome-editing technology, it is possible to artificially introduce two copies of the A2 allele into a human embryo. Doing so is likely to benefit the health and longevity of a baby born in northern Europe.

Likely (0) Unlikely (1)

APPENDIX B

DEMOGRAPHIC QUESTIONS

1. What is your academic year?
    a. First-year
    b. Sophomore
    c. Junior
    d. Senior
    e. 5th year or higher

2. I most closely identify as:
    a. Female
    b. Male
    c. Nonbinary
    d. Decline to state
    e. Please describe your gender identity if the best option is not listed: ________

3. What is your ethnicity? Please select all that apply.
    a. American Indian, Native American, or Alaskan Native
    b. Asian or Asian American
    c. Black or African American
    d. Hispanic or Latino/Latina
    e. Native Hawaiian or Other Pacific Islander
    f. White or European American
    g. Decline to state
    h. Other, not listed: ________________________

4. Are you a native English speaker?
    a. Yes
    b. No, but I'm very comfortable with understanding English
    c. No, I sometimes struggle to understand English, but only in the spoken form
    d. No, I sometimes struggle to understand English, but only in the written form
    e. No, I sometimes struggle to understand English, in both the written and spoken form
    f. Decline to state

5. I most closely identify as:
    a. Agnostic (does not have a definite belief about whether God exists or not)
    b. Atheist (believes that God does not exist)
    c. Buddhist
    d. Christian- Catholic
    e. Christian- The Church of Jesus Christ of Latter-Day Saints
    f. Christian- Protestant
    g. Christian- Other (please describe): ____________________
    h. Hindu
    i. Jewish
    j. Muslim

k. Nothing in particular
l. Other faith (please describe): ____________________
m. Decline to state

6. Do you identify as an Evangelical Christian?
    a. Yes
    b. No
    c. I'm not sure
7. Do you identify with your religion mainly on a cultural -but not spiritual- basis?
    a. Yes
    b. No
    c. I'm not sure

Please indicate how much you agree or disagree with the following statements:

8. I attend religious services regularly.
9. I believe in God.
10. I consider myself a religious person.
11. I consider myself a spiritual person.

Options:
    a. Strongly Disagree
    b. Disagree
    c. Neutral
    d. Agree
    e. Strongly Agree

How much conflict do you perceive...

11. Between your personal religious beliefs and evolution?
12. Between the teachings of your religion and evolution?
13. Between your belief in God and evolution?
14. Between your religious culture and evolution?

Options:
    a. None
    b. A little
    c. A moderate amount
    d. A lot

15. Which college or university do you attend? __________
16. What is your current major?
    a. Biology
    b. Other STEM, please describe: __________
    c. Other non-STEM, please describe: __________
17. How many biology courses have you taken in college?
    a. 0
    b. 1 – 2
    c. 3 – 4
    d. 5+

18. Have you ever taken a college science course that was primarily on evolution?

a. Yes
b. No
I’m not sure

APPENDIX C

EXPLORING THE VALIDITY OF EVOLUTION ACCEPTANCE MEASURES – FREE RESPONSE INTERVIEW QUESTIONS

The interviewer asked each student the following questions after the cognitive interview on either the GAENE or the I-SEA was completed. This was a semi-structured interview in which the interviewer would ask follow-up questions based on the student's answers. The questions were prefaced using the following script:

"Now that we have completed the survey, I would like to ask you a few open-ended questions about your views on evolution. The purpose of these questions is for you to have a chance to describe your views in your own words. There are no right or wrong answers, and I encourage you to elaborate on your answers until you feel that you have fully conveyed your views about each of the questions. I might also ask some follow-up questions to make sure that I fully understand your answer."

1. What are your thoughts on human evolution?

2. Do you think that all of life on earth descended from one ancestral (original) species? If not, what types of species DO share a common ancestor? What types of species DON'T share a common ancestor?

APPENDIX D

SUPPLEMENTAL TABLE 1. PROFILES OF INTERVIEW PARTICIPANTS ON THE I-SEA AND GAENE.

I-SEA							
ID	**Pseud.**	**Gender**	**Race/ Ethnicity**	**Religion**	**Evo. Ed.**	**Interview-based Acceptance**	**Composite Score**
001	Deniz	F	Asian	Hindu	High	Full Acceptance	Macro: 4.5 Micro: 5.0 Human: 5.0
002	Darya	F	Hispanic	Agnostic	High	Full Acceptance	Macro: 4.2 Micro: 5.0 Human: 4.9
003	Ocean	F	Black	Christian - Protestant	High	Human: accept Non-human: higher taxa	Macro: 3.6 Micro: 5.0 Human: 4.9
004	Anup	M	Hispanic	Christian - LDS	High	Full Acceptance	Macro: 4.6 Micro: 5.0 Human: 4.5
005	Jafar	M	White	Agnostic	High	Full Acceptance	Macro: 4.9 Micro: 5.0 Human: 5.0
006	Moana	F	Hispanic	Atheist	Med	Full Acceptance	Macro: 4.5 Micro: 4.9 Human: 4.9
007	Poseidon	M	Asian	Muslim	High	Human: reject Non-human: reject	Macro: 1.6 Micro: 2.8 Human: 1.5
008	Marisol	F	White	Agnostic	High	Full Acceptance	Macro: 4.5 Micro: 4.6 Human: 4.5
009	Fishel	M	White	Muslim	Med	Full Acceptance	Macro: 4.8 Micro: 5.0 Human: 4.8
020	Anemone	F	Black	Christian-Other (Spiritual)	Low	Full Acceptance	Macro: 4.6 Micro: 4.6 Human: 4.8
021	Triton	M	Asian	Christian - Protestant	Med	Human: undecided Non-human: undecided	Macro: 3.3 Micro: 3.6 Human: 3.3
022	River	F	White	Other faith (Spirituality)	Med	Human: reject Non-human: only micro	Macro: 3.0 Micro: 3.3 Human: 2.5
023	Nemo	M	Prefer not to answer	Christian - Protestant	Low	Human: reject Non-human: only micro	Macro: 2.4 Micro: 4.6 Human: 3.1
026	Neptune	M	Hispanic, White	Agnostic	Low	Full Acceptance	Macro: 5.0 Micro: 5.0 Human: 4.5
027	Tethys	F	Hispanic	Christian - Catholic	Low	Full Acceptance	Macro: 3.9 Micro: 4.0 Human: 4.1

