A Revised Measure of Acceptance of the Theory of Evolution: Introducing the MATE 2.0

M. Elizabeth Barnes,[†‡*] Taya Misheva,[‡§] K. Supriya,[§‖] Michael Rutledge,[¶] and Sara E. Brownell[§]

[†]Social Perceptions of Science Lab, Department of Biology, Middle Tennessee State University, Murfreesboro, TN 37132; [§]Biology Education Research Lab, Research for Inclusive STEM Education Center, School of Life Sciences, Arizona State University, Tempe, AZ 85282; [‖]Center for Education Innovation and Learning in the Sciences, University of California, Los Angeles, Los Angeles, CA 90095; [¶]Department of Biology, Middle Tennessee State University, Murfreesboro, TN 37132

ABSTRACT

Hundreds of articles have explored the extent to which individuals accept evolution, and the Measure of Acceptance of the Theory of Evolution (MATE) is the most often used survey. However, research indicates the MATE has limitations, and it has not been updated since its creation more than 20 years ago. In this study, we revised the MATE using information from cognitive interviews with 62 students that revealed response process errors with the original instrument. We found that students answered items on the MATE based on constructs other than their acceptance of evolution, which led to answer choices that did not fully align with their actual acceptance. Students answered items based on their understanding of evolution and the nature of science and different definitions of evolution. We revised items on the MATE, conducted 29 cognitive interviews on the revised version, and administered it to 2881 students in 22 classes. We provide response process validity evidence for the new measure through cognitive interviews with students, structural validity through a Rasch dimensionality analysis, and concurrent validity evidence through correlations with other measures of evolution acceptance. Researchers can now measure student evolution acceptance using this new version of the survey, which we have called the MATE 2.0.

INTRODUCTION

Researchers have conducted hundreds of studies over the past 30 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 (Nadelson and Southerland, 2012; Short and Hawley, 2012; Pew, 2013; Smith *et al.*, 2016; Gallup, 2019; Glaze *et al.*, 2020), which makes it difficult to compare research findings that conflict. 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 and Warden, 1999), may have limitations that could be causing confusion about how to increase evolution acceptance (Lloyd-Strovas and Bernal, 2012; Sickel and Friedrichsen, 2013; Glaze and Goldston, 2015; Barnes *et al.*, 2019). However, after 20 years of MATE 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 among a population of undergraduate biology students so that we could present a revised instrument with validity evidence.

William Romine, *Monitoring Editor*

Submitted May 18, 2021; Revised Nov 15, 2021; Accepted Nov 17, 2021

CBE Life Sci Educ March 1, 2022 21:ar10

DOI:10.1187/cbe.21-05-0127

[†]These authors contributed equally to this work.

*Address correspondence to: M. Elizabeth Barnes (liz.barnes@mtsu.edu).

© 2022 M. E. Barnes, T. Misheva, *et al.* CBE—Life Sciences Education © 2022 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 4.0 Unported Creative Commons License (http://creativecommons.org/licenses/by-nc-sa/4.0).

"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

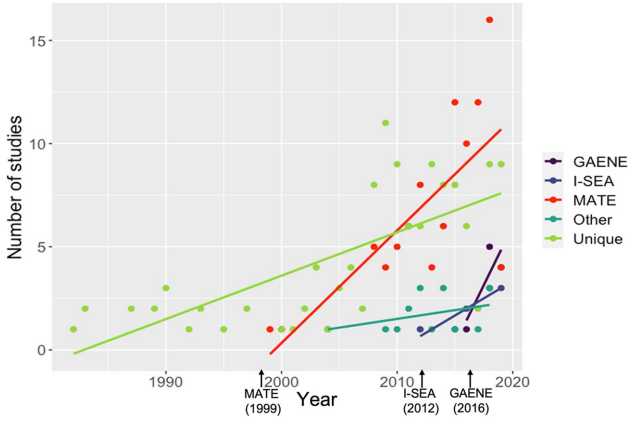

FIGURE 1. The use of instruments for measuring acceptance of evolution in peer-reviewed studies. Marked time points are the publications of the MATE (Rutledge and Warden, 1999), the I-SEA (Nadelson and Southerland, 2012), and the GAENE (Smith *et al.*, 2016). Measures labeled "other" include the Evolutionary Attitudes and Literacy Survey (EALS) and sociological polls such as the Gallup and the Pew.

