

RESEARCH ARTICLE

Open Access



Different evolution acceptance instruments lead to different research findings

M. Elizabeth Barnes¹ , Hayley M. Dunlop¹, Emily A. Holt², Yi Zheng³ and Sara E. Brownell^{1*}

Abstract

Background: Despite widespread concern about the differential measurement of evolution acceptance among researchers, no one has systematically explored how instrument choice can impact research results and conclusions in evolution education studies. In this study, we administered six evolution acceptance instruments in a single survey to students in undergraduate biology courses at universities in Arizona, Colorado, and Utah. We conducted separate analyses with the same students for the six different evolution acceptance instruments to understand how different results and conclusions may arise based on different evolution acceptance instruments used.

Results: We found statistically significant differences in levels of evolution acceptance across the three student populations when using a human evolution acceptance instrument, but not when using a microevolution acceptance instrument. Further, significance/effect sizes of variables associated with evolution acceptance differed beyond sampling variation depending on the evolution acceptance instrument used. The results of analyses using different evolution acceptance instruments were most often dissimilar when examining the effect of evolution understanding and identifying as Protestant/Mormon on evolution acceptance.

Conclusions: We found that different instruments used to measure evolution acceptance sometimes led to different research results and conclusions. The extent to which variables predicted evolution acceptance was dependent on the instrument used to measure acceptance, which has the potential to explain over 30 years of conflicting research on the relationship between evolution acceptance and understanding. These results indicate that before researchers may be able to determine how to best improve evolution acceptance, the evolution education community may need to articulate a consistent definition of evolution acceptance and identify a singular valid and reliable instrument to quantify evolution acceptance so results can be compared across studies.

Keywords: Evolution acceptance, Evolution understanding, Religiosity, Nature of Science, Undergraduates, Survey instruments, MATE, GAENE

Introduction

Decades of research has resulted in little consensus on which factors are most important for student evolution acceptance (Barnes et al. 2017; Mead et al. 2018; Smith 2009a) and how to best increase evolution acceptance (Barnes and Brownell 2017; Mead et al. 2018). One explanation for this lack of consensus is that researchers use different instruments to measure evolution acceptance that were designed using different definitions of evolution

acceptance (Glaze and Goldston 2015; Lloyd-Strovas and Bernal 2012; Smith 2009a). In this study, we administered a survey containing multiple evolution acceptance instruments to undergraduate students in Arizona, Colorado, and Utah. We illustrate the similarities and differences in the results from each instrument and how the instrument choice can impact the results and conclusions of a study.

Background

Due to the low levels of evolution acceptance among members of the public (Gallup 2017; Pew 2013) and college students (Brem et al. 2003; Ingram and Nelson 2006; Rice et al. 2010; Walter et al. 2013), research on how to increase evolution acceptance has become

*Correspondence: sara.brownell@asu.edu

¹ School of Life Sciences, Biology Education Research Lab, Arizona State University, Tempe, AZ, USA

Full list of author information is available at the end of the article



one of the predominant subfields of evolution education. Despite over 30 years of research on how to improve evolution acceptance, rates of acceptance in the United States have remained relatively unchanged (Gallup 2017). Further, although at least 300 articles have been published examining evolution acceptance, little consensus has emerged regarding the relationship between different variables and evolution acceptance. While some studies have found large positive relationships between evolution acceptance and evolution understanding (Rutledge and Warden 1999; Trani 2004), some studies report a weak relationship between acceptance and understanding (Athanasios and Papadopoulou 2012; Cavallo et al. 2011; Deniz et al. 2008; Großschedl et al. 2014; Nadelson and Sinatra 2009), and other studies report no relationship between evolution acceptance and understanding (Bishop and Anderson 1990; Brem et al. 2003; Lawson 1983; Sinatra et al. 2003). Further, what best predicts evolution acceptance is variable across studies; religiosity, evolution understanding, and Nature of Science (NOS) understanding have each been reported as the biggest predictors for evolution acceptance in different studies (Carter and Wiles 2014; Dunk et al. 2017; Glaze et al. 2014; Mead et al. 2018; Weisberg et al. 2018). This lack of overall consensus on the relationships between evolution acceptance and other variables could be one reason why we have seen so little change in evolution acceptance in the United States for 30 years. How can educators determine the best methods for increasing evolution acceptance if the research community has not reached consensus on how variables relate to evolution acceptance?

One explanation for these inconsistencies in evolution acceptance findings is that researchers measure evolution acceptance differently and that this could lead to disparate results and conclusions. Prior to the publication of peer-reviewed evolution acceptance instruments, evolution education researchers used dozens of different evolution acceptance instruments that were usually constructed for use in a single study (Bishop and Anderson 1990; Johnson and Peeples 1987; Lawson 1983; Sinatra et al. 2003). The Measure of the Theory of Acceptance of Evolution (MATE) was published in 1999 and it has slowly gained popularity in the evolution education community to measure evolution acceptance (Rutledge and Warden 1999). More recently, the Inventory of Student Evolution Acceptance (I-SEA; Nadelson and Southerland 2012) and the Generalized Acceptance of Evolution Exam (GAENE; Smith et al. 2016) have been published to measure evolution acceptance. The availability of different instruments to measure evolution acceptance means that evolution education researchers have to

make a decision about the best way to measure evolution acceptance.

Evolution education researchers have expressed repeated concern about how evolution acceptance is measured (Glaze and Goldston 2015; Lloyd-Strovas and Bernal 2012; Sickel and Friedrichsen 2013; Smith 2009a). Dating back to more than 20 years ago, review articles have suggested problems with the measurement of evolution acceptance. In Smith et al. (1995) the authors highlighted that individuals with different levels of Nature of Science (NOS) understanding could be confused by the wording of questions in instruments meant to measure evolution acceptance, which could lead to inflated or even erroneous correlations between Nature of Science (NOS) understanding and evolution acceptance. In 2009, Smith et al. wrote a review article in which they also expressed concern that some instrument items meant to capture evolution acceptance may measure evolution understanding, which could lead to inflated correlations between evolution understanding and evolution acceptance. When Lloyd-Strovas and Bernal (2012) reviewed the literature on undergraduate evolution education, they found it hard to detect patterns because the instruments used to measure evolution acceptance were so different that they claimed the studies were “not comparable.” In Sickel and Friedrichsen (2013) the authors raised concerns that instruments used to measure evolution acceptance of the respondent include items about whether the respondent thinks that scientists accept evolution, which could lead to inflated rates of evolution acceptance in research findings. Nadelson and Southerland (2012) further expressed concern that many evolution acceptance instruments do not disentangle the role of context (e.g. evolution occurring in humans or evolution occurring in plants) in evolution acceptance. However, despite the prevalence of these concerns, the evolution education community has not reached a consensus for how we should measure evolution acceptance and researchers continue to use a variety of instruments to measure student evolution acceptance. Further, researchers often compare conclusions from studies that use different evolution acceptance instruments (Wiles and Alters 2011; Glaze and Goldston 2015). This practice implies that conclusions are comparable across studies using different evolution acceptance instruments, but little prior research has determined whether using different evolution acceptance instruments leads to the same conclusions about evolution acceptance (Romine et al. 2018; Sbeglia and Nehm 2018).

In our literature review, we found preliminary evidence that different instruments to measure evolution acceptance could be a cause of inconsistent research results in the literature. When we examined research studies

exploring the relationship between evolution acceptance and evolution understanding, we found that Bishop and Anderson (1990), Sinatra et al. (2003), and Hermann (2012) all used similar measures of evolution acceptance in which students were asked the extent to which they believed/accepted evolution or thought evolution was true/credible. All three of these studies found no relationship between evolution acceptance and evolution understanding. However, we found that in studies in which researchers used the Measure of Acceptance of the Theory of Evolution (MATE) to measure evolution acceptance, researchers consistently found a positive relationship between acceptance and understanding (Rutledge and Warden 2000; Trani 2004); no studies to our knowledge that use the MATE have ever reported an insignificant relationship between evolution acceptance and understanding. However, these studies were conducted with different populations of students, so researchers cannot determine from these studies alone whether it was the evolution acceptance instruments used that led to the conflicting research findings.

There have been very few published studies that have used the I-SEA or the GAENE to measure evolution acceptance because they are fairly new. Nadelson and Hardy (2015) have used the I-SEA to show that introductory undergraduate psychology students' acceptance of microevolution, macroevolution, and human evolution are related to more trust in science and scientists, less conservative political orientations, and weaker religious commitment, similar to findings using other instruments (Dunk et al. 2017; Glaze and Goldston 2015). Using the I-SEA, other studies have shown that undergraduate students have higher acceptance of microevolution than macroevolution and human evolution (Nadelson and Hardy 2015; Nadelson and Southerland 2012; Schleith 2017). In one study, researchers showed that scores on the GAENE and MATE were strongly correlated among introductory undergraduate health sciences students (Metzger et al. 2018) and one study showed that GAENE scores increased among some high school students after evolution instruction (Pobiner et al. 2018). No peer-reviewed published study to our knowledge has reported the results of the relationship between evolution understanding and acceptance using either the GAENE or I-SEA.