028	Marinus	M	NA	Hindu	Low	Full Acceptance	Macro: 4.1 Micro: 4.9 Human: 4.6
032	Azure	F	Black	Christian - Protestant	High	Human: undecided Non-human: higher taxa	Macro: 4.0 Micro: 4.9 Human: 3.6
035	Sapphira	F	Asian	Christian - Protestant	Low	Human: reject Non-human: higher taxa	Macro: 3.6 Micro: 4.1 Human: 1.8
036	Rio	NA	NA	NA	NA	Full Acceptance	Macro: 4.5 Micro: 4.6 Human: 4.1
037	Tiamat	F	White	Christian - Other (Nondenom.)	Med	Human: reject Non-human: micro & limited macro	Macro: 3.4 Micro: 4.0 Human: 2.3
038	Ariel	F	Asian	Christian - Protestant	Low	Human: reject. Non-human: higher taxa	Macro: 3.0 Micro: 4.9 Human: 3.0
039	Cyan	F	Hispanic, White	Christian - Catholic	Low	Human: undecided. Non-human: higher taxa.	Macro: 4.0 Micro: 4.8 Human: 4.0
GAENE 2.1							
ID	**Pseud.**	**Gender**	**Race/ Ethnicity**	**Religion**	**Evo. Ed.**	**Interview-based Acceptance**	**Composite Score**
010	Jocasta	F	White	Atheist	High	Full Acceptance	4.3
011	Helen	F	White	Agnostic	Med	Full Acceptance	4.7
012	Adrastea	F	Black	Christian - Protestant	High	Human: undecided Non-human: undecided	3.9
013	Alecto	F	Asian	Christian - Catholic	High	Full Acceptance	4.5
014	Andromache	F	Hispanic	Christian - Other (Nondenom.)	Med	Full Acceptance	4.3
015	Clio	F	Hispanic	Atheist	High	Full Acceptance	4.6
016	Paris		N/A	N/A		Full Acceptance	4.2
017	Cora	F	Hispanic, White	Christian - Other (Progressive)	High	Full Acceptance	4.9
018	Daphne	F	Asian	Hindu	Low	Full Acceptance	4.9
019	Achilles	M	Native	Christian - Protestant	Med	Human: reject Non-human: higher taxa	3.5
024	Adonis	M	White	Christian - Protestant	Low	Human: undecided	4.5

						Non-human: accept	
025	Penelope	F	Black	Christian - Protestant	Med	Human: undecided Non-human: accept	3.5
029	Agamemnon	M	White	Christian - Protestant	Low	Human: reject Non-human: only micro	3.8
030	Eudora	F	Black	Christian - Catholic	High	Human: undecided Non-human: higher taxa	4.6
031	Electra	F	White	Christian - Protestant	Low	Human: reject Non-human: higher taxa	4.2
033	Ajax	M	White	Christian - Catholic	Low	Human: reject Non-human: higher taxa	3.7
034	Ariadne	F	White	Other (Bahai)	High	Human: reject Non-human: higher taxa	3.5
GAENE 3.0 (new items only)							
ID	**Pseud.**	**Gender**	**Race/ Ethnicity**	**Religion**	**Evo. Ed.**	**Interview-based Acceptance**	**Composite Score**
040	Europa	F	White	Christian - Other (Nondenom.)	Med	Full Acceptance	4.4
041	Callisto	F	Black	Nothing in particular	High	Full Acceptance	4.1
042	Antigone	F	Asian, White	Agnostic	Med	Full Acceptance	3.2
043	Minos	M	Hispanic	Christian - Catholic	High	Human: accept Non-human: undecided	4.4
044	Dido	F	White	Christian - Catholic	Med	Full Acceptance	3.5
045	Circe	F	Hispanic, White	Christian - Other (Christian/Agnostic)	Med	Full Acceptance	4.5
046	Pandora	F	Asian, White	Agnostic	High	Full Acceptance	4.6
047	Theseus	M	Asian	Decline to state	Med	Full Acceptance	4.3
048	Clytemnestra	F	White	Agnostic	Low	Full Acceptance	4.3
049	Hecuba	F	White	Atheist	Med	Full Acceptance	4.6
050	Hector	M	Asian	Muslim	Med	Human: undecided	3.9

						Non-human: higher taxa	
051	Odysseus	M	White	Agnostic	Low	Full Acceptance	4.6
052	Alcyone	F	Black	Other faith (Ifa)	Low	Full Acceptance	3.7
053	Orpheus	M	White	Agnostic	Low	Full Acceptance	4.4
054	Damocles	M	Asian	Muslim	Low	Human: reject Non-human: higher taxa	3.2
055	Hippolyta	F	White	Christian - Protestant	High	Human: undecided Non-human: higher taxa	3.5
056	Leda	F	White	Atheist	Med	Full Acceptance	4.5
057	Eurydice	F	Asian	Agnostic	Med	Full Acceptance	2.7
058	Cassandra	F	White	Christian-Other (Southern Baptist)	Low	Human: reject Non-human: limited macro	3.7
059	Hermione	F	Black	Christian - Other (Nondenom.)	Low	Full Acceptance	4.6
060	Phaedra	NA	NA	NA	NA	Human: reject Non-human: accept	4.7

APPENDIX E

CODING RUBRIC FOR I-SEA AND GAENE COGNITIVE INTERVIEWS

This is the coding rubric used to analyze cognitive interviews on the I-SEA and the GAENE. It is the final rubric that was developed after all of the interviews were concluded. In it, each primary code from the pre-interview codebook is divided into several sub-codes. Note that some sub-codes are not discussed in the article because they arose in a relatively small number (<10%) of interviews. This codebook was developed using inductive methods, so each sub-code arose at least twice.