BACKGROUND

Acceptance of Evolution Survey Tools

Before 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 five-point Likert scale (Koevering and 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 five-point Likert scale; it was the first measure of evolution acceptance that had substantial validation evidence (Rutledge and 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 and 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 *et al.*, 2020). Figure 1 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 in future studies (Romine *et al.*, 2016; Rutledge and Warden, 1999). Further, multiple researchers in the field have voiced concerns about the limitations of this instrument (Smith, 2009b; Nadelson and Southerland, 2012; Sickel and Friedrichsen, 2013; Wagler and Wagler, 2013; Romine *et al.*, 2018; Barnes *et al.*, 2019). 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 (NOS), 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 it may need to be revised for use among populations of undergraduate biology students.

Researchers have often questioned whether the MATE measures student conceptions that are not acceptance of evolution. One concern has been that the MATE conflates understanding of the 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—acceptance of the idea that evolution occurs (Smith, 2009a). 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 and 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 and Southerland, 2012). For example, the MATE item "The theory of evolution is incapable of being scientifically tested" requires the survey respondents to picture their own definitions 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 *et al.*, 2019, 2020a,b).

A recent study from our research group revealed that using different surveys to measure evolution acceptance with the same population of students can lead to different research results (Barnes *et al.*, 2019); for example, the I-SEA (Nadelson and Southerland, 2012), the GAENE (Smith *et al.*, 2016), the MATE (Rutledge and Warden, 1999; Rutledge and Sadler, 2007), and others. While prior studies have taken quantitative approaches to examine the dimensionality of the MATE (Wagler and Wagler, 2013; Romine *et al.*, 2016, 2018) and how results from the MATE compare with those from other evolution acceptance instruments (Metzger *et al.*, 2018; Barnes *et al.*, 2019), 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, we explored potential weaknesses of the MATE through student cognitive interviews to illuminate potential response process errors such as conflation of evolution acceptance with other constructs (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 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 principal 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 and Warden, 1999). The authors also showed that the measure was internally consistent (Rutledge and Warden, 1999) and then later showed evidence of test–retest reliability among university students (Rutledge and Sadler, 2007). In future studies, researchers found evidence that the MATE might be better analyzed as a multidimensional instrument (Wagler and Wagler, 2013; Romine *et al.*, 2016, 2018), 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 in analyses (Metzger *et al.*, 2018; Romine *et al.*, 2018; Sbeglia and Nehm, 2018; Barnes *et al.*, 2019), indicating evidence for concurrent validity of the MATE with other measures of evolution acceptance (American Educational Research Association *et al.*, 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 and response process errors identified when a participant responds to an item for reasons other than what is intended by the researchers and can indicate that an item is measuring extraneous information other than the targeted construct (American Educational Research Association *et al.*, 2014; Mead *et al.*, 2019). 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; Willis, 2004; García, 2011). Students taking the survey may be the most direct source of this 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, we 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). We designed the interviews to be semistructured so that we 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, we 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 (e.g., "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 we took into account these discussions while constructing this definition (Smith and Siegel, 2004, 2016; Smith, 2009a,b). Also, the original authors of the MATE did not specify whether their definition of evolution was that of microevolution, macroevolution, or human evolution acceptance, and these have since been shown to be separate constructs (Nadelson and Southerland, 2012), so we incorporated these critiques when constructing our definition of evolution acceptance. Members of our authorship team, including the lead author of the original publication of the MATE, 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 our definition of acceptance of evolution, because microevolution acceptance is relatively high among students and thus may not be the most impactful target for evolution acceptance studies in the future (Barnes *et al.*, 2019, 2020a,b). This definition includes the term "species"; although 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 and Rieseberg, 2020; Wu *et al.*, 2020).

METHODS AND ANALYSES

Cognitive Interviews with Original MATE Survey

The first step of the study was to explore the process validity of the original MATE so that we could revise items based on any weaknesses we found. We 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, we recruited from an upper-level biology course for biology majors, four introductory-level science, technology, engineering, and mathematics (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. 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 Material 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 interviews.

Students were asked to fill out a brief demographic survey at the end of the interview (see Supplemental Material 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 proxy of their prior exposure.

To find any process errors in how students respond to MATE items, we qualitatively analyzed the cognitive interviews. We used a combination of deductive and inductive coding (Krippendorff, 2012; Cho and Lee, 2014). Student responses were initially coded using a deductively developed, relatively broad codebook with codes we expected to emerge from the data based on prior critiques of the MATE. Based on these prior published critiques (Smith, 2009a,b; Smith *et al.*, 2016; Nadelson and Southerland, 2012; Sickel and Friedrichsen, 2013; Rissler *et al.*, 2014), we began the interviews with a codebook that included codes to be applied if students answered questions based on their understanding of the 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 Judeo-Christian.

