Research Article

Understanding Patterns of Evolution Acceptance—A New Implementation of the Measure of Acceptance of the Theory of Evolution (MATE) With Midwestern University Students

William L. Romine,[1] Emily M. Walter,[2] Ephiram Bosse,[2] and Amber N. Todd[3]

[1]*Department of Biological Sciences, Wright State University, 3640 Colonel Glenn Hwy., Dayton 45435, Ohio*
[2]*California State University—Fresno, M/S SB 73, Fresno, California, 93740*
[3]*Boonshoft School of Medicine, Office of Medical Education, 290 White Hall, Wright State University, Dayton, Ohio*

Received 5 February 2016; Accepted 12 November 2016

Abstract: We validate the Measure of Acceptance of the Theory of Evolution (MATE) on undergraduate students using the Rasch model and utilize the MATE to explore qualitatively how students express their acceptance of evolution. At least 24 studies have used the MATE, most with the assumption that it is unidimensional. However, we found that the MATE is best used as two separate dimensions. When used in this way, the MATE produces reliable ($\alpha > 0.85$) measures for (i) acceptance of evolution facts and data and (ii) acceptance of the credibility of evolution and rejection of non-scientific ideas. Using *k*-means cluster analysis, we found students express their acceptance of evolution in five distinct profiles: (i) uniform high acceptance; (ii) uniform moderate acceptance; (iii) neutral acceptance; (iv) acceptance of facts, but rejection of credibility; and (v) rejection of both facts and credibility. Furthermore, we found that knowledge of macroevolution moderately explains students' acceptance profiles, corroborating previous claims that teaching macroevolution may be one way to improve students' acceptance. We use these findings to express the first set of operational definitions of evolution acceptance and propose that educators continue to explore additional ways to operationalize evolution acceptance. J Res Sci Teach 9999:XX–XX, 2016
Keywords: evolution acceptance; college; biology; natural selection; assessment; Rasch model; clustering

Despite the prevalence of anti-evolution views in the United States, American education policy emphasizes the importance of evolution. Evolution is one of four Disciplinary Core Ideas in life sciences in the Next Generation Science Standards (NGSS Lead States, 2013) and both the National Association of Biology Teachers (NABT) and AAAS have formal statements identifying evolution as a foundational component of scientific literacy (Brewer & Smith, 2011; NABT, 2002). Evolution is also one of five core content areas outlined in *Vision and Change in Undergraduate Biology Education* (Brewer & Smith, 2011) which is

William L. Romine and Emily M. Walter contributed equally to this work.
Correspondence to: E. M. Walter; E-mail: ewalter@csufresno.edu
DOI 10.1002/tea.21380
Published online in Wiley Online Library (wileyonlinelibrary.com).

a testament not only to the importance of evolution as a foundation for understanding the biological sciences, but also to the deficiency in understanding and acceptance of evolution at the post-secondary level.

Although 98% of American Association for the Advancement of Science (AAAS) scientists agree that humans and other living things have evolved over time, only two-thirds (65%) of the American public agrees (Pew Research Center, 2015). Americans of all ages and levels of education struggle with accepting evolution as driven by natural selection; that is, they do not consider evolution to be the best explanation for the diversity of life on Earth. This includes high school students (e.g., Flanagan & Roseman, 2011), high school biology teachers (e.g., Berkman & Plutzer, 2011), pre-service teachers (e.g., Losh & Nzekwe, 2011), and college science majors and non-science majors (e.g., Dagher & BouJaoude, 1997; Ingram & Nelson, 2006; Nadelson & Southerland, 2010b). Although low levels of acceptance are found in other countries, the United States ranks among the lowest in acceptance among developed countries worldwide (Miller, Scott, & Okamoto, 2006). Public agreement with evolution wanes further when considering American acceptance of human evolution. Some estimates suggest that as few as one in five college educated American adults agree that humans evolve (Lovely & Kondrick, 2008).

Valid and reliable assessments are key for measuring the impact of any educational initiative (Brewer & Smith, 2011), including those designed to address evolution literacy at the post-secondary level. One popular measurement tool which has been used extensively at both the secondary and post-secondary levels is the Measure of Acceptance of the Theory of Evolution (MATE; Rutledge & Sadler, 2007; Rutledge & Warden, 1999), a 20-item Likert-style instrument.

The MATE has been used frequently in the evolution acceptance literature as a tool for measuring students' evolution acceptance; it has been used as a measurement tool in at least 24 studies (Table 1).

Although a thorough meta-analysis of these studies is beyond the scope of this article, Table 1 is shown to document the extent to which the MATE has been utilized since the turn of the century, even in light of extensive criticism. Although Rutledge and Sadler (2007) report the MATE as internally consistent ($\alpha = 0.94$) with strong test–retest reliability ($r = 0.92$), the MATE is not without criticism. These concerns surround four primary issues (Hogan, 2000; Smith, 2010; Smith, Snyder, & Devereaux, 2016; Wagler & Wagler, 2013): (i) Rutledge and coworkers did not provide an operational definition of *acceptance* as assessed by the MATE; (ii) the MATE may conflate acceptance with other knowledge domains and/or religious beliefs; (iii) the MATE has inadequate construct validation; and (iv) MATE items load onto unpredictable dimensions.

Fortunately, many of these concerns can be addressed by examining the MATE through Rasch analysis of item- and test-level construct validity. We were interested in whether the MATE could measure multiple dimensions of evolution acceptance. We could then validate these constructs and operationalize what form(s) of evolution acceptance the MATE measured, thereby addressing some of the concerns in the literature. Furthermore, given moderate correlations between content knowledge and evolution acceptance (Nadelson & Southerland, 2010b; Walter, 2013), we also explored how knowledge of macroevolution relates to MATE acceptance dimension(s).