KNOWLEDGE ABOUT EVOLUTION
The student's answer is influenced by their factual understanding of evolutionary processes and/or evidence for evolution separate from their general acceptance of evolution. Code does not apply if misconceptions about evolution directly inform rejection of evolution.

> Low certainty: The student recognizes that they have limited knowledge about certain evolutionary facts (ex: origins of life) or certain lines of evidence (ex: fossils) and thus does not want to take a strong stand on a particular statement.
>
> Misconception: The student's answer is based on a misconception about how evolutionary processes work (ex: if the environment is stable, species stop evolving), or other misunderstandings of biology.
>
> Concept mix-up: The student confuses evolution for other natural processes, such as prenatal development or the central dogma. As a result, their answer reflects their views on the other biological process, and not necessarily their views on the evolutionary concept in question.

DEFINITION OF EVOLUTION
The item refers to evolution in general, but the student's answer is based on their exclusion of certain concepts or species from their interpretation of the term "evolution." (ex: student's answer indicates high acceptance because they have removed ideas that they disagree with from their idea of evolution).

> Micro: The student indicates that they're only thinking about microevolution when answering an item that refers to evolution as a whole, or to macroevolution in particular.
>
> Not Single Common Ancestor: Th students says they're including at least some macroevolution (ex: speciation) in their definition of "evolution," but they're not including the shared ancestry of higher taxa.
>
> Not Human: The student indicates that they're applying the concept of evolution only to non-human species when answering an item that refers to evolution as a whole, or to human evolution in particular.

Other: The student has another non-standard interpretation of "evolution" that influences how they answer (ex: they interpret "evolution" and "theory of evolution" as referring to different concepts).

Artificial: The student conceptualizes artificial selection as a process that is related to, but separate from, evolution.

UNDERSTANDING NATURE OF SCIENCE

The student's answer is influenced by their factual understanding of NOS separate from their overall acceptance of evolution. Code does not apply if misconceptions about NOS directly inform rejection of evolution.

Misconception: The student has a misconception about how science works, which influences their answer.

Example: The student claims to accept all aspects of evolution, but answers an item based on the idea that it's "just a theory," or says that when more evidence accumulates, evolution will be referred to as a fact and not as a theory.

Evidence Types: The student points out that the form of evidence referred to in the item is one relevant form of evidence, but that other forms of evidence are also necessary. This is an issue if the student says they accept the relevant evolutionary idea, but don't strongly agree/disagree because the mentioned evidence isn't the *only* evidence.

Fact vs. Theory: The student points out that "theory" & "fact" are different concepts in science, and evolution is a theory. (This mainly applies to GAENE #12)

Note: If a student says that evolution is "just" a theory and that it can become fact with more evidence, code response as "NOS: misconception."

Plausible: The student's answer is based on acknowledging that evolution is a scientifically plausible theory, even when they personally do not fully accept it.

WORDING

The student's answer is influenced by confusion over the wording of an item. Code does not apply if the student doesn't understand a scientific concept essential to the item (ex: unaware of how to establish "scientific validity").

Definition: The student is confused about the meaning of a particular word or phrase, and struggles to answer because they don't know what the item is referring to.

Strong Language: The student says that they are picking a more moderate answer than they would otherwise because the phrasing is too absolute or extreme (ex: because something is described as "dramatic").

Confusing Phrasing: The student struggles to answer because they are confused by phrasing that contrasts terms with potentially similar meanings (ex: adapt vs. evolve vs. change), OR the whole item is phrased in such a way that it's not clear what it means to "agree" or "disagree;" OR the student finds the item's choice of concepts irrelevant, or otherwise unsuitable. In short, the concepts are unclear.

Complex: The wording is generally hard to decipher; it takes the students a few tries to make sense of the item as a whole. In short, the wording as a whole is unclear.

Partial: The student selected "agree" or "strongly agree" for GAENE #3 because they agree that some parts of evolution are true, and some are false.

PERSONAL CONTEXT

The item refers to actions that may be taken on the basis of acceptance or rejection of evolution, and the student's answer is influenced by considering interpersonal dynamics in addition to their own views on evolution.

Not Insistent: The student says that they personally accept evolution but ensuring that other people accept and/or understand evolution is not a priority for them.

Comfort: The student says that their willingness to *discuss evolution with others* depends on whether the social context is welcoming or unwelcoming of their views, as a separate matter from the content of their personal views (ex: the students says they accept evolution, but are reluctant to argue with a hostile audience).

Whole Life: The student says that they find evolution to be broadly relevant to science & society, but it does not play a very large part in their personal life as a whole.

Betting: The student says that they're very sure that evolution is true, but they're disinclined to bet their life on anything.

No Change: For GAENE 3.0 (Item 6), the student says that understanding evolution hasn't changed their life because they never consciously rejected evolution, so accepting/understanding evolution did not constitute a change in their views.

SCIENTIFIC CONTEXT

Proximate: The student answers an item based on picturing evolution as a module in a biology course, and not as a major scientific theory (ex: the student says that they are able to learn about the proximate explanations for processes such as photosynthesis without knowing their evolutionary origins).

General Science: Student states that evolution is very important, but other major scientific theories (ex: general relativity) are equally important for science and/or society.

APPENDIX F

SUPPLEMENTAL TABLE 2. PROFILES OF INTERVIEW PARTICIPANTS ON THE ORIGINAL MATE AND MATE 2.0.