To establish the general interview protocol, M.E.B. and T.M. 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. T.M. then conducted the next three interviews alone, while M.E.B. listened to the interview recordings afterward. Again, we each took notes and came to agreement on how to apply the deductively derived codes. T.M. then conducted all the remaining interviews.

Because we also wanted to identify any weakness in the original MATE that we did not hypothesize before data collection, after the interviews were complete, we proceeded with inductive coding to analyze the interview data. We 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 with existing codes to determine whether an existing code is applicable or a new code is warranted (Cho and 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 Material for the full codebook).

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

Creating the MATE 2.0

Based on the findings from our cognitive interviews, the prior literature on evolution acceptance measurement, and our new definition of evolution acceptance, we 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 versus 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, we conducted cognitive interviews with 29 undergraduate students to confirm we had sufficiently improved the process validity of the items. We interviewed students across different religious affiliations, levels of acceptance, and levels of knowledge about evolution. The first set of students were recruited 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, we sent individual emails to recruit additional students who received low scores on measures of evolution acceptance on another survey. To include more students with lower prior exposure to evolution, we also recruited students from an introductory biology course for nonmajors 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). Our interview process and data analysis methods for the MATE 2.0 remained largely identical to our 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, we 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 (American Educational Research Association *et al.*, 2014). We expected the new MATE 2.0 to have higher bivariate correlations with measures of macroevolution and human evolution acceptance 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, we 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.*, 2020) and conducted a PCA of the residuals of this model. Low eigenvalue of the first contrast, that is, a value less than 2, indicates that the residuals are small and without structure and, therefore, data fit a unidimensional model (Boone, 2016; Sbeglia and 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 no. 00010903). We present the results and discussion together so that we can 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, we interviewed a total of 62 students for this portion of the study. Table 1 outlines the diversity of the sample in terms of student religious affiliation, gender identity, race/ethnicity, academic year, and evolution education exposure. See Supplemental Table 1 for each participant's gender identity, race/ethnicity, religious affiliation, evolution education experience, and average agreement with items on the MATE (range = 1–5). We report average composite scores, because they reflect an individual's average answer choice on the Likert scale, that is, a 4.0 out of 5.0 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).

We 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 1). 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.

In the following sections, we present our findings from the cognitive interviews with the original MATE. First, we present response errors present across items on the MATE (findings 1 and 2), and then we present response process errors specific to particular items on the MATE (finding 3). A list of items on the MATE and the most frequent codes applied to each specific item can be found in Supplemental Table 2.

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

TABLE 1. Characteristics of participants in cognitive interviews with the original MATE[a]

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	
Multiracial	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)

[a]Percentages may not add up to 100% due to rounding.

"microevolution," which does not include macroevolution or human evolution. That is to say, some students' definitions of evolution included only the evolutionary *processes*, 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 and 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 forms. 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 nonhuman 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 overly high MATE scores that suggest 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 NOS. 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 expected MATE scores and had item answers that were influenced by NOS misconceptions. The presence of these misconceptions typically resulted in students selecting answers that indicate a lower acceptance of evolution than their actual views.

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 and 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 did not strongly disagree.

> 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 the 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 the NOS to be positively correlated with acceptance of evolution, as measured by the MATE (Rutledge and Warden, 2000; Rutledge and Mitchell, 2002; Dunk *et al.*, 2017); 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 subtrends: 1) items that 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. In the following sections, we 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 and 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 and 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 answers or went so far as to explicitly point out how their answers to this item did not reflect their own views.

In one pattern that we 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 (emphasis ours).

> 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, we 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 and 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 (emphasis ours):

> 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 support 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. Students' 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, researchers 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 ~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 the NOS and confusion about the wording of the item. 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 did not strongly agree—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 was 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 nonreligious 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 students 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 ~13% of introductory biology students based on unpublished data from our research group). For instance, Muslim students tend to have similar evolution acceptance levels as Christian students (Barnes *et al.*, 2021), but if items reference Christian-specific conflict with evolution, researchers may not detect low acceptance among Muslim students with this item.

MATE Item 15: "Humans Exist Today in Essentially the Same Form in which They Always Have." For item 15, we 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 it 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 [emphasis ours].