Purpose of the Research

In this study, we explore the validity of the Measure of Acceptance of the Theory of Evolution (MATE; Rutledge & Sadler, 2007; Rutledge & Warden, 1999) and relate patterns in the data to current theoretical models of evolution acceptance

Table 1
Summary of studies which have employed the Measure of Acceptance of the Theory of Evolution (MATE; Rutledge & Warden, 1999), including their sample sizes, populations, and Cronbach's alpha (if reported)

Studies Which Have Employed the MATE	*N*	Population	Cronbach's Alpha (if Reported)
Athanasiou and Papadopoulou (2012)	81	K	0.79 (pre); 0.87 (post)
Barone, Petto, and Campbell (2014)	259	H	0.96
Coleman et al. (2015)	164	F	
Deniz and Sahin (2016)	120	F	0.90
Deniz, Cetin, and Yilmaz (2011)	147	F	0.93
Deniz et al. (2008)	132	B	0.92
Donnelly et al. (2009)	33	A	0.94
Glaze, Goldston, and Dantzler (2014)	115	F	
Großschedl, Konnemann, and Basel (2014)	180	F	
Ha et al. (2012)	124	F	0.94
Ha et al. (2015)	28	B, J	0.865 (pre); 0.928 (post)
Korte (2003)	87	B	
Manwaring et al. (2015)	1,104	F	0.914
Moore and Cotner (2009)*	728	C	
Nadelson and Sinatra (2009)	337	E	0.96
Nadelson and Southerland (2010b)	741	C, D	0.93
Nehm and Schonfeld (2007)	253	D	
Rice, Olson, and Colbert (2015)	309	I	
Rissler, Duncan, and Caruso (2014)	2,999	C, D	0.96
Rutledge and Sadler (2007)	61	C	0.941
Walter (2013)	268	C	
Wiles and Alters (2011)	81	A	
Wiley (2003)	47	B	
Yousuf, bin Daud, and Nadeem (2011)*	299	G	0.87

A, high school students; B, high school teachers; C, undergraduate non-science majors; D, undergraduate science majors; E, education graduate students and professionals; F, pre-service teachers; G, medical students; H, museum visitors; I, university faculty; J, elementary teachers; K, pre-service early childhood educators.
*Used only a subset of the original MATE items.

(Deniz, Donnelly, & Yilmaz, 2008; Ha, Haury, & Nehm, 2012). Although the MATE has been used extensively in evolution education research (see Table 1), it has not been validated with Rasch analysis. We, therefore, examine patterns in MATE data by exploring how item measures fit with Rasch expectations, explore the MATE as a unidimensional measure of evolution acceptance, consider patterns in acceptance produced by the MATE in respect to current theoretical models of acceptance, and highlight the relationship between Rasch measures of evolution knowledge and acceptance of evolution.

We address the following research questions:

(1) How do the MATE items and the MATE scale conform to Rasch validity expectations?
(2) What patterns exist regarding how college students express their acceptance of evolution, and how generalizable are these patterns across multiple universities?
(3) What are key similarities and differences among the patterns of evolution acceptance expressed by college students?
(4) What is the role of content knowledge in college students' acceptance of evolution?

Review of Literature

Methods of Measuring Evolution Acceptance

Researchers have been examining evolution acceptance for decades. In the United States, semi-regular national surveys of evolution acceptance include polls by Gallup (e.g., Newport, 2004) and the Pew Research Center (2015). In the research literature, both qualitative and quantitative methods have been used to examine evolution acceptance. Some studies explore evolution acceptance through interviews (e.g., Donnelly, Kazempour, & Amirshokoohi, 2009; Nehm & Reilly, 2007) or open-ended response instruments (e.g., Nehm & Schonfeld, 2007; Robbins & Roy, 2007). Others use single survey items to measure acceptance (Lawson, 1983; McKeachie, Lin, & Strayer, 2002; Rutledge & Mitchell, 2002), or design new surveys without documenting their psychometric properties (e.g., Crivellaro & Sperduti, 2014).

Evolution acceptance instruments with documented validity and reliability include Johnson and Peeples (1987), the MATE (Rutledge & Warden, 1999), an unpublished instrument by Ingram and Nelson (2006), the Inventory of Student Evolution Acceptance (I-SEA, Nadelson & Southerland, 2012), the Evolutionary Attitudes and Literacy Survey (EALS; Hawley, Short, McCune, Osman, & Little, 2011), and the Generalized Acceptance of Evolution Evaluation (GAENE; Smith et al., 2016). Smith et al. (2016) created a summary table to compare the development characteristics of the EALS, I-SEA, MATE, and GAENE. This summary includes validity and reliability of results established by the original authors of the instruments, but does not discuss results produced by the instruments in subsequent studies.

As the MATE was one of the only widely available evolution acceptance instruments available for 10+ years (1999–2011), it has been one of the most common methods to measure evolution acceptance. The MATE has been used in at least 24 studies since its publication and has well-documented reliability in studies beyond Rutledge and coworkers (Table 1). For this reason, we chose to explore properties of the MATE, both for enhancing our understanding of the nature of evolution acceptance and so that others may consider their MATE results in novel ways.

Development of the MATE

Rutledge and Warden (1999) developed and validated the MATE using Classical Test Theory (CTT). It uses a 20-item Likert-scale to measure acceptance of six evolution concepts: (i) the process of evolution; (ii) the scientific validity of evolutionary theory; (iii) the evolution of humans; (iv) evidence of evolution; (v) how scientists view evolution; and (vi) the age of the Earth. Although these topics are discussed on the MATE, they should not be viewed as separate constructs. Rather, these elements contribute to a single score for "evolution acceptance."

The original publication of the MATE (Rutledge & Warden, 1999) examined its reliability and validity for measuring evolution acceptance of high school teachers ($N = 552$). Content validity was established by asking five university professors with expertise in evolutionary biology to rate each item on a scale of 1 (strongly disagree) to 5 (strongly agree), and only items with a rating higher than 3.5 were included. Validation included exploratory factor analysis to establish dimensionality and an internal consistency definition of reliability. The MATE was found to be unidimensional (principle components analysis revealed 71.6% variance explained by one factor) and produce reliable results ($\alpha_c = 0.98$; Rutledge & Warden, 1999). Rutledge and Warden called for future researchers to strengthen the MATE through subsequent analyses, as instrument development is always a work in progress.

Rutledge and Sadler (2007) have since validated the MATE as a measure of evolution acceptance for post-secondary students (non-biology majors; $N = 61$). Reliability of their results was evaluated through internal consistency (Cronbach's alpha) and test–retest (Pearson's r)

perspectives. They found the MATE produced results with high internal consistency ($\alpha = 0.94$) and high test–retest reliability ($r = 0.92$). Other authors beyond the Rutledge research group have also calculated reliability of the MATE for a variety of populations. Generally, Cronbach's alpha for the MATE is above 0.90 (Table 1), and no study has found the MATE to have a Cronbach's alpha less than 0.79 (value for pre-test MATE scores of early childhood education teachers; Athanasiou & Papadopoulou, 2012).