Original MATE							
ID	**Pseud.**	**Gender**	**Race/Ethnicity**	**Religion**	**Evolution Education**	**Interview-based Acceptance**	**Average Score on MATE items**
1	Acacia	F	Asian	Atheist	High	Full Acceptance	4.6
2	Ash	M	Asian	Agnostic	High	Full Acceptance	4.6
3	Daisy	F	Hispanic or Latinx	Agnostic	High	Full Acceptance	4.4
4	Aspen	F	White	Agnostic	Medium	Full Acceptance	4.6
5	Azalea	F	Asian	Hindu	High	Full Acceptance	4.4
6	Bryony	F	White	Nothing in particular	High	Full Acceptance	4.4
7	Berry	M	White	Atheist	Medium	Full Acceptance	4.2
8	Clover	F	Hispanic or Latinx	Christian – Catholic	High	Full Acceptance	4.4
9	Coral	F	White	Christian – Catholic	Medium	Full Acceptance	3.9
10	Daffodil	F	Asian	Muslim	High	Full Acceptance	4.6
11	Dahlia	F	White	Agnostic	Medium	Full Acceptance	5.0
12	Amber	F	White	Agnostic	High	Full Acceptance	4.5
13	Eartha	F	White	Agnostic	Medium	Full Acceptance	4.6
14	Fern	F	White	Agnostic	Medium	Full Acceptance	4.6
15	Fleur	F	White	Christian – Greek Orthodox	High	Full Acceptance	4.8
16	Basil	M	Asian	Other – Jainist	High	Full Acceptance	4.6
17	Ginger	F	Hispanic or Latinx	Christian – Protestant	High	Human Exception	3.8
18	Hazel	F	White	Agnostic	High	Full Acceptance	5.0
19	Heather	F	White	Agnostic	Medium	Full Acceptance	4.9
20	Holly	F	White	Agnostic	Medium	Full Acceptance	4.7
21	Iris	F	Hispanic or Latinx; White	Christian – Protestant	High	Creation of Higher Taxa	4.4
22	Laverne	F	Hispanic or Latinx; White	Agnostic	Medium	Full Acceptance	4.6
23	Jasmine	F	White	Christian – undecided	High	Rejection	3.8
24	Lake	M	White	Atheist	High	Full Acceptance	4.9
25	Juniper	F	Asian	Atheist	Medium	Full Acceptance	4.3
26	Ivy	F	White	Atheist	Medium	Full Acceptance	4.8
27	Liana	F	White	Agnostic	Medium	Full Acceptance	4.8

28	Cedar	M	Asian	Nothing in particular	High	Undecided	4.1
29	Dale	M	Hispanic or Latinx	Christian – Catholic	High	Full Acceptance	4.4
30	Lilac	F	White	Christian – Catholic	Medium	Full Acceptance	4.3
31	Lily	F	White	Agnostic	High	Full Acceptance	4.8
32	Lavender	F	White	Agnostic	Medium	Full Acceptance	4.9
33	Linden	M	Hispanic or Latinx	Nothing in particular	High	Full Acceptance	4.6
34	Heath	M	Other: Middle Eastern	Decline to state	Medium	Full Acceptance	4.5
35	Reed	M	Native American	Agnostic	High	Full Acceptance	4.2
36	Magnolia	F	Asian	Hindu	Medium	Full Acceptance	4.5
37	Marigold	F	Black	Christian - Protestant	None	Undecided	3.6
38	Myrtle	F	Asian	Agnostic	Low	Creation of Higher Taxa	4.3
39	Haywood	M	Hispanic or Latinx	Agnostic	Low	Full Acceptance	5.0
40	Pansy	F	White	Nothing in particular	Low	Full Acceptance	4.9
41	Jasper	M	Asian	Agnostic	Low	Full Acceptance	4.9
42	Petunia	F	White	Other - spiritual	Low	Full Acceptance	4.8
43	Forrest	M	White	Agnostic	Low	Full Acceptance	3.7
44	Poppy	F	Asian	Buddhist	Low	Full Acceptance	4.7
45	Primrose	F	Hispanic or Latinx	Christian - nondenomination al	Low	Creation of Higher Taxa	4.4
46	Rosemary	F	Black	Agnostic	Low	Human Exception	4.1
47	Sage	F	Asian; White	Atheist	Low	Full Acceptance	4.4
48	Briar	M	White	Christian – nondenomination al	None	Full Acceptance	4.7
49	Moss	M	Asian	Nothing in particular	None	Full Acceptance	4.7
50	Savannah	F	White	Agnostic	None	Human Exception	3.9
51	Oliver	M	Asian; Pacific Islander	Christian – Catholic	None	Full Acceptance	4.4
52	Herb	M	White	Christian - Protestant	Low	Rejection	1.4

53	Violet	F	Black	Christian - Baptist	Low	Rejection	2.9
54	Willow	F	White	Jewish	None	Full Acceptance	4.1
55	Zinnia	F	Asian	Buddhist	None	Full Acceptance	4.4
56	Robin	M	Asian	Hindu	Low	Full Acceptance	4.4
57	Rowan	M	Hispanic or Latinx	Christian - Catholic	None	Rejection	3.6
58	Silvester	M	Asian	Other - Sikhism	Low	Full Acceptance	4.1
59	Dove	F	Hispanic or Latinx	Christian - nondenomination al	Low	Rejection	4.0
60	Lark	F	White	Christian – Latter-Day Saints	Low	Undecided	4.8
61	Raven	F	Hispanic or Latinx	Christian - Protestant	Low	Creation of Higher Taxa	4.0
62	Wren	F	White	Christian - Lutheran	Low	Full Acceptance	4.5
MATE 2.0							
ID	**Pseud.**	**Gender**	**Race/Ethnicity**	**Religion**	**Evolution Education**	**Interview-based Acceptance**	**Average Score on MATE items**
010	Bjork	F	White	Atheist	High	Full Acceptance	4.6
011	Hadas	F	White	Agnostic	Medium	Full Acceptance	4.6
012	Iva	F	Black	Christian - Protestant	High	Undecided	3.3
013	Sawda	F	Asian	Christian - Catholic	High	Full Acceptance	4.8
014	Kalina	F	Hispanic	Christian - nondenomination al	Medium	Full Acceptance	4.8
015	Liepa	F	Hispanic	Atheist	High	Full Acceptance	4.9
016	Palmer	N/A	N/A	N/A	N/A	Full Acceptance	4.9
017	Melia	F	Hispanic or Latinx; White	Christian – Progressive Christian	High	Full Acceptance	4.9
018	Ornella	F	Asian	Hindu	Low	Full Acceptance	4.9
019	Alon	M	Native	Christian - Protestant	High	Creation of Higher Taxa	3.9
024	Aritz	M	White	Christian - Protestant	Low	Human Exception	4.4
025	Pomona	F	Black	Christian - Protestant	Medium	Undecided	3.8