> 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 only *Homo sapiens* 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 forms over a brief time span. 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 that were not related to one another and that 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

We 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 removed nine items, so the revised MATE consisted of 11 items, but after further cognitive interviews and Rasch dimensionality analyses, we removed two additional items due to response process errors and marginal acceptable fit statistics. Thus, the final version of the MATE 2.0 consists of nine items (Table 2).

Validity Evidence for the MATE 2.0

In the following sections, we report process validity, structural validity, and concurrent validity evidence for the MATE 2.0.

Finding 4: Cognitive Interviews Provide Process Validity Evidence for the MATE 2.0. To see whether we sufficiently revised items on the MATE to resolve process errors, we conducted cognitive interviews with a total of 29 students using the new items from the MATE 2.0. Of these students, five identified as nonreligious, 18 as Christian (including Catholic, Protestant, Latter-day Saints, 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 these 14 students, nine stated that nonhuman 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 one research-intensive institution and from a nationwide set of students from PhD-granting universities who had previously participated in related research. Items 10 and 11 arose as potentially problematic items during the first round of these interviews. 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 "cavemen" 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 versus prehistoric *Homo sapiens*. We flagged these items for potential deletion but kept them in the pilot to further assess their performance in the quantitative analyses.

TABLE 2. The MATE 2.0[a]

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 nonhuman 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

[a]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.

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 a research-intensive institution and as well as from a nationwide 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 (see section on finding 5 below).

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. For the MATE 2.0, students' limited understanding of the NOS influenced their answers for three items (items 3, 5, and 8), but this occurred only for 10% of students for items 3 and 8 and 14% of students for item 5. In contrast, seven items were influenced by NOS misconceptions on the original MATE, and this occurred for an average of 24% of students for these seven items. All other response process errors on the MATE 2.0 occurred in 10% of responses or fewer, which included low confidence in one's own knowledge for item 2 and a tendency to interpret "explains the diversity of life" as "explains *some* of the diversity of life" for item 7. As discussed in findings 1 and 2, misconceptions about the NOS and the use of incomplete definitions of evolution on the original MATE most often impacted particular students' answers across many items, rather than many students' answers on individual items. Misconceptions about the NOS impacted 10% of students on the MATE 2.0, down from 24% of students on the original MATE. Similarly, use of an incomplete definition of evolution impacted 3% of students on the MATE 2.0, down from 11% on the original MATE.

Additionally, the cognitive interviews revealed that, for some students, items measure acceptance of macroevolution at certain taxonomic levels. For example, Jelena had a mean composite score of 4.3 and selected "strongly agree" for item 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. Exploring students' conceptions of the common ancestry of life separate from their conceptions of speciation may be a fruitful area for future research.

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 to 1.37, which is largely within the acceptable range (i.e., 0.7–1.3 logits). However, items 10 and 11 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 Weighted Liklihood Estimate (WLE) person separation index, which estimates whether a similar order of person abilities would be generated by items of similar difficulty was 0.88. Because items 10 and 11 were marginally outside acceptable fit statistics and also showed some response process errors during cognitive interviews, we decided to remove them from the final version of the instrument. The eigenvalue of the first contrast for the Rasch model without items 10 and 11 was 0.87 and WMNSQ item fit statistics were all within the acceptable range of 0.7–1.3 logits. The EAP/PV reliability index was 0.91, and the 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 our sample of 2881 students, we 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 I-SEA (Nadelson and Southerland, 2012) were high (macroevolution: $r = 0.81$, $p < 0.001$; human evolution: $r = 0.82$, $p < 0.001$). 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 ($r = 0.67$, $p < 0.001$). 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 nonhumans 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 five-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), we 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 cutoff for low, medium, high, and very high scores (Rutledge and 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; that is, an average score of 4 across items would indicate a participant, on average, "agreed" with each item on a scale from 1 (strongly disagree) to 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 five-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). Finally, Rasch models yield equal-interval logit-scale measures, which are more suitable for parametric analyses such as regression analyses (Boone, 2016; also see Sbeglia and Nehm, 2019; Barnes *et al.*, 2020b). For these reasons, researchers can convert MATE scores using Rasch analysis to "ability" scores and use those scores for input in analyses. However, an evolution instructor who wants to measure the evolution acceptance of students in a course will likely not want to use Rasch and can simply use average composite scores as described above. A review of instructions for how to administer the MATE can be found in the Supplemental Material.