Theoretical Models of Evolution Acceptance

The relationship between knowledge of evolution and evolution acceptance is not straightforward. Some studies suggest little or no relationship between natural selection knowledge and evolution acceptance (e.g., Bishop & Anderson, 1990; Butler, 2009; Demastes, Settlage, & Good, 1995; Lawson & Worsnop, 1992). Other studies suggest a significant relationship between acceptance and knowledge of macroevolutionary concepts (Nadelson & Southerland, 2010b; Walter, 2013; Walter, Halverson, & Boyce, 2013).

Regardless of a clear relationship with evolutionary knowledge, the literature documents a complicated picture of evolution acceptance. There are at least two theoretical models that explain the nature of an individual's evolution acceptance. We begin this discussion with a description of the *conceptual ecology of evolution thinking* developed by Deniz et al. (2008). Deniz et al. (2008) developed their model by considering cognitive, affective, and contextual domains related to how people think about evolution, building upon the revisionist theory of conceptual change (Strike & Posner, 1992). To construct their framework, they used multiple regression to frame evolution thinking across multiple variables (domains). They placed evolution acceptance (noted as a component of evolution thinking) as an output of both cognitive and affective domains, noting that acceptance was an embodiment of *plausibility* conditions for conceptual change (i.e., an idea with explanatory power is plausible; Strike & Posner, 1992).

Ha et al. (2012) focus their theoretical model (Figure 1 in that manuscript) on the nature of evolution acceptance (as opposed to evolution thinking in general). Their model takes a neurological approach to examining evolution acceptance, describing that acceptance is mediated through two paths: (i) an conscious, reflective thinking process (mediated by reasoning and understanding) and (ii) a separate, non-conscious pathway that gives rise to passive, intuitive feelings of certainty (a new, previously unexplored variable). Their data support the idea that both active and intuitive lines of thinking can inform individuals' evolution acceptance.

Our Theoretical Approach. As we considered the nature of evolution acceptance and how to measure it, we wondered if existing theoretical frameworks could explain patterns in evolution acceptance data produced by the MATE. To consider this question, we drew from traditions of grounded theory as we approached our study of the MATE. We chose this approach as we did not see the Deniz et al. (2008) and Ha et al. (2012) models as necessarily different. One model placed acceptance between cognitive and affective domains of evolutionary thinking and the other established acceptance as an output of active thinking or passive, intuitive feelings of certainty. We wondered if evolution acceptance as informed by cognitive and affective domains (Deniz et al., 2008) and the two paths described by Ha et al. (2012) could be overlapping ideas. For example, could feelings of certainty (Ha et al., 2012) be an element of affect-driven acceptance (Deniz et al., 2008; Strike & Posner, 1992) and/or could active, conscious thinking mediated by reasoning and understanding (Ha et al., 2012) be an element of cognitively driven acceptance (Deniz et al., 2008; Strike & Posner, 1992)? Regardless of how the frameworks interact, it seemed plausible that elements of evolution acceptance described by the theoretical models could manifest in a set of evolution acceptance data.

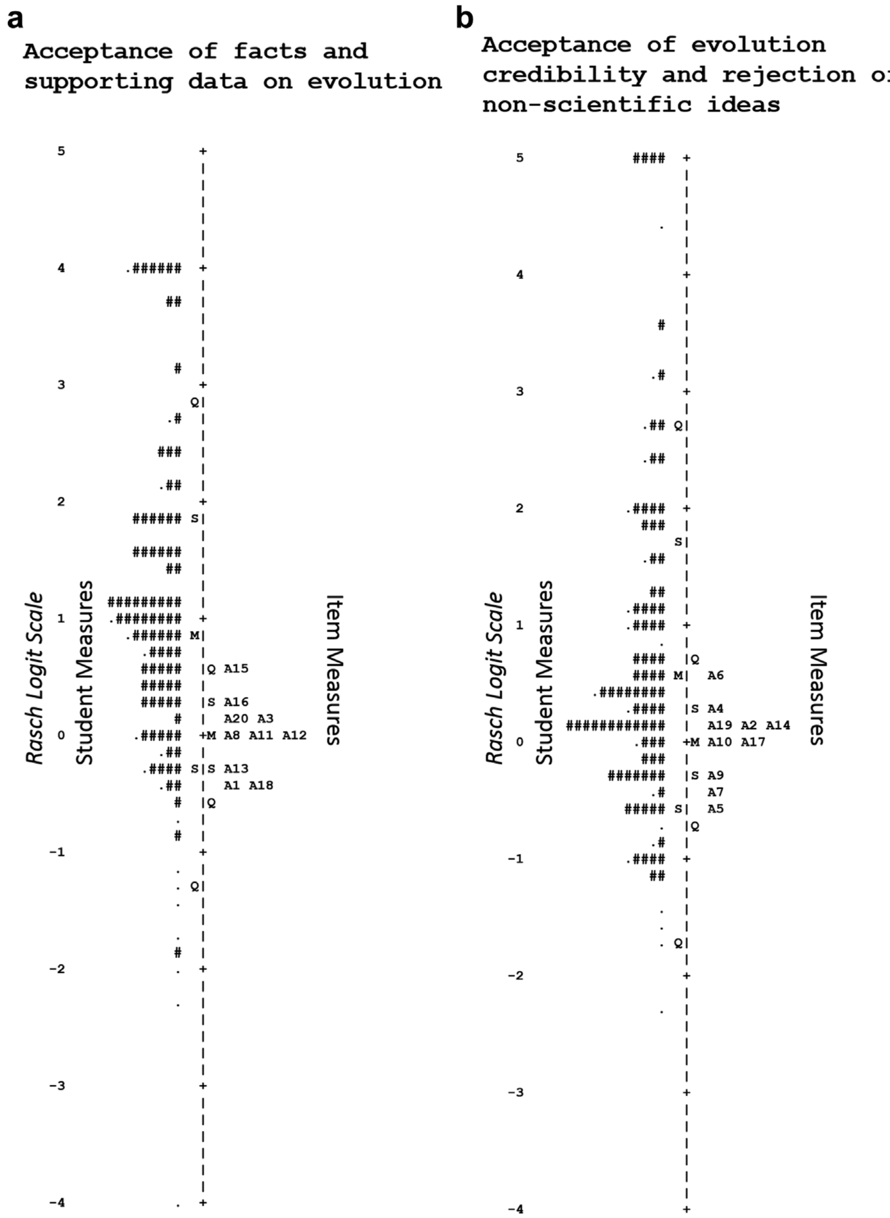

Figure 1. Person-item maps of the two MATE scales: (**a**) acceptance of facts and supporting data around evolution and (**b**) acceptance of evolution credibility and rejection of non-scientific ideas.