040	Pihla	F	White	Christian - nondenomination al	High	Rejection	2.4
041	Randa	F	Hispanic or Latinx	Christian - Latter-Day Saints	High	Creation of Higher Taxa	2.2
042	Jelena	F	Black	Christian – Church of Christ	Low	Creation of Higher Taxa	4.3
043	Boris	M	Prefer not to answer	Christian - Protestant	Low	Rejection	2.6
044	Taimi	F	White	Christian - nondenomination al	Medium	Creation of Higher Taxa	3.1
045	Anargul	F	Asian	Muslim	Medium	Creation of Higher Taxa	3.6
046	Blodwen	F	White	Spiritual	Medium	Full Acceptance	4.8
047	Anthea	F	White	Agnostic	Low	Full Acceptance	4.3
048	Hanako	F	Black	Christian - Protestant	Low	Creation of Higher Taxa	4.8
049	Elon	M	White	Buddhist	Low	Full Acceptance	3.9
050	Tomer	M	Black	Nothing in particular	Low	Full Acceptance	4.7
051	Ione	F	White	Christian – Latter-Day Saints	Low	Creation of Higher Taxa	2.9
052	Laleh	F	Black	Christian – non-Orthodox	None	Creation of Higher Taxa	4.8
053	Vipin	M	Asian	Hindu	Low	Full Acceptance	4.7
054	Yasen	M	Black; Hispanic or Latinx	Christian - Catholic	Low	Full Acceptance	3.8
055	Leilani	F	Black; Hispanic or Latinx	Christian - Lutheran	None	Creation of Higher Taxa	2.8
056	Lys	F	Hispanic or Latinx	Christian - Catholic	None	Full Acceptance	4.7

APPENDIX G

CODING RUBRIC FOR MATE REVISION COGNITIVE INTERVIEWS

This is the coding rubric used to analyze cognitive interviews with the original MATE and MATE 2.0. It is the final rubric that was developed after all of the MATE 1.0 interviews were concluded; it was used for the MATE 2.0 interviews as well. In it, each primary code from the pre-interview codebook is divided into several sub-codes. Note that some sub-codes are not discussed in the article because they arose in a relatively small number (<10%) of interviews. This codebook was developed using inductive methods, so each sub-code arose at least twice.

DEFINITION OF EVOLUTION

Species Specific: Student's answer depends on whether evolution is applied to humans.

Microevolution: Student defines "evolution" as evolutionary processes. Code is applicable if a student's answer is based only on microevolution, or if they say that their answer would depend on whether macroevolution is part of the evolution definition.

Socially Modern: Student interprets the word "modern" in an item as a reference to cultural or technological modernity, rather than anatomical modernity.
Extra Theories: Student includes concepts that are not actually part of evolutionary theory in their definition of evolution (e.g., origins of life, the Big Bang).

UNDERSTANDING NATURE OF SCIENCE

Tentative Nature: Student states that all scientific theories are falsifiable, which in itself is not a misconception. Rather, this code applies when a student emphasizes falsifiability to the point of avoiding answers of "strongly agree" and "strongly disagree." In their explanation, the student may mention either of the following:

- Current evidence supports evolution, but evidence against evolution could theoretically be found in the future.
- Current evidence supports evolution, but *more* supporting evidence will be found in the future.

Scientific Testing: Student states that some aspects of evolution cannot be tested. Their answer is based on a misconception about what counts as scientific testing. Code does not apply if the student says that they are generally unaware of how evolution can be tested. This trend takes four main forms:

- Evolution cannot be tested because we cannot go back in time to observe extinct species.
- Evolution cannot be tested because we as individuals cannot see one species evolve into another.

- Evolution cannot be tested because the only way to test a hypothesis is through a controlled experiment. Observations do not count as scientific testing.
- A scientific prediction is a prediction of what will happen *in the future* in the natural world. Predictions are not made about present-day processes or past events (i.e., evolutionary history).

NOS Unaware: Student is broadly unaware of some aspect of the nature of science (NOS) and acknowledges their lack of knowledge. Code takes two main forms:

- Student does not know how evolution can be tested.
- Student does not know what counts as scientific validity.

Speculation: Student states that science always involves some speculation. They equate "speculation" with generating new hypotheses.

Final Answer: Student states that evolution is not well supported until everything about evolution is discovered. In other words, a theory is not fully valid if scientists are still generating and testing new hypotheses.

Just A Theory: Student states that evolution is not fully supported by the evidence because it is "just a theory" that has not been declared a scientific fact. Unlike "Final Answer," this code applies only when the student clearly has a misconception about fact vs. theory in science.

Factual: Student is confused by the idea of factual vs. non-factual data. Their answer may be influenced by the idea that *factual* data must be 100% correct. Code is mainly applicable to Item 16 of MATE 1.0.

KNOWLEDGE ABOUT EVOLUTION

Unfamiliar Data: Student states that they do not know evolution-related data/evidence well enough to give a decisive answer. This code is similar to "NOS Unaware." Differences:

- DO NOT use this code if they do not know how evolution could be tested. "NOS Unaware" applies.
- DO NOT use this code if a student says that they have not made up their mind about evolution because they do not know whether there is evidence to support the theory. This is an accurate measure of uncertainty.