One goal of the current study was to explore whether inconsistent research conclusions about evolution understanding and evolution acceptance could arise because of different instruments used to measure evolution acceptance. However, we also explore whether research findings could be inconsistent when using different evolution acceptance instruments to examine relationships with

Nature of Science (NOS) understanding, religiosity, religious affiliation, political affiliation, and race/ethnicity.

Research questions and methods

The overarching goal for the study was to compare research findings from different instruments that have been previously used to measure evolution acceptance. We administered six different multi-item instruments to measure evolution acceptance in a single survey to the same students and then compared the findings from each instrument using predefined criteria to determine what results and conclusions were different.

Our specific research questions were:

1. Do various instruments lead to different conclusions about the level of evolution acceptance among populations?
2. Do various instruments lead to different results and conclusions about the relationships between student variables (e.g., evolution understanding, religion) and levels of evolution acceptance?

Survey and administration

In the fall of 2017 and spring of 2018, we sent a survey to ~2300 students from nine introductory biology courses at a research-intensive university in urban Arizona, ~190 students in two introductory biology courses at a comprehensive institution in rural Colorado, and ~200 students in four introductory biology courses at a primarily undergraduate institution in suburban Utah. Data were collected mid-semester. Students were offered a small amount of extra credit for completing the survey. We collected student evolution acceptance with six evolution acceptance instruments and also collected demographic information from students. To ensure differences in results were not due to an instrument order effect, students were given the evolution acceptance instruments in a random order. Demographic questions were presented at the end of the survey. For Arizona students, we also collected data on their evolution understanding and Nature of Science (NOS) understanding. See Table 1 for a list of data we collected for this study. The research study was approved by the Arizona and Colorado institutional review boards for all research in this manuscript, Protocol Numbers 00007719 and 1131916-2, respectively.

Evolution acceptance instruments

We administered six evolution acceptance instruments to students. Below we describe each instrument, including its prior prevalence in the literature, the motivation

Table 1 Data collected in the current study for each student population

	Arizona survey	Colorado survey	Utah survey
Evolution understanding	✓	X	X
Nature of Science understanding	✓	X	X
Randomized			
I-SEA: macroevolution	✓	✓	✓
I-SEA: microevolution	✓	✓	✓
I-SEA: human evolution	✓	✓	✓
GAENE	✓	✓	✓
MATE	✓	✓	✓
100-pt	✓	✓	✓
Parent education	✓	✓	✓
Gender	✓	✓	✓
Political affiliation	✓	✓	✓
Race/ethnicity	✓	✓	✓
Religious affiliation	✓	✓	✓
Religiosity	✓	✓	✓
Course	✓	✓	✓

A “✓” means that the data were collected for the indicated population and “X” means that the data were not collected in that student population. Instruments are presented here in the same order as the survey administered to students. Evolution acceptance instruments were presented in a random order. Students were presented with the I-SEA macroevolution, microevolution, and human evolution instruments in that order in every survey

*GPA was collected from university registrar for Arizona students only

behind the instrument’s construction, its unique features compared to other instruments, and its validity and reliability evidence.

Measure of Acceptance of the Theory of Evolution (MATE; Rutledge and Warden 1999)

The MATE is the most popular instrument used to measure evolution acceptance in the evolution education literature (Smith et al. 2016); we found 51 studies published in academic journals that have used the MATE to measure evolution acceptance.

The MATE has 20-items and was originally designed to measure high school biology teachers’ evolution acceptance. Items on the MATE were reviewed and approved by a group of evolutionary biologists, science educators, and a philosopher of science to establish content validity (Rutledge and Warden 1999). The authors of the MATE reported that a factor analysis of their instrument revealed only one factor, and the authors deemed this single construct was evolution acceptance. However, more recent analyses have suggested the MATE is multi-dimensional (Metzger et al. 2018). The authors of the MATE reported acceptable reliability with high school biology teachers (Rutledge

and Warden 1999) and in a subsequent publication from Rutledge and Sadler (2007), the authors reported acceptable reliability of the MATE with college students. High reliability of the MATE has been confirmed by multiple subsequent studies (Barone et al. 2014; Manwaring et al. 2015; Nadelson and Sinatra 2009).

Rutledge and Warden (1999) did not provide an explicit definition of “evolution acceptance” for the basis of their instrument, but did cite Schwabb (1968) as justification for what they included in the MATE, in the following passage:

“because informed decisions of acceptance or rejection of a scientific theory are based on evaluations of substantive and syntactical elements of a domain, fundamental concepts of evolutionary theory and the nature of science were selected to be addressed by the MATE: the processes of evolution, the available evidence of evolutionary change, the ability of evolutionary theory to explain phenomena, the evolution of humans, the age of the earth, the independent validity of science as a way of knowing, and the current status of evolutionary theory within the scientific community.” (pg. 14)

Rutledge and Warden define what concepts are included in the definition of “evolution acceptance,” but do not define what is meant by “acceptance.” Based on the strongly agree-strongly disagree Likert-style response scale of the MATE, acceptance is implicitly defined as level of agreement with each of the concepts included in the MATE.

The MATE has been criticized for including questions that could measure evolution understanding, Nature of Science (NOS) understanding, religiosity, and perceptions of scientists’ acceptance of evolution (Smith 2009a). If items on the MATE measure these constructs in addition to evolution acceptance, then we may see inaccurate correlations between evolution acceptance and these constructs. For instance, one item on the MATE asks students if they agree that “the age of the earth is at least 4 billion years”; to be scored as an acceptor on this question, students would need to use their understanding of the age of Earth. A student could have an inaccurate understanding of the age of the Earth (e.g. that it is 2 billion years old) and choose “disagree” even though they accept that the Earth is old and accept evolution. The use of the MATE is often justified by its prior prevalent use in the literature, even though many criticisms of this instrument have been published (Metzger et al. 2018; Nadelson and Southerland 2012; Romine et al. 2017; Smith et al. 2016).

Inventory of Student Evolution Acceptance (I-SEA: microevolution, macroevolution, and human evolution; Nadelson and Southerland 2012)

The I-SEA was published by Nadelson and Southerland in 2012 and we identified two studies published in the academic literature (Nadelson and Southerland 2012; Nadelson and Hardy 2015) and three dissertations/theses that have used the I-SEA to measure evolution acceptance.

The I-SEA is a 24-item Likert-style instrument developed to measure evolution acceptance among high school and college students. The I-SEA was constructed to provide a new measure of acceptance that addressed what the authors considered two shortcomings of prior instruments used to measure evolution acceptance. First, they argued that other instruments conflated students' evolution acceptance with student evolution understanding, so the I-SEA was designed to not conflate these. Second, other instruments did not disaggregate microevolution from macroevolution from human evolution, so the I-SEA was developed with three distinct subscales of student acceptance of microevolution, macroevolution, and human evolution.

Nadelson and Southerland (2012) defined evolution as being comprised of three distinct contexts that are relevant for evolution acceptance: microevolution; defined as the results of evolution in the short term, macroevolution; defined as the results of evolution in the long term, and human evolution; defined as the evolution of the human species specifically. They defined acceptance as "the examination of the validity of the knowledge supporting the construct, the plausibility of the construct for explaining phenomenon, persuasiveness of the construct and fruitfulness or productivity of the empirical support for the construct." (pg. 1639)

The I-SEA definition of evolution acceptance is different from other instruments in that it discriminates acceptance of different evolutionary contexts. Macroevolution and human evolution tend to be in direct conflict with commonly held religious beliefs in the United States, but microevolution is not (Pobiner 2016; Scott 2005). Therefore, levels of evolution acceptance are often higher for microevolution than macroevolution and human evolution (Nadelson and Hardy 2015; Nadelson and Southerland 2012) and relationships between predictor variables and evolution acceptance may change depending on the different subscale.

Exploratory and confirmatory factor analyses of the I-SEA have confirmed a three factor structure of microevolution, macroevolution, and human evolution to the I-SEA and the three resulting subscales had high internal consistency coefficients of $>.80$. Further, content validity of the I-SEA was supported by a group of experienced biology teachers, science teacher educators, and college

biology faculty members who reviewed the items on the I-SEA (Nadelson and Southerland 2012). The authors of the I-SEA recommended that the instrument could be used as three separate instruments to measure acceptance of evolution or as an aggregate single instrument. Because research findings may be different for different subscales on the I-SEA, we treat the I-SEA as three distinct instruments: I-SEA microevolution acceptance, I-SEA macroevolution acceptance, and I-SEA human evolution acceptance.

Generalized Acceptance of Evolution Evaluation (GAENE; Smith et al. 2016)

The GAENE was published in 2016 and we have identified three studies (Metzger et al. 2018; Pobiner et al. 2018; Smith et al. 2016) published in academic journals that used the GAENE to measure evolution acceptance.

The GAENE is a 13-item Likert-style instrument originally designed to measure high school and college students' evolution acceptance. Items on the GAENE went through an iterative construction process, with two rounds of pilot testing with students, two rounds of validation with science education experts, and two rounds of reliability, factor, and Rasch analyses, the second set of which showed acceptable reliability and validity of the GAENE with high school and college students (Smith et al. 2016).