In our case, we wondered how patterns in the data produced by the MATE might relate to the theoretical literature. However, in lieu of the possibility that existing acceptance theories may overlap, we decided to approach our study using an inductive approach, without strict adherence to a given theoretical framework. In the tradition of grounded theory (Glaser, 1978, 1992; Glaser & Strauss, 1967, 2009; Strauss, 1987), we allowed for patterns in the acceptance to emerge from our analysis, thereby building empirical support for the nature of evolution acceptance as measured by the MATE. This led us to explore the MATE with a Rasch framework, so that we could explore if evolution acceptance was a single, unidimensional construct (as originally documented by Rutledge & Warden, 1999) or if evolution acceptance was a multidimensional construct, with dimensions potentially attributable to current or new theoretical models. This also led us to use clustering, an unsupervised quantitative approach that focuses on identifying qualitative groups in the data without reference to an external variable or criterion.

Contextualizing This Study

Educational researchers often select and trust the MATE as a valid measure of acceptance of evolution based on its high reliability (Table 1). As no study has documented reliability for the MATE below 0.79 (Athanasiou & Papadopoulou, 2012), it is likely capable of producing results that are precise (Nunnally, 1978) and more than sufficient for individual comparisons (Tennant & Connaghan, 2007). Furthermore, the sum score provided by the MATE has traditionally been treated as a unidimensional measure of evolution acceptance.

Given the criticisms of the MATE as an overall instrument (Hogan, 2000; Smith, 2010; Smith et al., 2016; Wagler & Wagler, 2013), and the domains of acceptance described by current theoretical models of evolutionary thinking (Deniz et al., 2008; Ha et al., 2012), it is possible that correlations of evolution acceptance and knowledge (e.g., Nadelson & Southerland, 2010b; Walter, 2013) are confounded. In other words, the correlations may be artifacts produced by the MATE measuring more than one facet of acceptance, or possibly by the MATE measuring evolution knowledge. Additionally, the theoretical models supported the possibility of multi-dimensionality, as they describe evolution acceptance as both an output of cognitive and affective domains (Deniz et al., 2008) or as influenced by active thinking and passive feelings of uncertainty (Ha et al., 2012). To address these ideas, it was our goal in this study to explore the dimensionality of MATE through a Rasch framework, provide insight into the data using theoretical models of evolution acceptance, and investigate some of the criticisms of the MATE.

Methods

Context and Participants

We administered the MATE and a revised, revalidated version of the Measure of Understanding of Macroevolution (MUM; Nadelson & Southerland, 2010a; Romine & Walter, 2014; Re-MUM Walter & Romine, in preparation) to a convenience sample of undergraduate students from the Midwest United States ($N = 194$). Since evolution is a critical and foundational component of scientific literacy (Brewer & Smith, 2011; Bybee, 1997; National Association of Biology Teachers, 2010), and the general education biology course is one of the final opportunities to influence scientific literacy of college educated individuals, it is vital to understand what non-science majors know and how they learn about evolution in general education biology courses. Forty-four students were sampled from an urban public research-intensive university (enrolling approximately 12,000 undergraduate students) in the Midwestern United States. The student body at this university is about 80% White. Minority students hail from African American (10%), Hispanic (3%), Asian (3%), and multi-racial (3%) backgrounds. About 48% of the student body is male, and 52% is female. The remaining 150 students were sampled from a rural private liberal arts college (enrolling approximately 2,000 students) in the Midwestern United States. The student body is about 70% White. A majority of nonwhite students hail from African American (15%), Hispanic (5%), and Asian (3%) backgrounds. About 55% of students are male, and 45%

are female. Of the 194 students sampled, 51 reported majoring in biology and 143 were pursuing other majors.

Cross-validation of the clustering model (research question 2) was undertaken on a sample of 1,045 students from a large, public, research-extensive university enrolling over 20,000 students in the western United States. This sample was quite different from the previous sample with regard to ethnic background. A majority of responding students were Hispanic/Latino (52%). Minority students included Asian (20%), White (20%), and African American (2%); the remaining students reported other ethnicities. The gender distribution was similar to the previous schools, however, at 41% male and 59% female.

RQ1: How Do the MATE Items and the MATE Scale Conform to Rasch Validity Expectations?

There are many aspects of measurement validity to consider when evaluating an instrument (Popham & Popham, 2005). Since there are no published evaluations of the MATE's construct validity under Rasch or Item Response Theory (IRT) frameworks, our validity analyses focused on testing construct validity with respect to the Rasch model. Specifically, we focused on teasing out the dimensionality of the MATE and then evaluating the validity of its individual items with respect to the Rasch validity framework (Wright & Stone, 1979). We note that Deniz et al. (2008) used Rasch student ability measures, but they used the classical factor analytical approach to make the case that the MATE was reliable and unidimensional.

Rasch Analysis. In *BIGSTEPS*, we used the Rasch partial credit model (Masters, 1982) to provide a framework to quantify item- and scale-level construct validity. Since the purpose of this part of the study was validation, we preferred the Rasch model over the comparatively data-dependent traditions of Item Response Theory (IRT) and Classical Test Theory (CTT). Using the Rasch model to calibrate a test or survey is like calibrating a machine in the lab based on an accepted standard. The goal of the Rasch model is not to fit the data, but rather to evaluate how data fit with what would be expected from a well-constructed measurement tool. Specifically, the probability of selecting a higher level of acceptance of evolution should be proportional only to the difference between the item's agreeability and the student's level of acceptance (Wright & Stone, 1979). To date, the validity of the MATE has not been evaluated in this way; hence the Rasch model will give us new insights into the validity of the MATE based on this accepted standard. Through fit of our data with the Rasch partial credit model, we can proceed to evaluate the usefulness of the MATE in generating measures of evolution acceptance for undergraduate students using a data-independent criterion which is accepted as true by our community (Boone & Scantlebury, 2006; Boone, Townsend, & Staver, 2011).