Earth Age: Student accepts the idea that the earth is old (i.e., millions or billions or years), but they are factually unaware of whether it is over 4 billion years old.

Counts as Evolution: Student's answer is affected by the misconception that evolution necessarily involves a "progression" from less complex to more complex, meaning that that which does not superficially change does not evolve. Example: Early life was unicellular; this means that humans have evolved, but bacteria have not evolved. Code is mainly applicable to Item 9 of MATE 1.0.

KNOWLEDGE ABOUT SCIENTIST VIEWS

Rejecting Scientists: Student's answer is informed by the impression that some scientists do not fully accept evolution. This code applies if the student says that not all scientists accept evolution, or if they say that they don't know whether most scientists accept evolution (which implies that some may not). Code is applicable to Items 5 & 17 of MATE 1.0.

Accepting Scientists: Student's answer is informed by the impression that a majority of scientists do fully accept evolution. Code is applicable to Items 5 & 17 of MATE 1.0.

General Public: Student bases their answer on what the non-scientist public think about evolution, rather than on their own views. Example: The student says that the evidence for evolution is unconvincing/unclear because *other people* reject evolution; if there was no fault in the data, everyone would accept evolution. Student's answer may or may not reflect their personal view.

CHRISTIAN ASSUMPTION

Other Religion: Code is applicable only to Item 14 of MATE 1.0. Student says that they would answer this item differently if it said "my religion's account of creation" instead of "the Biblical account of creation" because they follow a religion other than Christianity.

Non-Christian Background: Code is applicable only to Item 14 of MATE 1.0. Student says that they are *unfamiliar* with the Biblical account of creation because they do not come from a Christian background. Code is applicable for both non-religious students and students who follow a religion other than Christianity. Code does not necessarily apply to every non-Christian religious student.

New Earth: Student says that they believe in a Young Earth but are not committed to the Earth being less than 20,000 years (i.e., it could be 25,000 years old). This influences their answer. Code is mainly applicable to Item 7 of MATE 1.0.

WORDING

Millions: Code is applicable to Items 1 and 3 of MATE 1.0. Student interprets "millions of years" to be a reference to how long individual species have existed on Earth.

- Example: Student interprets Item 3 as saying, "Homo sapiens as a single species have existed for millions of years."

Organisms: Code is applicable to Items 1, 9, 18, and 19 of MATE 1.0. Student displays confusion about the terms "organisms" or "living forms."

Who Scientist: Code is applicable to Items 5 and 17 of MATE 1.0. Student states that either a) they are uncertain about who counts as a scientist, or b) their answer would depend on how "scientist" is defined. Example: They might draw a distinction between biologists vs. scientists as a whole.

Much/Most: Code is applicable to Items 5 and 17 of MATE 1.0. Student states that they find the terms "much" or "most" to be vague, which makes it difficult for them to answer the item.

Data Clarity: Code is mainly applicable to Item 6 of MATE 1.0. Student interprets the item as asking, "Have you ever seen evolutionary data that is confusing?" They do *not* interpret it as, "Is evolution supported by data?"

Species Definition: Code is applicable to Items 9 and 15 of MATE 1.0. Student interprets "organism" or "human" as referring to a single species. They give the reasoning that the "form" of a species does not radically change as long as it is still the same species. When the form of a species changes greatly, it becomes a new species.

Conserved Traits: Code is applicable to Item 9 of MATE 1.0. Student points out that certain biological features (e.g., ATP, ribosomes) have been highly conserved throughout evolutionary history. Their answer is affected by the idea that life exists in "essentially the same form" at the biomolecular level.

Sound: Code is applicable to Item 12 of MATE 1.0. Student displays confusion about the term "sound."

Current Evolution: Code is applicable to Item 12 of MATE 1.0. Student displays confusion about the term "current evolutionary theory." This may take the form of the student trying and struggling to draw a distinction between *current* vs. *old* evolutionary theory.

Characteristics of Life: Code is applicable to Item 13 of MATE 1.0. Student displays confusion about the term "characteristics of life."

Testable Predictions: Code is applicable to Item 13 of MATE 1.0. Student does not know the literal meaning of the term “testable predictions.”

Respect: Code is applicable to Item 13 of MATE 1.0. Student displays confusion about the term “with respect to.” Some misinterpret it as “acts kindly towards.”

Historical: Code is applicable to Item 16 of MATE 1.0. Student displays confusion about the term “historical,” and this impacts their answer.

Doubt: Code is applicable to Item 17 of MATE 1.0. Student interprets “doubt” to mean that there is debate within the scientific community about specific evolutionary hypotheses. They do not interpret “doubt” to mean questioning whether evolution occurs at all.

Brings Meaning: Code is applicable to Item 18 of MATE 1.0. Student displays confusion about the term “brings meaning.” They may point out that there are two alternative definitions: “explains” or “provides philosophical value or purpose.”

Opposite: Code is applicable to Item 19 of MATE 1.0. Student interprets this item as saying that all of life on earth descended from one common ancestor, which by definition was alive at one point in time. This is the exact opposite of what the survey authors intended.

Exceptions: Code is applicable to Item 19 of MATE 1.0. Student displays confusion about the term “with few exceptions.” They may ask something like, “What sort of exceptions?”

General Confusion: Student has no idea what the item as a whole is attempting to say and cannot pinpoint one word or phrase that is confusing. Code is applicable to all items.

APPENDIX H

SUPPLEMENTAL TABLE 3. PARTICIPANTS' MOST FREQUENT USES OF CONSTRUCTS OTHER THAN PERSONAL ACCEPTANCE OF EVOLUTION WHEN ANSWERING ITEMS ON THE MATE.