The GAENE was constructed to provide a new measure of evolution acceptance that addressed what the authors identified as two major weaknesses of other instruments used to measure evolution acceptance. First, the authors of the GAENE built the instrument so that it would not conflate evolution understanding with evolution acceptance, and second, it was built from an explicit definition of evolution acceptance while other instruments used to measure evolution acceptance often operate from an implicit definition of evolution acceptance (Smith et al. 2016). The authors explicitly define 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." (pg. 8)

Items on the GAENE, unlike items on other instruments used to measure evolution acceptance, gauge the extent to which an individual is willing to *advocate* for evolution. For instance, students are asked to agree or disagree with statements such as, "It is important to let people know about how strong the evidence for evolution is," and "I would be willing to argue in favor of evolution

in a public forum such as a school club, church group, or meeting of public school parents,” which is different from other evolution acceptance instruments in which students are asked how much they agree with the current claims of evolutionary theory.

100-point instrument of self-defined acceptance

We constructed an evolution acceptance instrument with the purpose of mimicking the ones used by Bishop and Anderson (1990), Sinatra et al. (2003), and Hermann (2012) described in the introduction. We found five studies that used similar instruments to measure evolution acceptance. Our instrument was a composite score of three items in which students used a slider scale to indicate from 0 to 100 points: (1) “To what extent do you accept evolution?” (2) “To what extent do you believe evolution?” and (3) “To what extent do you think evolution is true?” In the study by Sinatra et al. (2003), students placed an “X” on a horizontal number line to indicate the extent to which they thought evolution was credible and in Bishop and Anderson (1990) students were asked, “Do you believe the theory of evolution to be truthful?” In Hermann (2012) students were asked “To what extent do you accept (believe) evolution?” None of these three studies found a significant relationship between evolution understanding and acceptance. The instrument used in the current study was constructed with the purpose of mimicking an instrument like these that does not provide any definition of evolution acceptance and only relies on the student’s definition of evolution acceptance. This is the only instrument we use in this study that relies completely on the respondents’ own definition of evolution acceptance; other instruments generally provide students with specific contexts about evolution acceptance with which to agree or disagree.

We conducted think-aloud interviews (Willis 2004) with 25 undergraduate students with these items from this instrument. We instructed students to read each question aloud, then explain what they perceive the question is asking, and then to walk us through their reasoning as they answer the question. Finally, we asked students if there was anything confusing about the questions and if they had any suggestions for how they could be improved. Interviews indicated that students did not misinterpret the questions and answered the questions in a manner we would expect given their reasoning (e.g., one student answered “100” to “to what extent do you think evolution is true?” because she thought the scientific evidence was strong for evolution while another student answered “7” because she thought there was evidence for some aspects of evolution but not others). No student found the questions confusing or in need of clarification. However, six out of 25 students did indicate

that the questions were repetitive and seemed to be asking the same question. Inter-item correlation coefficient ranges for items on this instrument were very high ($r = .85-.89$) and the reliability coefficient also very high ($\alpha = .95$). Unlike the other evolution acceptance instruments used in this study, this instrument did not undergo formal evaluation by experts and was not published as a peer-reviewed paper dedicated to the instrument’s development. Nonetheless, it is similar to instruments used in other peer-reviewed studies published by leaders in the field of evolution education and these studies have been highly cited and thus fairly influential in the evolution education literature (Bishop and Anderson 1990; Sinatra et al. 2003).

Demographics and other predictor variables

We collected information on race/ethnicity, parent education level, political affiliation, religious affiliation from all students. Religiosity was also collected from all students and was defined as the extent which students see religion as important to their identity and the extent to which they participate in religious activities. Those associated with a religious denomination and those who selected “nothing in particular” as their religious affiliation were categorized based on their agreement on a 5-point Likert-scale from strongly disagree to strongly agree with the statements “I attend religious services regularly” and “my religion or faith is an important part of my identity” (Cohen et al. 2008). Religiosity was treated as an ordinal variable in which atheists and agnostics were categorized at the lowest end.

To measure Arizona students’ evolution understanding, we used two subscales on the Evolutionary Attitudes and Literacy Instrument (EALS; Hawley et al. 2010). We only used the two subscales (13 items) from the instrument that measure “Evolutionary Knowledge” (e.g., “In most populations, more offspring are born than can survive”) and “Evolutionary Misconceptions” (e.g., “Evolution is a linear progression from primitive to advanced species”). Students were asked to decide whether each item was true or false based on their evolution understanding, rather than on a Likert scale from strongly agree-strongly disagree so that the student would not answer based on their personal opinion but based on their conceptual understanding of the scientific theory of evolution. Students’ scores were calculated by determining the number of items answered correctly. We chose to use the EALS to measure evolution understanding because it has been used in other evolution education studies (Dunk et al. 2017; Short and Hawley 2015), has shown evidence of reliability and validity among college students (Hawley et al. 2010), and the items do not appear to conflate evolution acceptance with evolution understanding.

Table 2 Demographic characteristics of three student populations

	Arizona (n = 742)	Colorado (n = 102)	Utah (n = 79)
Politics			
Republican	16.0%	15.7%	57.0%
Other affiliation	84.0%	76.5%	33.0%
Race/ethnicity			
Black/African Am.	3.9%	3.9%	0%
Hispanic	22.4%	14.7%	1.3%
Asian	15.0%	2.0%	.0%
White	52.3%	69.8%	86.1%
Other race/ethnicity	6.5%	17.6%	8.9%
Gender			
Male	35.1%	26.7%	36.7%
Female	62.9%	73.3%	60.8%
Religion			
Mormon/LDS	1.8%	1.0%	92.2%
Protestant	21.0%	33.3%	.0%
Catholic	24.7%	21.6%	.0%
Other religion	9.6%	4.9%	.0%
No religion	42.2%	33.3%	7.6%
Religiosity on scale of 1–10: Mean (SD)	4.45 (3.59)	4.74 (3.27)	7.85 (3.29)

To measure Arizona students' Nature of Science (NOS) understanding, we used the Rutledge and Warden (2000) modified version of the Johnson and Peeples (1987) instrument. This 20-item instrument probes students about a wide range of characteristics of science, including but not limited to: the tentative nature of science (e.g., "A fact in science is a truth that can never be changed"), the scientific method (e.g., "The initial step of the scientific method is to test a hypothesis"), the techniques of science (e.g., "To make any determinations about historic occurrences in nature, there must be direct observations"), and the limits of science (e.g., "Scientists must limit their investigations to the natural world"). Students' scores were calculated by determining the number of items answered correctly (True-or-False response scale). We chose to use this instrument because it has been repeatedly used in the evolution education literature to measure Nature of Science (NOS) understanding and has shown large correlations with evolution acceptance in several studies (Dunk et al. 2017; Glaze et al. 2014; Rutledge and Warden 2000).

We also collected Arizona students' cumulative end of semester college GPA from the university's registrar.

Analyses and results

We wanted to compare research findings from different instruments when the student sample remained constant. Therefore, we only included data from students that provided answers for all questions used in the analyses, so the sample would be exactly the same across the analyses using different instruments.¹ This resulted in 742 complete responses from students in Arizona, 102 complete responses from students in Colorado, and 79 complete responses from students in Utah. We used SPSS version 25 for all analyses.

The religious background and political beliefs of students in Arizona, Colorado, and Utah were notably different. Ninety-two percent of students who we surveyed in Utah identified as Mormon/LDS, while only 1–2% of students in Arizona and Colorado identified as Mormon/LDS. Further, students in Utah scored one standard deviation higher on the measure of religiosity compared to students in Arizona and Colorado (Table 2). Students in Utah were also more likely to identify as Republican (57%) compared to students in Arizona (16%) and Colorado (16%). Past research shows that the general Mormon population has some of the lowest evolution acceptance rates (Baker et al. 2018), so we would expect a priori that the Utah student population would be less accepting of evolution compared to students in Arizona and Colorado. Arizona and Colorado students scored similarly on mean religiosity (Table 2), which was surprising because Colorado is generally rated as a less religious state

¹ The majority of the excluded responses had data that was missing at random.

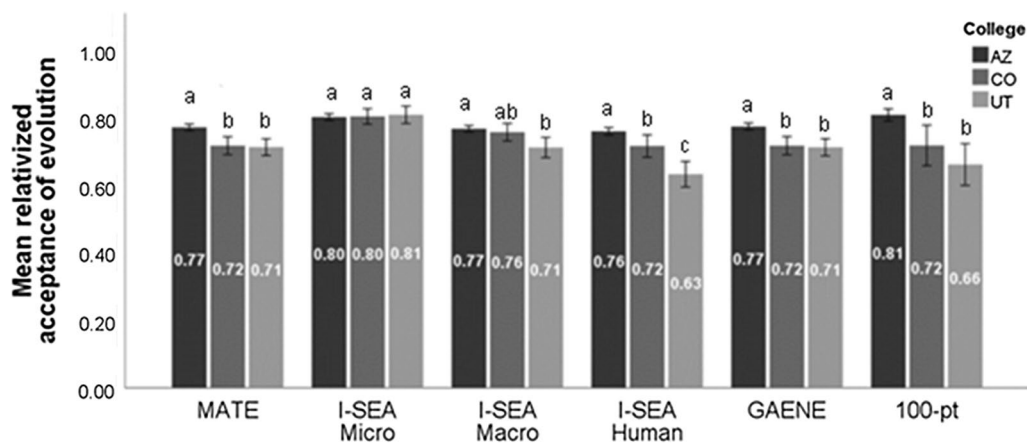


Fig. 1 Results comparing mean relativized evolution acceptance scores on six different evolution acceptance instruments from students in Arizona, Colorado, and Utah using Welch's test and Games-Howell post hoc comparisons tests. Error bars represent the 95% confidence interval. Different letters represent statistically different scores among populations for each instrument. Comparisons were made across populations for a single instrument and not across instruments

compared to Arizona (Pew 2016). However, the institution sampled in Colorado was located in a rural part of Colorado, which could explain the similarity in Arizona and Colorado students' religiosity scores. See Table 2 for a comparison of the demographics collected from all three student populations.