The usefulness of individual items was evaluated using mean squares infit and outfit statistics, which have expected values of 1.0, but can fall between 0.5 and 1.5 for items with constructive measurement properties (Wright & Linacre, 1996). A fit value near 1.0 indicates that the item on the MATE is useful in helping the researcher distinguish between students of high and low acceptance, but also does not contain wording biases favoring students with high acceptance (indicated by a fit below 0.5) which may artificially inflate the MATE's reliability, giving the researcher an overestimation of its usefulness as a measurement tool (Masters, 1988). Test-level structural validity was also deduced from the Rasch model; namely, the extent to which the measurement model used for the MATE is unidimensional and locally independent. Unidimensionality indicates that a single score is sufficient to explain all meaningful variation between a set of items. Items that are locally independent are related only through acceptance of evolution and not other variables (Collins & Lanza, 2010). The hypothesis of unidimensionality was investigated by observing variance in the data unaccounted for by the Rasch model using principal components analysis on residuals. Residual correlations for individual items were used to quantify local dependency between items.

RQ2: What Patterns Exist Regarding How College Students Express Their Acceptance of Evolution?

Clustering of Response Patterns. Although Rasch analysis is a method of mapping students onto a latent continuum based on their responses to individual items, this cannot lend direct insight into discrete categories of responses that exist in the data. To partition students' responses into common patterns regarding their acceptance of evolution, we utilized clustering, an unsupervised geometric classification technique (Theodoridis & Koutroumbas, 2001). By "unsupervised," we mean that there is no reference dependent variable by which to solve model parameters. This can be contrasted with supervised techniques (e.g., regression, including Rasch modeling) where independent variable parameters are chosen to achieve fit with a dependent variable. In this way, unsupervised classification is a purely exploratory approach with an underlying motivation similar to grounded theory; our goal in using clustering was to quantify and visualize the MATE's 20-item universe of responses without prior assumptions about what this universe should like. While the Rasch model is a variable-centered approach which imposes the criterion that students with higher evolution acceptance will express higher agreement to individual items, clustering, a person-centered approach, holds no such assumptions (Collins & Lanza, 2010). Instead of placing students along a scale, clustering assigns students to categories, and can therefore be used to detect more nuanced patterns in how students express their acceptance (or non-acceptance) of evolution on the MATE.

We utilized the *k*-means clustering procedure (Theodoridis & Koutroumbas, 2001) to classify students into discrete categories based on their responses to the 20 items on the MATE. See Supplementary Materials for additional details on the *k*-means algorithm.

Finding the Best Number of Clusters. One problem with clustering procedures including *k*-means is that the optimal number of clusters is not known a priori. The optimal number of clusters to describe the data was decided using two methodologies: (i) the sum of squares error (SSE) and (ii) the Dunn index (Dunn, 1973). While the SSE will tend to decrease with an increased number of clusters, the best number of clusters is found when the SSE stops decreasing a large amount when a new cluster is added. Since what constitutes a large amount is somewhat subjective, we used the Dunn index as an additional validity measure. The Dunn index is the ratio of the smallest distance between observations not in the same cluster (cluster separation) and the largest distance between observations in the same cluster (lack of cluster compactness).

The best solution yields compact clusters that are well separated from neighboring clusters; hence, the best number of clusters is obtained by maximizing the Dunn index. Once a solution for the best number of clusters is found, it can be interpreted as follows: students within the same cluster express their acceptance of evolution similarly, while students in different clusters possess important differences in how they express their acceptance of evolution.

Cross-Validation of the Cluster Solution. Using an inductive, unsupervised approach to pattern recognition in data allows generation of a model from data through a process similar to the inductive coding of qualitative data inherent in grounded theory traditions (Glaser, 1978, 1992; Glaser & Strauss, 1967, 2009; Strauss, 1987). One important part of this process is to see how the model performs on new data, thereby evaluating the generalizability, robustness, and usefulness of the model (Gasson, 2004). In quantitative research, this is accomplished through cross-validation (Diamantidis, Karlis, & Giakoumakis, 2000). While there are many methods to perform cross-validation, we chose to evaluate the efficacy of our cluster model using an external student body in an ethnically distinct university within a different part of the United States. This was done in three steps. In the first step, we assigned students in the new dataset to clusters using the cluster centers derived from the previous analysis. In the second step, we assigned students to clusters based on cluster centers best-fitting the new dataset. In the third step, we evaluated the agreement between students' cluster assignments using the two models, which we quantified through two statistics: percent agreement and Fleiss' kappa (Fleiss, 1971; McHugh, 2012). Landis and Koch (1977) suggest that a kappa value above 0.6 is indicative of substantial agreement between the two models.

RQ3: What Are Key Similarities and Differences Between the General Patterns of Evolution Acceptance Expressed by Students?

Evaluation of Cluster Uniqueness. Separation or uniqueness of clusters with respect to the MATE was evaluated using a 1-way multivariate analysis of variance (MANOVA) procedure. We used MANOVA to test the null hypothesis that the average response pattern for the 20 items on the MATE was not significantly different between students in different clusters at the 95% confidence level using the F-statistic calculated from Wilk's Lambda. If the multivariate null hypothesis was rejected, then univariate ANOVA's were used to test the null hypothesis that the average response on each item did not differ significantly between students in different clusters. If this null hypothesis was rejected for a single item at the 95% confidence level, then the Scheffe test for differences in means was used to quantify significance of differences in the average response score on the item between individual clusters. Given that the MATE solicits students' affective responses (e.g., Deniz et al., 2008) to a variety of ontological and epistemological factors related to evolution, it was also useful to quantify the effect size of differences between clusters on each item. We used η^2_{partial} to quantify the proportion of variance in students' responses to each item that can be attributed to differences between clusters.

Descriptive Analyses of Each Cluster. The presence of statistically separable clusters is reflective of general patterns by which the college students in this study express their acceptance of evolution in light of the ideas assessed by the MATE. A descriptive analysis of each cluster was generated using the mean level of acceptance expressed on each item by students in each cluster. By observing the average acceptance scores across items expressed by students in each cluster and differences in scores between clusters, we were able to create a qualitative acceptance profile for each group of students and describe important differences between them.

RQ4: What Is the Role Content Knowledge in Students' Acceptance of Evolution?