MATE Item	**Theme 1: Code (Subcode)**	**N**	**Theme 2: Code (Subcode)**	**N**
1. Organisms existing today are the result of evolutionary processes that have occurred over millions of years.				
2. The theory of evolution is incapable of being scientifically tested.	Understanding of NOS (Scientific Testing)	13/62		
3. Modern humans are the product of evolutionary processes that have occurred over millions of years.				
4. The theory of evolution is based on speculation and not valid scientific observation and testing.				
5. Most scientists accept evolutionary theory to be a scientifically valid theory.	Perception of Scientists' Views (Rejecting Scientists)	31/47*	Perception of Scientists' Views (Accepting Scientists)	6/15**
6. The available data are ambiguous (unclear) as to whether evolution actually occurs.	Knowledge About Evolution (Unfamiliar Data)	9/62		
7. The age of the earth is less than 20,000 years.				
8. There is a significant body of data that supports evolutionary theory.	Knowledge About Evolution (Unfamiliar Data)	9/62		
9. Organisms exist today in essentially the same form in which they always have.				
10. Evolution is not a scientifically valid theory.				
11. The age of the earth is at least 4 billion years.	Knowledge About Evolution (Age of Earth)	30/62		
12. Current evolutionary theory is the result of sound scientific research and methodology.				

13. Evolutionary theory generates testable predictions with respect to the characteristics of life.	Understanding of NOS (Scientific Testing)	22/62	Wording (Characteristics of Life)	21/62
14. The theory of evolution cannot be correct since it disagrees with the Biblical account of creation.	Christianity is Assumed	5/9***		
15. Humans exist today in essentially the same form in which they always have.	Wording (Species Definition)	9/62		
16. Evolutionary theory is supported by factual historical and laboratory data.				
17. Much of the scientific community doubts if evolution occurs.	Perception of Scientists' Views (Rejecting Scientists)	18/47*	Perception of Scientists' Views (Accepting Scientists)	9/15**
18. The theory of evolution brings meaning to the diverse characteristics and behaviors observed in living forms.	Wording (Brings Meaning)	18/62		
19. With few exceptions, organisms on earth came into existence at about the same time.	Wording (Opposite Interpretation)	11/62		
20. Evolution is a scientifically valid theory.				

*Out of 47 who fully accept evolution. **Out of 15 who do not fully accept evolution. ***Out of 9 who are affiliated with a non-Christian religion.

APPENDIX I

CODING RUBRIC FOR FREE RESPONSE INTERVIEW QUESTIONS

This is the coding rubric for the own-views interview questions which were presented after the cognitive interview with the original MATE. This portion of the interview was coded holistically; one of the five codes listed below was assigned based on a student's full set of answers in response to all of the interview questions. Note that the "Rejection" code allows for acceptance of microevolution, since not a single student rejected evolutionary change within species.

FULL ACCEPTANCE

Student expresses the following views throughout the own-views interview:

- Humans share a common ancestry with other animals, such as primates.
- New species arise through the process of speciation.
- All species descended from single-celled ancestors.

HUMAN EXCEPTION

Student expresses the following views throughout the own-views interview:

- New species arise through the process of speciation.
- All <u>non-human</u> species descended from single-celled ancestors.
- Humans do not share a common ancestry with other animals.

CREATION OF HIGHER TAXA

Student expresses the following views throughout the own-views interview:

- God created a relatively small number of "original ancestors" that do not share a common ancestry with each other.
- Evolutionary processes have caused the "original ancestors" to diversify into the species we see today.
- Examples of independent evolutionary trees that students might mention:
    - Primates, carnivores, etc.
    - Mammals, reptiles, fish, insects, etc.
- Humans do not share a common ancestry with other animals. The student may say that humans were created in their present form OR they may say that God created hominins separately from primates, and the hominins underwent their own internal evolution (i.e., Neandertals existed as a separate population/subspecies).

REJECTION

Student expresses the following views throughout the own-views interview:

- God created a relatively large number of initial species that resembles the diversity of life that we see today.
- God created humans in more-or-less their present form.

UNDECIDED

- Do use this code if the student says that they have not made up their mind about one or more major aspects of evolution that are *necessary for choosing between the other codes.* Examples:
  - Undecided about whether humans share a common ancestry with other animals, or if hominins were created separately
  - Undecided about whether life evolved by natural processes alone, or if God created several major taxa separately
- Do use this code if the student says that they are oscillating between accepting the scientifically accurate version of evolution vs. believing a creationist account that is not consistent with the scientific evidence (e.g., creation of higher taxa or rejection).

Do <u>not</u> use if the student is factually uncertain about whether plants and animals share a common ancestry without bringing God into the explanation.

APPENDIX J

ADDITIONAL TABLES AND FIGURES FOR MATE 2.0 VALIDATION

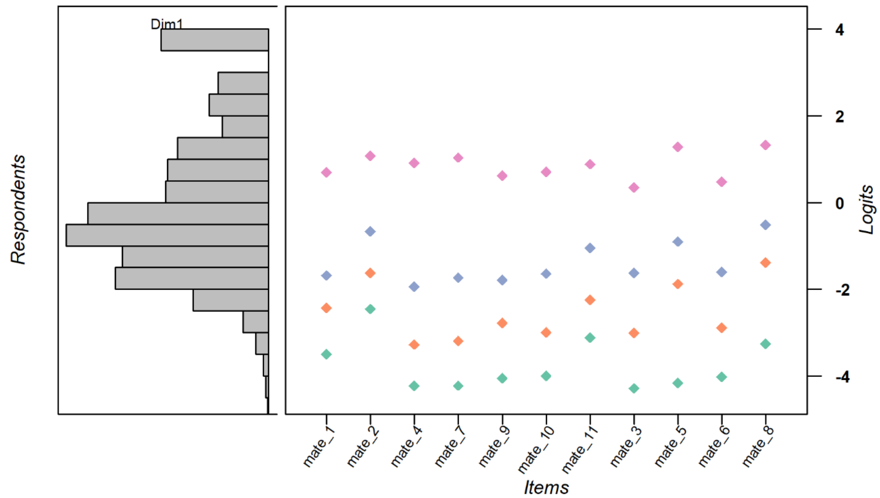
Supplemental Table 4. Unweighted and Weighted mean squares item fit statistics (equal to outfit and infit MNSQ respectively) for the unidimensional partial credit Rasch model for MATE 2.0. Values of 0.7-1.3 are considered to indicate good fit. Values outside of this range are underlined.