For ease of interpretation, all scores from evolution acceptance instruments were relativized by the maximum scores so that scores from each instrument ranged from 0 to 1. From this point onwards, when we refer to instrument scores we are referring to the relativized scores. Internal consistency analyses revealed that all six evolution acceptance instruments had acceptable reliability ($\alpha = .84-.98$). All alpha values for each evolution acceptance instrument by student population can be found in the Additional file 1: Table S1.

Bivariate correlations of students' scores on the six different evolution acceptance instruments revealed that scores from the six different instruments were moderately to strongly correlated with one another in Arizona ($r = .49-.83$, $p < .001$), Colorado ($r = .49-.82$, $p < .001$), and Utah ($r = .36-.79$, $p < .001$). Correlation coefficients for each instrument by population of student can be found in the Additional file 1: Tables S2–S4.

Research question #1: do various instruments lead to different conclusions about the level of evolution acceptance among populations?

One research question that evolution education researchers often explore is whether evolution acceptance differs across students in different contexts (Short and Hawley 2015). We examined whether different evolution acceptance instruments lead to different conclusions about

levels of evolution acceptance among three undergraduate student populations.

The samples in Arizona, Colorado, and Utah did not meet the homogeneity of variance assumptions needed to run one-way ANOVA's when exploring population level differences with the MATE (Levene's statistic = 7.95, $p < .001$), the GAENE (Levene's statistic = 3.94, $p = .02$), the I-SEA microevolution (Levene's statistic = 4.92, $p = .007$), or the 100-pt scale (Levene's statistic = 11.18, $p < .001$), according to Levene's test. Therefore, to determine if there were differences in evolution acceptance across the three student populations, we ran Welch's robust tests of equality of means (Field 2009). Following Welch's test, we then conducted Games-Howell post hoc tests that corrected for multiple comparisons (Field 2009) to identify which populations were statistically significantly different from one another. We conducted a separate test for each of the six different evolution acceptance instruments as the dependent variable with student population (AZ, CO, or UT) as the predictor variable. We consider a conclusion using one instrument to be different from a conclusion using a different instrument when one instrument detects statistically significant differences between two populations and another instrument does not.

Welch's test detected differences in mean evolution acceptance scores across populations using all evolution acceptance instruments (Welch's range (2, 920) = 14.61–22.02, $p < .01$), except the I-SEA microevolution instrument (Welch's (2, 920) = .123, $p = .88$) in which student scores in AZ, CO, and UT were not deemed statistically significantly different from one another, indicating that each instrument showed

similar patterns (see Fig. 1). Games-Howell post hoc tests revealed that Arizona students scored higher than Utah students on evolution acceptance when using every instrument ($p < .01$), except for the I-SEA microevolution instrument ($p = .88$). When comparing Arizona and Colorado populations, the test revealed higher scores among students in Arizona when using the MATE ($p = .001$), the I-SEA human evolution instrument ($p = .039$), the GAENE ($p < .001$), and the 100-point instrument ($p = .014$) but not using the I-SEA microevolution instrument ($p = .986$) or the I-SEA macroevolution instrument ($p = .754$). When comparing Colorado and Utah students, the test only detected differences using scores from the I-SEA human instrument ($p = .003$), in which Colorado students scored higher. Figure 1 summarizes the results comparing mean levels of evolution acceptance across populations by evolution acceptance instrument.

Research question #1 discussion We found that most evolution acceptance instruments, except the I-SEA microevolution and I-SEA macroevolution, showed that AZ had the highest evolution acceptance, so many of the instruments showed similar patterns. However, differences in evolution acceptance across populations were found with some evolution acceptance instruments but not others, indicating the results were inconsistent with some instruments. This was most notable within the three subscales of the I-SEA. Differences in levels of evolution acceptance between populations were consistently found using the I-SEA human instrument, but not the I-SEA microevolution instrument, which supports that there is higher variation in levels of acceptance of human evolution across different populations of students than microevolution.

Students in Utah scored higher on religiosity, so we expected that there would be higher perceived conflict between evolution and these students' religious beliefs compared to the students in Arizona and Colorado who were less religious on average. As Nadelson and South-erland (2012) argued when they created the I-SEA, religious individuals may perceive more conflict between their religious beliefs and human evolution than macroevolution or microevolution and this may explain the difference we see in results from the I-SEA microevolution and I-SEA human evolution instruments.

These results suggest that using the I-SEA subscales, and specifically separating the analyses by subscale, may give researchers a more nuanced understanding of the differences in evolution acceptance across populations. Further, if researchers use instruments for evolution acceptance that are more focused on microevolution, then researchers could report higher rates of evolution acceptance and could also miss differences

in overall evolution acceptance between populations of students. Further, if researchers do not inquire about human evolution acceptance with their instrument, they may not identify population level differences in evolution acceptance that would be apparent with a different instrument that included questions about human evolution.

Research question # 2: do various instruments lead to different results and conclusions about relationships between student variables (e.g. evolution understanding, religion) and levels of evolution acceptance?

Another common aim of evolution education research is to identify what variables are related to evolution acceptance and what positively or negatively predicts whether someone will be accepting of evolution. We examined whether researchers could get different results and come to different conclusions about what variables would predict evolution acceptance if they used different evolution acceptance instruments.

We restricted these analyses to students in Arizona for which we collected student evolution understanding and Nature of Science (NOS) understanding. We used linear regression analyses to explore whether variables predicted evolution acceptance differentially depending on the instrument used to measure evolution acceptance. We inputted all predictor variables (evolution understanding, NOS understanding, GPA, course level, parent education level, religiosity, religious denomination, political affiliation and race/ethnicity) into regressions with scores from each evolution acceptance instrument as a dependent variable (six regressions total). Table 3 illustrates how each variable was inputted in the analyses.

To determine if study results were different across analyses we compared confidence intervals of the unstandardized beta coefficients from each independent variable across the six regressions and identified the cases in which the confidence intervals do not overlap. If the confidence intervals for a single independent variable do not overlap between analyses then we deem the revealed impacts of this independent variable on evolution acceptance to be different beyond sampling variation (Schenker and Gentleman 2001) and constitute a different "result" (Table 5). We then determined if researchers would make different conclusions based on these data by comparing whether a variable will be deemed statistically significant across analyses and thus whether the variable would be considered a significant predictor of evolution acceptance across analyses. If when using one instrument we would conclude that a variable was statistically significantly related to evolution acceptance, but with another instrument we would conclude that this same variable

Table 3 Description of categorical and ordinal variables in the regression analyses

Categories in analyses (categories in survey)
Course level
Non-majors = 1 (BIO 100)
Intro majors I = 2 (BIO 181, BIO 281)
Intro majors II = 3 (BIO 182, BIO 282)
Parent education
First generation = 1 (first generation)
Non-first generation = 2 (≥ 1 parent attended college)
Non-first generation = 3 (≥ 1 parent has college degree)
Religious affiliation
Catholic (Christian—Catholic)
Protestant or LDS (Christian—Protestant, Christian—LDS)
Other Religion (Jewish, Muslim, Buddhist, Hindu, Mormon, other religion)
Reference group—no religion (atheist, agnostic, nothing in particular)
Political affiliation
Republican (Republican)
Democrat (Democrat)
Reference group: other affiliation—(independent, libertarian, not political, other)
Race/ethnicity
Asian (Asian)
Black/African American (Black or African American)
Hispanic (Hispanic)
Other race (American Indian or Alaska Native, Pacific Islander, mixed race, other race)
Reference group—White (white)

is not statistically significantly related to acceptance, we deemed that this would be a different “conclusion” drawn based on different instruments used (Table 6).

Students had an average GPA of 3.3 ($SD = .61$), scored an average of 9 out of 13 on the understanding of evolution measure ($SD = 2.14$), and scored an average of 13 out of 20 on the Nature of Science (NOS) understanding measure ($SD = 2.68$).

Overall, the collective variables explained a statistically significant amount of variation in evolution acceptance from six models across all evolution acceptance instruments (range of adjusted $R^2 = .29-.42$, $p < .001$). The variables explained 42% of the variation in MATE scores ($F(16, 725) = 34.20$, $p < .001$), thirty-five percent of the variation in I-SEA microevolution scores ($F(16, 725) = 26.36$, $p < .001$), thirty-one percent of the variation in macroevolution scores ($F(16, 725) = 21.82$, $p < .001$), thirty-five percent of the variation in I-SEA human evolution scores ($F(16, 725) = 26.04$, $p < .001$), twenty-nine percent of the variation in GAENE scores ($F(16, 725) = 18.31$, $p < .001$), and thirty-one percent of the variation in 100-pt evolution acceptance scores ($F(16,$

$725) = 21.87$, $p < .001$). All six full regression tables with r -squared values, F -statistics, standardized and unstandardized coefficients, standard errors, and t -statistics, can be found in the Additional file 1: Tables S5–S10.