Following previous studies, we began by computing the Pearson correlation of students' positions along re-MUM knowledge and MATE acceptance Rasch scales. While this statistic is informative, it only gives a first approximation of the role of content knowledge in acceptance. Using clustering results, we were able to extract categorical response patterns for acceptance of evolution. Evaluation of differences between the average knowledge levels of students within each cluster category for acceptance enables a more in depth, descriptive analysis of the relationship between knowledge and acceptance of evolution. Using a one-way ANOVA procedure, we evaluated this relationship under the null hypothesis of no significant difference in re-MUM Rasch measures between acceptance profiles at the 0.05 alpha confidence level. If the null hypothesis was rejected, we proceeded to evaluate differences in knowledge between individual groups using the Scheffe post-hoc procedure. We then calculated ω^2 as a population measure of the percent variance in knowledge of evolution that can be attributed to differences between acceptance profiles.

Results

RQ1: How Do the MATE Items and the MATE Scale Conform to Rasch Validity Expectations?

Fitting the Single-Scale Model. Rasch analysis indicates that a single dimension is not enough to explain the data completely. PCA on Rasch residuals generated a first eigenvalue of 2.73 items of variance. Since this value is well above 2 (Raîche, 2005), we concluded that there are additional important underlying dimensions not explained by a one-dimensional measurement model. Local dependency as measured by item residual correlations ranged from −0.32 to (A2 and A17) to 0.33 (A13 and A16; A1 and A3). Although these values are above 0, they account for only about 16% of the shared variance. Linacre (2005) states that item residual correlations below 0.7 (i.e., 49% shared variance) are relatively low and document non-problematic local dependency.

Treated as a single 20-item scale, the MATE exhibits high person and item reliability of 0.92, and all items also exhibit satisfactory fit with the Rasch model, with the exception of item A5 which displays misfit (infit = 1.52, outfit = 1.47). This analysis shows that the MATE works in providing a precise one-dimensional measure for acceptance of evolution. However, a measurement model that acknowledges the MATE's multidimensional nature may allow extraction of more meaningful information.

Exploring a Two-Scale Model. Since a single scale did not adequately account for the data from the Rasch perspective, we explored breaking the MATE into two scales. Given that the MATE has yet to be treated as multiple subscales, we first performed principle components exploratory factor analysis with promax rotation which suggested that items A1, A3, A8, A11, A12, A15, A16, A18, and A20 should load onto the first scale and items A2, A4, A5, A6, A7, A9, A10, A14, A17, and A19 should load onto the second scale. What constructs do these scales measure?

Scale 1—Facts

Ten items define the first scale, including A1, A3, A8, A11–A13, A15, A16, A18, and A20 (Tables 2 and 3). We refer to this first construct as the *Facts* dimension, as its items elicit ontological views about scientific facts and supporting data for evolution. Items in this scale measure agreement with correct scientific facts, including statements about evolutionary processes over millions of years (A1), human evolution (A3), and the age of the Earth (A11). Other items in this scale describe quality evolution research, including statements about supporting and significant data (A8, A16), methodological rigor (A12), scientific validity (A20), and generating testable predictions (A13).

Scale 2—Credibility

The remaining 10 items (A2, A4–A7, A9, A10, A14, A17, and A19) define a second scale addressing the *credibility* of evolution and rejection of non-scientific ideas. We refer to this second construct hereon as the *Credibility* dimension. This scale is predominantly affective in nature, and underscores the distrust individuals place in evolutionary science and scientists. Items in the *Credibility* scale include several negative terms and phrases, including "incapable" (A2), "not valid" (A4, A10), "ambiguous" (A6), "cannot be correct" (A14), and "doubt" (A17). This scale also includes agreement with non-scientific ideas about the age of the Earth (A7), essential, non-changing forms of species (A9), and sudden appearance of diverse life forms (A19).

Table 2
Rasch item statistics for the two-dimensional MATE

Item	Measure	SE	Infit	Outfit	PtBis	Description
Dimension 1-*Facts*—Acceptance of Evolution Facts and Supporting Data ($\rho_{person} = 0.87$, $\rho_{item} = 0.89$)						
A1	−0.46	0.10	0.64	0.57	0.74	Organisms existing today are the result of evolutionary processes that have occurred over millions of years.
A3	0.14	0.08	0.69	0.65	0.71	Modern humans are the product of evolutionary processes which have occurred over millions of years.
A8	0.01	0.10	1.19	1.22	0.43	There is a significant body of data which supports evolutionary theory.
A11[a]	0.04	0.10	1.28	1.52	0.42	The age of the Earth is at least 4 billion years.
A12	0.07	0.11	1.16	1.26	0.42	Current evolutionary theory is the result of sound scientific research and methodology.
A13	−0.24	0.11	0.89	0.91	0.57	Evolutionary theory generates testable predictions with respect to the characteristics of life.
A15[a, b]	0.51	0.09	1.31	1.48	0.40	Humans exist today in essentially the same form in which they always have.
A16	0.25	0.09	0.84	0.85	0.62	Evolutionary theory is supported by factual, historical, and laboratory data.
A18	−0.43	0.10	0.90	0.90	0.58	The theory of evolution brings meaning to the diverse characteristics and behaviors observed in living forms.
A20	0.12	0.09	1.04	1.13	0.53	Evolution is a scientifically valid theory.
Dimension 2-*Credibility*—Acceptance of Evolution Credibility and Rejection of Non-Scientific Ideas ($\rho_{person} = 0.88$, $\rho_{item} = 0.94$)						
A2[b]	0.19	0.09	0.87	0.85	0.61	The theory of evolution is incapable of being scientifically tested.
A4[b]	0.22	0.08	0.89	0.91	0.61	The theory of evolution is based on speculation and not valid scientific observation and testing.
A5[a]	−0.64	0.13	1.44	1.40	0.20	Most scientists accept evolutionary theory to be a scientifically valid theory.
A6[b]	0.57	0.10	1.19	1.21	0.42	The available data are ambiguous as to whether evolution actually occurs.
A7[b]	−0.49	0.08	1.03	1.11	0.53	The age of the Earth is less than 20,000 years.
A9[b]	−0.26	0.08	0.87	0.95	0.61	Organisms exist today in essentially the same form in which they always have.
A10[b]	0.04	0.08	0.83	0.85	0.64	Evolution is not a scientifically valid theory.
A14[b]	0.16	0.08	0.90	0.86	0.62	The theory of evolution cannot be correct since it disagrees with the Biblical account of creation.
A17[b]	0.01	0.09	0.98	0.94	0.56	Much of the scientific community doubts if evolution occurs.
A19[b]	0.21	0.08	0.98	1.01	0.57	With few exceptions, organisms on Earth came into existence at about the same time.