Item	Outfit	Infit
mate_1	0.97	0.97
mate_2	1.15	1.13
mate_3	0.83	0.90
mate_4	0.84	0.86
mate_5	1.43	1.37
mate_6	0.82	0.89
mate_7	1.12	1.10
mate_8	1.39	1.33
mate_9	0.94	0.94
mate_10	0.79	0.81
mate_11	0.99	0.99

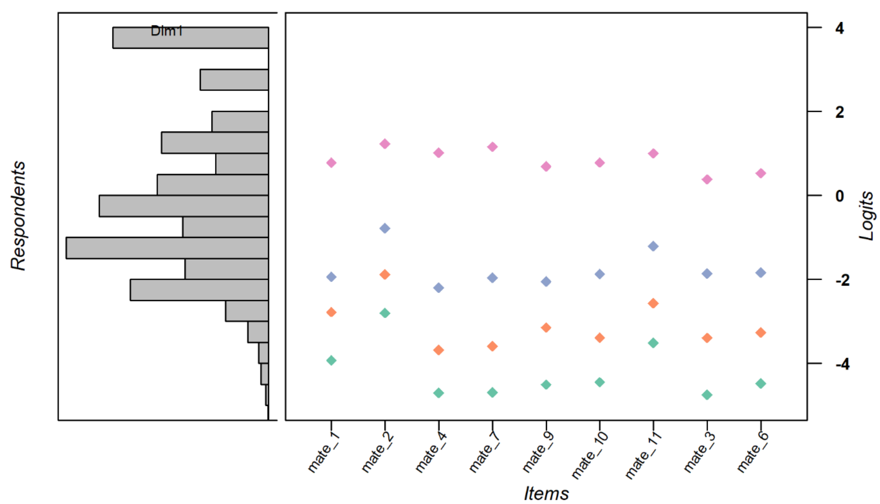
Supplemental Table 5. Unweighted and Weighted mean squares item fit statistics (equal to outfit and infit MNSQ respectively) for the unidimensional partial credit Rasch model for MATE 2.0 without Items 5 and 8. Values of 0.7-1.3 are considered to indicate good fit.

Item	Outfit	Infit
mate_1	1.02	1.04
mate_2	1.30	1.28
mate_3	0.92	1.00
mate_4	0.86	0.89
mate_6	0.94	1.00
mate_7	1.16	1.16
mate_9	0.98	0.98
mate_10	0.78	0.83
mate_11	1.04	1.06

Supplemental Figure 1. Wright map of MATE 2.0 data. The data points on the right represent item difficulties and the histogram on the left shows the distribution of person abilities. Higher points and higher respondents indicate more difficult items, i.e. high evolution acceptance. Colors indicate various points on the Likert scale: green = "disagree," orange = "neutral," blue = "agree," pink = "strongly agree." Comparison of the histogram with the item difficulties shows that Rasch item difficulties are below most person abilities, indicating that most students in our sample were accepting of evolution.



Supplemental Figure 2. Wright map of MATE 2.0 data without Items 5 and 8. The data points on the right represent item difficulties and the histogram on the left shows the distribution of person abilities. Higher points and higher respondents indicate more difficult items, i.e. high evolution acceptance. Colors indicate various points on the Likert scale: green = "disagree," orange = "neutral," blue = "agree," pink = "strongly agree." Comparison of the histogram with the item difficulties shows that Rasch item difficulties are below most person abilities, indicating that most students in our sample were accepting of evolution.



APPENDIX K

CO-AUTHOR PERMISSIONS

Chapter 4 of this dissertation has been published as a research article in the journal Cell Biology Education – Life Sciences Education (doi: 10.1187/cbe.21-05-0127). My committee member Dr. Elizabeth Barns is co-first author along with myself; my committee chair Dr. Sara Brownell, Dr. Supriya, and Dr. Michael Rutledge are also co-authors on this publication. All co-authors have approved the use of this publication in my dissertation.

APPENDIX L

IRB APPROVAL DOCUMENTS

EXEMPTION
GRANTED

Dear Sara Brownell:

On 10/16/2019 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	A Qualitative Assessment of the Measure of Acceptance of the Theory of Evolution (MATE) Instrument
Investigator:	Sara Brownell
IRB ID:	STUDY00010903
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• Recruitment form_10.15.2019.pdf, Category: Recruitment Materials; • Protocol_10.15.2019.docx, Category: IRB Protocol; • Consent form_10.15.2019.pdf, Category: Consent Form; • MATE project_MATE instrument.pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • MATE project_Demographic form.pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • MATE project_Interview questions.pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions);

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (2) Tests, surveys, interviews, or observation on 10/16/2019. In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,
IRB Administrator

cc:Anastasia Misheva
Sara Brownell
Maryann Barnes

EXEMPTION
GRANTED

Dear Sara Brownell:

On 10/3/2019 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Test of Validated Assessment in Evolutionary Medicine
Investigator:	Sara Brownell
IRB ID:	STUDY00010655
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• EvMedTest_Assessment, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • EvMedTest_Demographic, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • EvMedTest_Consent, Category: Consent Form; • EvMedTest_IRB, Category: IRB Protocol;

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (1) Educational settings, (2) Tests, surveys, interviews, or observation on 10/3/2019.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

IRB Administrator

cc:
Anastasia Misheva
Sara Brownell
Daniel Grunspan