Religiosity was the strongest predictor of evolution acceptance across all instruments (β range = $-.29$ to $-.25$, $p < .001$), except for the I-SEA microevolution acceptance instrument where evolution understanding was the strongest predictor ($\beta = .26$, $p < .001$).

We found that the *results* across analyses were often different. We compared the confidence intervals of the unstandardized beta coefficients for each predictor variable across the six regression models. Non-overlapping confidence intervals from unstandardized beta coefficients indicated that the regression coefficients for several predictor variables were different based on the evolution acceptance instrument used. We identified non-overlapping confidence intervals for the effect of parent education, evolution understanding, Nature of Science (NOS) understanding, religiosity, and identifying as Protestant or LDS, Muslim, Republican, Hispanic, or another race/ethnicity. All calculated confidence intervals and the p -values for coefficients can be found in Table 4. A summary of cases in which confidence intervals overlap and do not overlap can be found in Table 5.

Of particular note was the differential effect of evolution understanding and religious denomination across evolution acceptance instruments. Evolution understanding had a noticeably variable effect on evolution acceptance depending on the evolution acceptance instrument used. Confidence intervals for the effect of evolution understanding on MATE scores and I-SEA microevolution scores were almost identical ($CI = .015, .019$), but confidence intervals for the effect of evolution understanding on I-SEA macroevolution scores ($CI = .010, .014$), I-SEA human evolution scores ($CI = .009, .014$), GAENE scores ($CI = .003, .008$), and 100-pt scores ($CI = .004, .012$) were lower, thus indicating that evolution understanding is a weaker predictor with these instruments. Additionally, the confidence interval for the effect of identifying as Protestant/LDS was lower for I-SEA microevolution scores ($CI = -.040, -.009$) than MATE scores ($CI = -.076, -.046$), I-SEA macroevolution scores ($CI = -.085, -.053$), I-SEA human evolution scores ($CI = -.092, -.056$), and GAENE scores ($CI = -.093, -.058$). Additionally, the effect of identifying as Protestant/LDS appeared to be particularly strong in predicting scores on the 100-pt scale ($CI = -.186, -.127$).

The differences in confidence intervals of regression coefficients indicate different study “results”, with a “result” defined as the coefficient rendered by the analyses. This is different from the colloquial usage of

Table 4 Calculated confidence intervals and p-values for *unstandardized* beta coefficients from six regressions

	MATE	I-SEA Micro	I-SEA Macro	I-SEA Human	GAENE	100-pt scale
College GPA	(.0137, .0274) (.003)	(.0124, .0266) (.006)	(.0156, .0301) (.002)	(.0085, .0249) (.043)	(.0062, .0220) (.074)	(.0170, .0439) (.024)
Course level	(.0057, .0163) (.038)	(.0012, .0121) (.225)	(.0058, .0169) (.043)	(.0033, .0160) (.128)	(.0045, .0166) (.083)	(.0105, .0311) (.045)
Parent education	(.0085, .0182) (.006)	(.0063, .0163) (.024)	(.0021, .0124) (.156)	(.0037, .0153) (.100)	(.0101, .0212) (.005)	(.0161, .0350) (.007)
Evolution understanding	(.0147, .0191) (.000)	(.0144, .0190) (.000)	(.0096, .0142) (.000)	(.0091, .0144) (.000)	(.0032, .0082) (.025)	(.0036, .0122) (.068)
NOS understanding	(.0082, .0116) (.000)	(.0089, .0124) (.000)	(.00504, .0074) (.001)	(.0067, .0108) (.000)	(.0040, .0079) (.002)	(.0066, .0130) (.003)
Religiosity	(−.0120, −.0083) (.000)	(−.0114, −.0076) (.000)	(−.0113, −.0075) (.000)	(−.0150, −.0106) (.000)	(−.0126, .0084) (.000)	(−.0220, .0149) (.000)
Catholic	(−.0188, .0095) (.743)	(−.0152, .0140) (.967)	(−.0146, .0153) (.983)	(−.0094, .0245) (.657)	(−.0078, .0247) (.605)	(−.0145, .0409) (.633)
Protestant/LDS	(−.0757, −.0456) (.000)	(−.0405, −.0093) (.110)	(−.0846, −.0527) (.000)	(−.0920, −.0558) (.000)	(−.0928, −.0582) (.000)	(−.1857, −.1266) (.000)
Muslim	(−.0715, −.0285) (.020)	(−.0312, .0132) (.684)	(−.0627, −.0173) (.079)	(−.1058, −.0542) (.002)	(−.0521, −.0027) (.268)	(−.1056, −.0214) (.132)
Other religion	(−.0534, −.0296) (.775)	(−.0284, .0572) (.736)	(−.0287, .0589) (.079)	(−.0314, .0681) (.712)	(−.0011, .0941) (.329)	(.0143, .1766) (.240)
Democrat	(.0158, .0331) (.005)	(.0110, .0289) (.026)	(.0070, .0253) (.078)	(.0183, .0391) (.006)	(.0162, .0360) (.009)	(.0096, .0435) (.117)
Republican	(−.0267, −.0025) (.227)	(−.0240, .0009) (.354)	(−.0370, −.0115) (.057)	(−.0508, −.0219) (.012)	(−.0386, −.0109) (.074)	(−.1066, −.0594) (.000)
Black/African Am.	(−.0381, .0033) (.401)	(−.0633, −.0206) (.050)	(−.0311, −.0125) (.671)	(−.0403, .0093) (.532)	(−.0409, .0066) (.471)	(−.0745, .0065) (.401)
Hispanic	(−.0080, .0149) (.762)	(.0001, .0237) (.314)	(−.0063, .0178) (.633)	(−.0277, −.0003) (.307)	(−.0277, −.0015) (.266)	(−.0543, .0097) (.152)
Asian	(−.0327, −.0086) (.087)	(−.0515, −.0277) (.002)	(−.0337, −.0083) (.098)	(−.0552, −.0263) (.005)	(−.0419, −.0143) (.042)	(−.0429, .0042) (.411)
Other race/ethnicity	(−.0097, .0233) (.679)	(−.0328, .0012) (.352)	(−.0209, .0138) (.837)	(−.0593, −.019803) (.046)	(−.0310, .0068) (.523)	(−.0582, .0063) (.421)

Each column is a separate regression using a different evolution acceptance instrument as the dependent variable. Each row is a predictor variable that was inputted into each regression. Each cell contains the calculated confidence interval and below is the *p* value for that variable in that particular regression. Italicized cells indicate that the coefficient was statistically significant at $p < .05$

the term “result” which suggests statistical significance. We consider differences in statistical significance across models more specifically as differences in *conclusions* rendered from the different *results* (coefficients). Would we have come to different conclusions about what variables are statistically significantly related to evolution acceptance, based on different evolution acceptance instruments? To answer this, we look to whether variables were deemed as statistically significant ($p < .05$) in each regression model.

Whether variables were statistically significant was different for different evolution acceptance instruments; some variables were significant and positive predictors across all instruments, some variables were significant but negative predictors of acceptance across all instruments, and finally some variables were significant predictors of acceptance with some instruments but not others. The directionality of the relationships between predictor

variables and acceptance was constant across all models; variables were never significant positive predictors in one model and then significant negative predictors in other models. See Table 6 for a summary of which variables were statistically significant across evolution acceptance instruments.

Religiosity was a statistically significant negative predictor of evolution acceptance across all instruments. Nature of Science (NOS) understanding was a statistically significant positive predictor of evolution acceptance across all evolution acceptance instruments (Table 6).