[a]Infit or outfit greater than 1.30.
[b]Negatively worded item.

Rasch Analysis of the Two-Scale Model

PCA on Rasch residuals lends support for the integrity of treating the MATE as two scales. While first eigenvalues for Rasch residuals on the one-dimensional MATE sat above 2.7, a two-scale MATE yields eigenvalues of 1.70 and 1.61 items of variance for the *Facts* and *Credibility* scales, respectively, indicating that there is no systematic

Table 3
Comparison of classification between the model derived from analysis of the Midwestern United States students (previous model) and the model best-fitting the Western United States students (new model)

	New Model					
Previous Model	1	2	3	4	5	% Agreement
1	**155**	0	0	8	0	95.1
2	0	**59**	1	0	6	89.4
3	0	0	**59**	0	0	100
4	7	54	0	**389**	32	80.7
5	0	69	47	4	**155**	56.4
Total						78.2[a]

Substantial agreement between the two solutions supports generalizability of the cluster model. Categories predicted most frequently are in bold.
[a]$\kappa = 0.703$.

variance left in the residuals unaccounted for by the two-scale model. The Rasch models suggest that while some local dependency between items exists, it does not significantly attenuate the usefulness of the MATE. The largest item residual correlation magnitude is 0.26 (A8 and A11) in the *Facts* scale and 0.22 (A2 and A7) in the *Credibility* scale. This correlation of 0.26 suggests 7% shared variance between item residuals which does not significantly attenuate the usefulness of items (Linacre, 2005).

This result lends an argument that two scales most accurately reflect the true structure of the MATE, and the usefulness of this two-scale structure will be illustrated definitively in the cluster analyses to follow. We find that Rasch measures for the *Facts* and *Credibility* dimensions have a correlation of 0.77, meaning that approximately one half of the variance is independent between measures for *Facts* and *Credibility*.

Rasch modeling also shows that the two-scale MATE exhibits satisfactory construct validity, meaning that it is a useful tool for distinguishing between students with high and low acceptance (Table 2). Measures for both the *Facts* ($\rho_{\text{person}} = 0.87$, $\rho_{\text{item}} = 0.89$) and *Credibility* ($\rho_{\text{person}} = 0.88$, $\rho_{\text{item}} = 0.94$) dimensions exhibit reliability sufficient for meaningful comparison of individual students for purposes of hypothesis testing and items along the Rasch scale for analyses of construct validity. Analysis of dispersion of measures for students and items along the Rasch continuum for the two scales (Figure 1) lends a more qualitative look at construct validity. While we see considerable overlap for person and item distributions near the center of the Rasch continuum (logit 0) along both scales, the fact that the item distribution is shifted downward indicates that the MATE provides the most useful and precise measures for students with low levels of evolution acceptance, and its usefulness diminishes in students with high levels of acceptance.

The person-item maps (Figure 1) also allow us to see concepts that are easy or difficult for students. Items A1 (Organisms existing today are the result of evolutionary processes that occurred over millions of years) and A18 (The theory of evolution brings meaning to the diverse characteristics and behaviors observed in living forms) on the *Facts* scale and A5 (Most scientists accept evolutionary theory to be a scientifically valid theory) on the *Credibility* scale are among the easiest items for students to agree with. On the contrary, items A15 (Humans exist today in essentially the same form as they always have) in the *Facts* scale and A6 (The available data are ambiguous as to whether evolution actually occurs) on the *Credibility* scale met the most persistent disagreement (we note that A6 is negatively worded, and was, therefore, reverse coded).

As with the single-scale construction of the MATE, a majority of the items in the two-scale construction fit well with the Rasch model. While none of the items display extreme misfit, A5, A11, and A15 display fits that push the high bound of what may be considered a useful item, with infits of 1.44, 1.28, and 1.31, and outfits of 1.40, 1.52, and 1.48, respectively.

The positive item-total correlations of these items (0.20, 0.42, and 0.40, respectively) indicate the general tendency for students with higher levels of evolution acceptance to express agreement. However, a closer look shows there are a number of instances where low acceptance students express agreement and high acceptance students express disagreement, meaning that these items are not as useful as the other items on the MATE.

RQ2: How Many Distinct Patterns Exist for How College Students Express Their Acceptance of Evolution?

Comparison of sum of squares error (SSE) and Dunn indices between 2- and 10-cluster solutions shows a 5-cluster solution as the optimal explanation of the data (Figure 2). Students' acceptance of evolution can be subsequently classified into five distinct clusters. The five-cluster solution maximizes the Dunn index at 0.20, indicating the most similarity among students in a given group, and maximization of the separation among students in different groups. The SSE also supports the five-cluster solution.

We also find substantial agreement between this cluster solution and the five-cluster solution derived from best-fit with the separate dataset from the Western United States (Table 3). In classifying students within one of the five clusters, these models have a percent agreement of 78.2% and a kappa value of 0.703, indicating that the model derived from the Midwestern students has substantial robustness and generalizability. We now proceed to discuss the evolution acceptance profile associated with each cluster.

RQ3: What Are Key Similarities and Differences of Evolution Acceptance Among the Cluster Profiles?