Naming of 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. We have 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 using newly developed items (I-SEA: Nadelson and Southerland, 2012; GAENE: Smith *et al.*, 2016), we addressed these concerns by identifying specific response process errors for each survey item in the original MATE, 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. In addition, we chose to build on to the most popular evolution acceptance instrument with a new version in hopes that others will use the MATE 2.0 instead of the MATE. Although this revised instrument has validation evidence and fewer response process errors, we encourage other researchers to build upon our work and, if additional issues arise in subsequent studies or with different student populations, consider revising it and naming it as another version.

CONCLUSIONS

In this study we explored the process validity of the MATE and created a new updated MATE 2.0. We find 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 NOS; 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. We 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, we hope that researchers will instead use this modified instrument to negate some of the limitations of the original MATE.

ACKNOWLEDGMENTS

We thank the students who took the time to share their thoughts and interpretations of the questions. We also thank Gale Sinatra, Randy Nesse, Erin Dolan, Emily Holt, Ryan Dunk, Dan Grunspan, and the ASU Biology Education Research Lab for thoughtful discussions about this project over the years. This project was supported by National Science Foundation grants IUSE 1818659, IUSE 1712188, and DGE-1311230.

REFERENCES

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

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. (2020a). "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., 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. (2020b). 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

Bishop, G. F. (1987). Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly*, *51*, 220–232

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

Cho, J. Y., & Lee, E.-H. (2014). Reducing confusion about grounded theory and qualitative content analysis: Similarities and differences. *Qualitative Report*, *19*(32), 1.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *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

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

Gallup. (2019). 40% of Americans believe in Creationism. *Gallup.Com*. Retrieved January 5, 2022, from https://news.gallup.com/poll/261680/americans-believe-creationism.aspx

Gao, L., & Rieseberg, L. H. (2020). While neither universally applicable nor practical operationally, the biological species concept continues to offer a compelling framework for studying species and speciation. *National Science Review*, *7*, 1398–1400.

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 & Practice*, *20*(7), 128–148. https://doi.org/10.33423/jhetp.v20i7.3157

Johns, R. (2005). One size doesn't fit all: Selecting response scales for attitude items. *Journal of Elections, Public Opinion & Parties*, *15*, 237–264. doi: 10.1080/13689880500178849

Koevering, T. E. V., & Stiehl, R. B. (1989). Evolution, creation & Wisconsin biology teachers. *American Biology Teacher*, *51*(4), 200–202. https://doi.org/10.2307/4448903

Krippendorff, K. (2012). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.

Kuschmierz, P., Meneganzin, A., Pinxten, R., Pievani, T., Cvetković, D., Mavrikaki, E., … & 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

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

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

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

Pew. (2013, December 30). Public's views on human evolution. *Pew Research Center's Religion & Public Life Project*. Retrieved January 5, 2022, from www.pewforum.org/2013/12/30/publics-views-on-human-evolution

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

Robitzsch, A., Kiefer, T., & Wu, M.. (2020). TAM: Test analysis modules (R Package Version 3.3, 9–35). Retrieved February 12, 2020, from https://CRAN.R-project.org/package=TAM

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/speglia

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*, *54*(5), 642–671. https://doi.org/10.1002/tea.21380

Rutledge, M. L., & Mitchell, M. A. (2002). High School Biology Teachers' Knowledge Structure, Acceptance & Teaching of Evolution. *The American Biology Teacher*, *64*(1), 21–28. https://doi.org/10.2307/4451231

Rutledge, M. L., & Sadler, K. C. (2007). Reliability of the Measure of Acceptance of the Theory of Evolution (MATE) instrument with university students. *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

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

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.

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

Smith, M. U. (2009a). 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. (2009b). Current status of research in teaching and learning evolution: II. Pedagogical issues. *Science & Education*, *19*(6–8), 539–571. https://doi.org/10.1007/s11191-009-9216-4

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 twelve years later. *Science & Education*, *25*, 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

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

Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.

Wu, C. I., Wang, X., He, Z., & Shi, S. (2020). Replies to the commentaries on the question of "Is it time to abandon the biological species concept?" *National Science Review*, *7*, 1407–1409.

Yasri, P., & Mancy, R. (2016). Student positions on the relationship between evolution and creation: What kinds of changes occur and for what reasons? *Journal of Research in Science Teaching*, *53*(3), 384–399. https://doi.org/10.1002/tea.21302

Zimmerman, M. (1987). The Evolution-Creation Controversy: Opinions of Ohio High School Biology Teachers. *The Ohio Journal of Science*, *87*(4), 115–125. https://kb.osu.edu/handle/1811/23208