GPA was a positive significant predictor of evolution acceptance across almost all instruments, but GPA was not significant for GAENE scores. Evolution understanding was a significant positive predictor of evolution acceptance across all instruments except it was not a significant predictor of 100-pt scores. Identifying as Protestant or LDS was a significant negative predictor of

Table 5 Evaluation of confidence intervals of unstandardized beta coefficients from predictor variables across evolution acceptance instruments used to determine if results across analyses are different

	MATE/ Micro	MATE/ Macro	MATE/ Human	MATE/ GAENE	MATE/100	Micro/ Macro	Micro/ Human	Micro/ GAENE	Micro/100	Macro/ Human	Macro/ GAENE	Macro/100	Human/ GAENE	Human/100	GAENE/100	# of diff
College GPA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0
Course level	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0
Parent education	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	X	✓	2
Evolution understanding	✓	X	X	X	X	X	✓	X	X	✓	X	✓	X	✓	✓	9
NOS understanding	✓	X	✓	X	✓	X	✓	X	✓	✓	✓	✓	✓	✓	✓	4
Religiosity	✓	✓	✓	✓	X	✓	✓	✓	X	✓	✓	X	✓	X	X	4
Catholic	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0
Protestant/LDS	X	✓	X	✓	X	X	X	✓	X	✓	✓	X	✓	X	X	9
Muslim	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	X	✓	✓	2
Other religion	X	X	✓	X	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	4
Democrat	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0
Republican	✓	✓	✓	✓	X	✓	✓	✓	X	✓	✓	X	✓	X	X	5
Black/Af. Am.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0
Hispanic	✓	✓	✓	✓	X	✓	X	✓	X	✓	✓	X	✓	✓	✓	4
Asian	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0
Other race/ethnicity	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	3

Each column shows the comparison between two evolution acceptance instruments. Each row is a predictor variable that was included in every regression. Each cell indicates whether the confidence intervals of two instruments overlap. If the confidence intervals overlap, it is indicated with a "✓"; a "X" indicates that the two confidence intervals do not overlap and thus the beta coefficients are different from one another beyond sampling variability. The far right column summarizes the number of instances where confidence intervals do not overlap. Note that this is a conservative approach and those with overlapped confidence intervals may also be different (Schenker and Gentleman 2001)

Table 6 Summary of the statistical significance of predictor variables across linear regressions used to determine if conclusions were different across analyses

	MATE	Micro	Macro	Human	GAENE	100-pt	Same/ different conclusion
College GPA	.091	.088	.105	.065	.061	.075	Different
Course Level	.060	.037	.064	.047	.056	.063	Different
Parent Education	.085	.074	.048	.054	.097	.091	Different
Evo Understanding	.261	.264	.191	.161	.086	.068	Different
NOS Understanding	.192	.211	.124	.150	.112	.106	Same
Religiosity	-.263	-.252	-.252	-.293	-.266	-.267	Same
Catholic	-.014	-.002	.001	.021	.026	.023	Same
Protestant/LDS	-.184	-.077	-.215	-.197	-.223	-.263	Different
Muslim	-.079	-.015	-.065	-.111	-.042	-.056	Different
Other religion	-.008	.010	.011	.011	.032	.037	Same
Democrat	.086	.071	.058	.088	.089	.052	Different
Republican	-.039	-.031	-.066	-.085	-.064	-.122	Different
Black/African Am.	-.024	-.060	-.013	-.019	-.023	-.026	Same
Hispanic	.010	.037	.018	-.037	-.043	-.054	Same
Asian	-.053	-.103	-.056	-.093	-.071	-.028	Different
Other race/ethnicity	.012	-.029	-.007	-.062	-.021	-.026	Different

Middle columns are *standardized* regression coefficients for six regressions. Each row is a predictor variable included in each regression model. Variables that were statistically significant at $p < .05$ are italicized. Far right column summarizes whether conclusions were different across models ("same" indicates that the variable was always statistically significant or was always not significant; "different" indicates that the variable was statistically significant in some models but not others)

evolution acceptance for almost all instruments except the I-SEA microevolution instrument (Table 6).

Identifying as a race/ethnicity in the other race/ethnicity category was a negative significant predictor of scores from the I-SEA human evolution acceptance instrument, but no other instrument. Identifying as Catholic, other religion, Black/African American, or Hispanic was not a significant predictor of evolution acceptance scores for any evolution acceptance instrument (Table 6).

Because the effect of evolution understanding on evolution acceptance has been a topic of contention in the evolution education literature, and the observation that the beta coefficients were different across instruments for this variable, we decided to explore the extent to which evolution understanding scores could independently explain evolution acceptance scores across evolution acceptance instruments. We conducted simple linear regressions using evolution understanding as the sole predictor variable for each evolution acceptance instrument. Evolution understanding alone was a statistically significant predictor of evolution acceptance across all instruments (F range = 35.39–188.62, $p < .001$), but the amount of variance in evolution acceptance that could be explained by evolution understanding ranged from 4 to 20% across evolution acceptance instruments. See Table 7 for a summary of the results for these regressions looking at the independent effect of understanding

evolution on evolution acceptance across evolution acceptance instruments.

Research question #2 discussion Research results and conclusions exploring relationships between variables and evolution acceptance differed depending on the instrument used to measure evolution acceptance. First, the coefficients for many predictor variables differed depending on the evolution acceptance instrument, most notably for evolution understanding and identifying as a Protestant/LDS. Further, whether college GPA, course level, parent education, evolution understanding, identifying as a Protestant/LDS, Muslim, Democrat, Republican, or Asian and other race/ethnicity predicted evolution acceptance was different depending on the evolution acceptance instrument used. Similar to what has been reported in the literature, we saw that the strength and statistical significance of the relationship between evolution understanding and evolution acceptance was variable across instruments. Why does evolution understanding predict evolution acceptance differentially according to these different instruments?

First, many researchers use the MATE intending to measure only evolution acceptance, but the MATE has items that require understanding of evolution. So, a student categorized as fully "accepting evolution," according to the MATE, needs to have an accurate understanding of evolution, which could have led to stronger relationships

Table 7 Summary of simple linear regressions exploring the independent effect of evolution understanding on evolution acceptance across evolution acceptance instruments

	Unstandardized coefficients		Standardized coefficients	t	Sig	Model statistics			
	B	Std. error				Beta	R ²	F	Sig
Effects of evolution understanding									
On MATE	.029	.002	.449	13.676	< .0001	.20	187.04	.000	740
On Micro	.029	.002	.451	13.734	< .0001	.20	188.62	.000	740
On Macro	.021	.002	.334	9.641	< .0001	.11	92.94	.000	740
On Human	.023	.003	.320	9.198	< .0001	.10	84.60	.000	740
On GAENE	.015	.002	.227	6.340	< .0001	.05	40.20	.000	740
On 100-pt	.025	.004	.214	5.949	< .0001	.04	35.39	.000	740

Each row contains the results from a linear regression with scores from a different evolution acceptance instrument as the dependent variable, indicated by the row heading. Each coefficient statistic indicates the extent to which evolution understanding scores independently predicted evolution acceptance scores. Each adjusted R^2 value indicates the amount of variance in evolution acceptance scores that can be explained by evolution understanding scores alone

between understanding and accepting evolution. Our results revealed that evolution understanding had among the strongest coefficients when we used the MATE to measure evolution acceptance.

However, we found that evolution understanding had similar beta coefficients for I-SEA microevolution scores. The I-SEA was designed to avoid questions that measure evolution understanding, so the relationship between evolution understanding scores and scores seen on the I-SEA microevolution instrument is likely not due to the same conflation issues that may be occurring with the MATE (Nadelson and Southerland 2012; Smith et al. 2016). However, many people do not see conflict between their religious beliefs and microevolution (Scott 2005). Therefore, evolution understanding may be more strongly related to I-SEA microevolution scores because a higher understanding is able to impact acceptance more strongly in the absence of a belief/identity barrier. Indeed, we found that evolution understanding was more predictive of I-SEA microevolution scores than I-SEA macroevolution and human evolution scores. This reasoning aligns with prior research that shows that understanding is more related to acceptance of a topic when an identity barrier is not present (Kahan and Stanovich 2016; Weisberg et al. 2018).

For the I-SEA macroevolution, I-SEA human evolution, and GAENE, our results indicate that evolution understanding is a significant, but weaker, predictor of evolution acceptance. Further, when controlling for other variables, evolution understanding was not at all predictive of 100-pt acceptance scores. But how do these results compare with the findings from prior evolution education literature?

The results of this study have the potential to explain the contradictory findings we have seen in the evolution

education literature over the last 30 years. In fact, the results we found here are strikingly similar to the findings we have seen in the literature with regards to evolution understanding. Sinatra et al. (2003), Bishop and Anderson (1990), and Hermann (2012) who used instruments similar to the 100-point instrument, found no relationship between evolution acceptance and evolution understanding, similar to our results in this study. However, researchers who have used the MATE to measure evolution acceptance have often found large significant relationships between acceptance and evolution understanding (Rutledge and Warden 2000; Trani 2004), which is also similar to what we found in this study. Our study, which examines these instruments in the same student population, suggests that different patterns noted in these prior studies may not just be an artifact of different study populations.

Future research

There were several limitations to this study that warrant future research. This study was restricted to exploring the differences in measurement of evolution acceptance, but researchers studying evolution education also use very different instruments to measure evolution understanding and Nature of Science (NOS) understanding. We measured these predictor variables with instruments that have been used in several other studies, but results could be different if we used different instruments. For instance, research findings may differ if we had used the Conceptual Inventory of Natural Selection (CINS) (Anderson et al. 2002) or the Measure of Understanding of Macroevolution (MUM) (Nadelson and Southerland 2009) to measure evolution understanding instead of the EALS because understanding of natural selection may be related to evolution acceptance differently than understanding

of macroevolution (Nadelson and Southerland 2010). Additionally, research findings could have been different if we used the Views of Nature of Science (VNOS) instrument (Abd-El-Khalick et al. 2001) or the Evolution and Nature of Science (ENOS) instrument to measure student Nature of Science (NOS) understanding. Future research should explore how differential measurement of other constructs other than evolution acceptance may lead to different research findings.

We reported the results from multi-item evolution acceptance instruments in this manuscript. Single item instruments, such as those used by the Gallup or Pew polls (Gallup 2017; Pew 2013), have been used in evolution education studies that are published in high impact journals (Miller et al. 2006; Weisberg et al. 2018) and are often widely cited as statistics for the US public's evolution acceptance. An analysis of different single item instruments and how they may lead to different research findings would be of value to the field in future studies.

We treated Likert scale scores as interval data. Although this is a common practice in the analyses of Likert scale data, we recognize that from a measurement theory perspective this assumption may not stand (Hambleton et al. 1991).

Finally, we explored how multi-item evolution acceptance instruments operate in three different undergraduate student populations and the differences we found are likely dependent on the specific populations from which we sampled. Future studies should confirm that differences in instruments, particularly those measuring acceptance of microevolution, macroevolution, and human evolution are present in other populations of students.

Discussion and conclusion

Evolution education researchers have many different options to measure evolution acceptance, yet this study highlights that the choice of the instrument used to measure evolution acceptance can influence the results and conclusions of a study. We found that whether and how much a variable predicted evolution acceptance was dependent on the instrument used to measure evolution acceptance. Further, we also found that in some cases, and particularly when examining the sub-scales of the I-SEA, whether a researcher would find differences in levels of evolution acceptance across populations of students was dependent on the evolution acceptance instrument used. These findings highlight that the diversity of instruments used to measure evolution acceptance could be an underlying reason for conflicting results in the evolution education literature because people are

comparing studies that used different evolution acceptance instruments.

Our study was not designed to identify the “best” instrument to measure evolution acceptance, so we cannot give a recommendation for what we think researchers should use. However, this study does highlight the need for the evolution education community to be more specific about what is meant by “acceptance of evolution” and to take steps to improve the alignment of their research questions with their evolution acceptance instrument. Some of these instruments use specific items to elicit student acceptance of evolution, whereas some instruments ask broad questions that allow students to use their own definitions of what it means to accept evolution. The underlying issue may be that there are many different definitions of acceptance of evolution (Smith 2009a, b; Smith et al. 1995, 2016). If different evolution acceptance instruments were built using different definitions of evolution acceptance, then these instruments were built to measure different constructs, even if researchers all use the term “evolution acceptance.” Does evolution acceptance only include microevolution or does evolution acceptance also include macroevolution and human evolution (Nadelson and Southerland 2012)? To what extent does someone need to understand evolution and the Nature of Science (NOS) to accept evolution (Rutledge and Warden 1999)? To what extent does someone need to advocate for evolution to accept evolution (Smith et al. 2016)? Is it enough to simply let respondents define evolution for themselves in surveys or do the researchers need to define evolution for the respondent (Bishop and Anderson 1990; Sinatra et al. 2003)? Our study provides empirical evidence that it is important for researchers to be thoughtful about what the evolution education community actually wants to measure when they measure evolution acceptance. Different instruments will measure different aspects of evolution acceptance and may then influence the conclusions that researchers will draw.

We encourage the evolution education community to consider these issues and reach consensus on a definition of evolution acceptance and a common instrument to measure evolution acceptance, based on this consensus definition. Much could be learned by comparing evolution acceptance across different institutions and contexts, but our study indicates that researchers may not be able to use different instruments to do so. We argue for increased dialogue across the evolution education community about the relative strengths and weaknesses of different evolution acceptance instruments. While we are hesitant to give any further recommendations, we encourage other researchers to be critical of existing

instruments and not use an instrument's popularity alone in making a decision about which instrument to use.

Additional file

Additional file 1. Additional tables.

Abbreviation

NOS: Nature of Science.

Authors' contributions

MEB helped conceptualize the study, collected and interpreted data, performed statistical analyses, and drafted the manuscript. HMD helped conceptualize the study, collected data, performed statistical analyses, and aided in revising drafts of the manuscript. EAH helped conceptualize the study, collected and interpreted data, performed statistical analyses, and aided in revising drafts of the manuscript. Yi Zheng helped to plan the analyses, reviewed the analyses and statistics for critical content, and aided in revising drafts of the manuscript. SB helped conceptualize the study, interpret the data, and drafted the manuscript. All authors read and approved the final manuscript.

Author details

¹ School of Life Sciences, Biology Education Research Lab, Arizona State University, Tempe, AZ, USA. ² School of Biological Sciences, University of Northern Colorado, Colorado, CO, USA. ³ Mary Lou Fulton Teachers College & School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ, USA.

Acknowledgements

We would like to acknowledge Katelyn Cooper, Ryan Dunk, Logan Gin, and Dan Grunspan for their friendly reviews of the manuscript. We would also like to thank Gale Sinatra and Heath Ogden for their help gathering preliminary data during the early stages of the project.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data analyzed in this manuscript is available upon request from Sara Brownell (sara.brownell@asu.edu).

Ethics approval and consent to participate

The Institutional Review Board of Arizona State University and University of Northern Colorado approved the procedures for this study (ASU IRB #00007719; UNC IRB #1131916-2). Written informed consent was obtained by all participating students at the beginning of the study.

Funding

National Science Foundation (Grant Numbers DGE-1311230 and 1712188), which provided graduate student and postdoctoral support for this project.

Publisher's Note

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

Received: 4 October 2018 Accepted: 2 January 2019

Published online: 31 January 2019

References

Abd-El-Khalick F, Lederman NG, Bell RL, Schwartz RS. Views of Nature of Science Questionnaire (VNOS): toward valid and meaningful assessment of learners' conceptions of Nature of Science. 2011.

- Anderson DL, Fisher KM, Norman GJ. Development and evaluation of the conceptual inventory of natural selection. *J Res Sci Teach*. 2002;39(10):952–78. <https://doi.org/10.1002/tea.10053>.
- Athanasidou K, Papadopoulos P. Conceptual ecology of the evolution acceptance among Greek education students: knowledge, religious practices and social influences. *Int J Sci Educ*. 2012;34(6):903–24. <https://doi.org/10.1080/09500693.2011.586072>.
- Baker JO, Rogers D, Moser T. Acceptance of evolution among American Mormons. *J Contemp Relig*. 2018;33(1):123–34. <https://doi.org/10.1080/13537903.2018.1408295>.
- Barnes ME, Brownell SE. A call to use cultural competence when teaching evolution to religious college students: introducing religious cultural competence in evolution education (ReCCEE). *CBE Life Sci Educ*. 2017;16(4):es4. <https://doi.org/10.1187/cbe.17-04-0062>.
- Barnes ME, Evans EM, Hazel A, Brownell SE, Nesse RM. Teleological reasoning, not acceptance of evolution, impacts students' ability to learn natural selection. *Evolution*. 2017;10:7. <https://doi.org/10.1186/s12052-017-0070-6>.
- Barone LM, Petto AJ, Campbell BC. Predictors of evolution acceptance in a museum population. *Evolution*. 2014;7(1):1–11. <https://doi.org/10.1186/s12052-014-0023-2>.
- Bishop BA, Anderson CW. Student conceptions of natural selection and its role in evolution. *J Res Sci Teach*. 1990;27(5):415–27. <https://doi.org/10.1002/tea.3660270503>.
- Brem SK, Ranney M, Schindel J. Perceived consequences of evolution: college students perceive negative personal and social impact in evolutionary theory. *Sci Educ*. 2003;87(2):181–206. <https://doi.org/10.1002/sce.10105>.
- Carter BE, Wiles JR. Scientific consensus and social controversy: exploring relationships between students' conceptions of the nature of science, biological evolution, and global climate change. *Evolution*. 2014;7(1):6. <https://doi.org/10.1186/s12052-014-0006-3>.
- Cavallo AML, White KJ, McCall David. The mismatch among students' views about nature of science, acceptance of evolution, and evolutionary science understandings. *Sci Educ Rev*. 2011;10(2):37–42.
- Cohen AB, Shariff AF, Hill PC. The accessibility of religious beliefs. *J Res Person*. 2008;42(6):1408–17. <https://doi.org/10.1016/j.jrp.2008.06.001>.
- Deniz H, Donnelly LA, Yilmaz I. Exploring the factors related to acceptance of evolutionary theory among Turkish preservice biology teachers: toward a more informative conceptual ecology for biological evolution. *J Res Sci Teach*. 2008;45(4):420–43. <https://doi.org/10.1002/tea.20223>.
- Dunk RD, Petto AJ, Wiles JR, Campbell BC. A multifactorial analysis of acceptance of evolution. *Evolution*. 2017;10:4. <https://doi.org/10.1186/s12052-017-0068-0>.
- Field A. *Discovering statistics using SPSS*. Thousand Oaks: Sage Publications; 2009.
- Gallup. In US, belief in creationist view of humans at new low. <http://www.gallup.com/poll/210956/belief-creationist-view-humans-new-low.aspx>. Accessed 8 Aug 2017.
- Glaze AL, Goldston MJ. U.S. Science teaching and learning of evolution: a critical review of the literature 2000–2014. *Sci Educ*. 2015;99(3):500–18. <https://doi.org/10.1002/sce.21158>.
- Glaze AL, Goldston MJ, Dantzer J. Evolution in the Southeastern USA: factors influencing acceptance and rejection in pre-service science teachers. *Int J Sci Math Educ*. 2014;13(6):1189–209. <https://doi.org/10.1007/s10763-014-9541-1>.
- Großschedl J, Konnemann C, Basel N. Pre-service biology teachers' acceptance of evolutionary theory and their preference for its teaching. *Evolution*. 2014;7(1):1–16.
- Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of item response theory*. Newbury Park: Sage Publications Inc; 1991.
- Hawley PH, Short SD, McCune LA, Osman MR, Little TD. What's the Matter with Kansas?: the development and confirmation of the evolutionary attitudes and literacy survey (EALS). *Evolution*. 2010;4(1):117–32. <https://doi.org/10.1007/s12052-010-0294-1>.
- Hermann RS. Cognitive apartheid: on the manner in which high school students understand evolution without believing in evolution. *Evolution*. 2012;5(4):619.
- Ingram EL, Nelson CE. Relationship between achievement and students' acceptance of evolution or creation in an upper-level evolution course. *J Res Sci Teach*. 2006;43(1):7–24. <https://doi.org/10.1002/tea.20093>.

- Johnson RL, Peeples EE. The role of scientific understanding in college: student acceptance of evolution. *Am Biol Teach*. 1987;49(2):93–8. <https://doi.org/10.2307/4448445>.
- Kahan DM, Stanovich KE (2016) Rationality and belief in human evolution (SSRN Scholarly Paper No. ID 2838668). Rochester: Social Science Research Network. <https://papers.ssrn.com/abstract=2838668>.
- Lawson AE. Predicting science achievement: the role of developmental level, disembedding ability, mental capacity, prior knowledge, and beliefs. *J Res Sci Teach*. 1983;20(2):117–29.
- Lloyd-Strovas JD, Bernal XE. A review of undergraduate evolution education in U.S. Universities: building a unifying framework. *Evolution*. 2012;5(3):453–65. <https://doi.org/10.1007/s12052-012-0435-9>.
- Manwaring KF, Jensen JL, Gill RA, Bybee SM. Influencing highly religious undergraduate perceptions of evolution: mormons as a case study. *Evolution*. 2015;8(1):23.
- Mead R, Hejmadi M, Hurst LD. Scientific aptitude better explains poor responses to teaching of evolution than psychological conflicts. *Nat Ecol Evol*. 2018;2(2):388–94. <https://doi.org/10.1038/s41559-017-0442-x>.
- Metzger KJ, Montplaisir D, Haines D, Nickodem K. Investigating undergraduate health sciences students' acceptance of evolution using MATE and GAENE. *Evolution*. 2018;11(1):10.
- Miller JD, Scott EC, Okamoto S. Public acceptance of evolution. *Science*. 2006;313(5788):765–6. <https://doi.org/10.1126/science.1126746>.
- Nadelson LS, Hardy KK. Trust in science and scientists and the acceptance of evolution. *Evolution*. 2015;8(1):9.
- Nadelson LS, Sinatra GM. Educational professionals' knowledge and acceptance of evolution. *Evol Psychol*. 2009;7(4):490–516. <https://doi.org/10.1177/147470490900700401>.
- Nadelson LS, Southerland SA. Development and preliminary evaluation of the measure of understanding of macroevolution: introducing the MUM. *J Exp Educ*. 2009;78(2):151–90. <https://doi.org/10.1080/00220970903292983>.
- Nadelson LS, Southerland SA. Examining the interaction of acceptance and understanding: how does the relationship change with a focus on macroevolution? *Evolution*. 2010;3(1):82–8. <https://doi.org/10.1007/s12052-009-0194-4>.
- Nadelson LS, Southerland S. A more fine-grained measure of students' acceptance of evolution: development of the inventory of student evolution acceptance—I-SEA. *Int J Sci Educ*. 2012;34(11):1637–66. <https://doi.org/10.1080/09500693.2012.702235>.
- Pew. Public's views on human evolution. 2013. <http://www.pewforum.org/2013/12/30/publics-views-on-human-evolution/>. Accessed 31 May 2018.
- Pew. How religious is your state? 2016. <http://www.pewresearch.org/fact-tank/2016/02/29/how-religious-is-your-state/>. Accessed 17 Nov 2017.
- Pobiner B. Accepting, understanding, teaching, and learning (human) evolution: obstacles and opportunities. *Am J Phys Anthropol*. 2016;159(S61):S232–74. <https://doi.org/10.1002/ajpa.22910>.
- Pobiner B, Beardsley PM, Bertka CM, Watson WA. Using human case studies to teach evolution in high school AP biology classrooms. *Evolution*. 2018;11(1):3.
- Rice JW, Olson JK, Colbert JT. University evolution education: the effect of evolution instruction on biology majors' content knowledge, attitude toward evolution, and theistic position. *Evolution*. 2010;4(1):137–44. <https://doi.org/10.1007/s12052-010-0289-y>.
- Romine WL, Walter EM, Bosse E, Todd AN. Understanding patterns of evolution acceptance—a new implementation of the measure of acceptance of the theory of evolution (MATE) with Midwestern university students. *J Res Sci Teach*. 2017;54(5):642–71.
- Romine WL, Todd AN, Walter EM. 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. *Evo Edu Outreach*. 2018;11:17.
- Rutledge ML, Sadler KC. Reliability of the measure of acceptance of the theory of evolution (MATE) instrument with university students. *Am Biol Teach*. 2007;69(6):332–5.
- Rutledge ML, Warden MA. The development and validation of the measure of acceptance of the theory of evolution instrument. *Sch Sci Math*. 1999;99(1):13–8. <https://doi.org/10.1111/j.1949-8594.1999.tb17441.x>.
- Rutledge ML, Warden MA. Evolutionary theory, the nature of science & high school biology teachers: critical relationships. *Am Biol Teach*. 2000;62(1):23–31. [https://doi.org/10.1662/0002-7685\(2000\)062%5b0023:ETNOS%5d2.0.CO;2](https://doi.org/10.1662/0002-7685(2000)062%5b0023:ETNOS%5d2.0.CO;2).
- Sbeglia GC, Nehm RH. Measuring evolution acceptance using the GAENE: influences of gender, race, degree-plan, and instruction. *Evo Edu Outreach*. 2018;11:18.
- Schenker N, Gentleman JF. On judging the significance of differences by examining the overlap between confidence intervals. *Am Stat*. 2001;55(3):182–6.
- Schleith D. A study of central Florida College Students' acceptance of the theory of evolution, microevolution, macroevolution, and human evolution. 2017.
- Schwabb JJ. The concept of the structure of a discipline. In: Herbert LJ, Murphy W, editors. *Structure in the social studies*. Washington, DC: National Council for the Social Studies; 1968. p. 43–56.
- Scott EC. *Evolution vs. creationism: an introduction*. California: University of California Press; 2005.
- Short SD, Hawley PH. The effects of evolution education: examining attitudes toward and knowledge of evolution in college courses. *Evol Psychol*. 2015;13(1):67–88.
- Sickel AJ, Friedrichsen P. Examining the evolution education literature with a focus on teachers: major findings, goals for teacher preparation, and directions for future research. *Evolution*. 2013;6(1):23. <https://doi.org/10.1186/1936-6434-6-23>.
- Sinatra GM, Southerland SA, McConaughy F, Demastes JW. Intentions and beliefs in students' understanding and acceptance of biological evolution. *J Res Sci Teach*. 2003;40(5):510–28. <https://doi.org/10.1002/tea.10087>.
- Smith MU. Current status of research in teaching and learning evolution: i. philosophical/epistemological issues. *Sci Educ*. 2009a;19(6–8):523–38. <https://doi.org/10.1007/s11191-009-9215-5>.
- Smith MU. Current status of research in teaching and learning evolution: ii. Pedagogical issues. *Sci Educ*. 2009b;19(6–8):539–71. <https://doi.org/10.1007/s11191-009-9216-4>.
- Smith MU, Siegel H, McInerney JD. Foundational issues in evolution education. *Sci Educ*. 1995;4(1):23–46. <https://doi.org/10.1007/BF00486589>.
- Smith MU, Snyder SW, Devereaux RS. The GAENE—generalized acceptance of evolution evaluation: development of a new measure of evolution acceptance. *J Res Sci Teach*. 2016. <https://doi.org/10.1002/tea.21328>.
- Trani R. I won't teach evolution; it's against my religion. And now for the rest of the story. *Am Biol Teach*. 2004;66(6):419–27. [https://doi.org/10.1662/0002-7685\(2004\)066%5b0419:IWTEAM%5d2.0.CO;2](https://doi.org/10.1662/0002-7685(2004)066%5b0419:IWTEAM%5d2.0.CO;2).
- Walter EM, Halverson KM, Boyce CJ. Investigating the relationship between college students' acceptance of evolution and tree thinking understanding. *Evolution*. 2013;6(1):26. <https://doi.org/10.1186/1936-6434-6-26>.
- Weisberg DS, Landrum AR, Metz SE, Weisberg M. No missing link: knowledge predicts acceptance of evolution in the United States. *Bioscience*. 2018;68(3):212–22.
- Wiles JR, Alters B. Effects of an educational experience incorporating an inventory of factors potentially influencing student acceptance of biological evolution. *Int J Sci Educ*. 2011;33(18):2559–85. <https://doi.org/10.1080/09500693.2011.565522>.
- Willis GB. *Cognitive interviewing: a tool for improving questionnaire design*. Thousand Oaks: Sage Publications; 2004.