Group-Level Response Variance. At the most general multivariate level, we found response patterns on the MATE are significantly different among groups at the 0.05 alpha level ($\Lambda = 0.015$, $F_{(80,661.2)} = 14.2$, $p << 0.001$), and that these group-level differences account for 63% of the

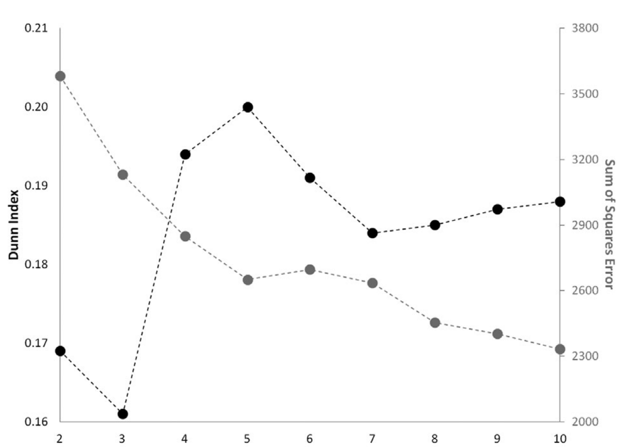

Figure 2. Plots of the Dunn index (black) and sum of squares error (gray) for multiple cluster solutions.

variation in the data ($\eta^2_{\text{partial}} = 0.63$). At the single-item level, we found significantly different between-group response levels on all items ($6.03 \le F_{(4,186)} \le 99.75$, $p << 0.001$), respectively accounting for between 11% and 68% of variation in item responses ($0.11 \le \eta^2_{\text{partial}} \le 0.68$; Table 4, Figures 3 and 4). Group differences accounted for over one half of the response variance on items A1, A3, A10, and A14. A1 and A3, within the *Facts* dimension, measure the extent to which students accept large time scales. Items A10 and A14, within the *Credibility* dimension, solicit acceptance based on both scientific (A10) and religious grounds (A14). Levels of acceptance were most similar on A5, indicating general agreement across all groups that scientists accept the scientific validity of evolution.

Description of the Cluster Profiles. Although separation among the cluster groups is significant under both statistical and practical considerations, we can begin to describe the acceptance qualities of students in each of the five cluster profiles. We note MATE item means for each cluster profile in Figure 3 and highlight the proportion of the study population from each cluster in Figure 5. We subsequently describe cluster profiles by highlighting patterns in mean MATE scores for each dimension (0 = strongly disagree; 4 = strongly agree) as well as unique item response patterns within each cluster profile (Tables 4 and 5).

As we discuss item means within acceptance profiles, the Likert-scale values for particular items are reversed (0 = strongly agree; 4 = strongly disagree). Reversed scales affect 9 of 10 items

Table 4
Mean acceptance scores on items on the MATE for students in each cluster

	Cluster 1 ($n = 44$)		Cluster 2 ($n = 23$)		Cluster 3 ($n = 12$)		Cluster 4 ($n = 78$)		Cluster 5 ($n = 34$)		
Item	*M*	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD	η^2_{Partial}
A1	3.91	0.29	3.74	0.45	0.50	0.90	3.33	0.62	2.41	0.74	0.68
A3	3.70	0.73	3.22	0.85	0.17	0.58	2.91	0.84	1.68	1.04	0.56
A8	3.59	0.66	2.52	1.12	1.58	1.16	2.88	0.76	1.71	0.76	0.41
A11	3.61	0.78	2.96	1.19	0.75	1.06	2.62	0.97	2.65	0.81	0.34
A12	3.23	0.83	3.30	0.70	1.08	1.16	2.62	0.71	2.24	0.78	0.34
A13	3.39	0.69	3.09	0.95	1.00	0.60	2.71	0.63	2.32	0.77	0.40
A15[a]	3.36	1.06	1.83	1.27	0.25	0.45	2.58	1.03	1.65	1.07	0.38
A16	3.36	0.81	2.61	1.03	0.83	0.58	2.62	0.84	2.03	0.83	0.36
A18	3.68	0.52	2.87	1.10	1.33	0.89	2.94	0.61	2.41	0.78	0.40
A20	3.50	0.93	2.83	1.11	0.92	1.00	2.78	0.82	1.97	0.83	0.36
M	3.53		2.90		0.84		2.80		2.11		
SD	0.20		0.52		0.45		0.23		0.35		
A2[a]	3.57	0.50	1.52	0.90	1.25	1.36	2.27	0.77	1.91	1.03	0.44
A4[a]	3.52	0.73	1.30	1.11	1.33	1.23	2.58	0.76	1.62	0.82	0.47
A5	3.41	0.73	2.65	1.07	3.17	0.72	2.82	0.73	2.74	0.75	0.11
A6[a]	2.93	0.87	1.39	0.66	1.58	0.90	2.15	0.90	1.56	0.86	0.29
A7[a]	3.86	0.41	1.43	1.24	1.25	1.60	3.09	1.01	3.15	0.89	0.42
A9[a]	3.70	0.73	1.26	1.29	1.17	0.94	3.03	0.76	2.56	0.86	0.48
A10[a]	3.64	0.65	1.48	1.34	0.75	0.75	2.81	0.65	1.97	0.87	0.52
A14[a]	3.66	0.57	1.17	0.89	0.50	0.90	2.47	0.94	2.03	1.17	0.50
A17[a]	3.52	0.66	0.87	0.69	2.25	0.97	2.60	0.78	2.12	1.12	0.47
A19[a]	3.50	0.63	1.22	1.13	0.92	1.00	2.41	0.83	1.97	1.09	0.44
M	3.53		1.43		1.42		2.62		2.16		
SD	0.25		0.47		0.78		0.31		0.50		

[a]Likert-scale reversed due to negative item loading to the dimension.

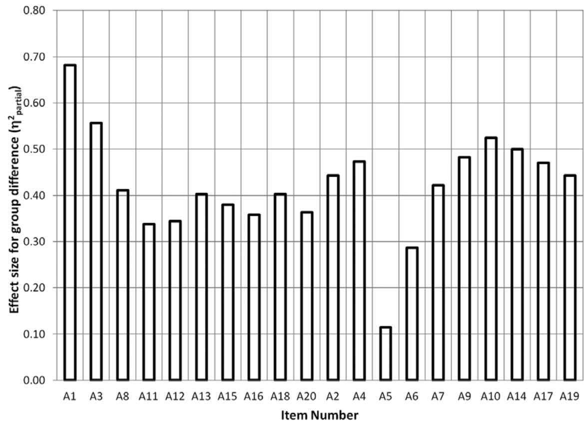

Figure 3. Cluster separation effect sizes for items on the MATE.

on the *Credibility* dimension and 1 item on the *Facts* dimension (see Table 2). All item scores and statistics should be interpreted as degree of acceptance (0 = no acceptance, 4 = maximum acceptance) expressed, not students' raw responses.

Cluster 1: Uniform, High Acceptance

Cluster 1 (23.0% of the population; $n = 44$) is exemplified by uniform and consistently high acceptance of evolution. Mean acceptance scores for this cluster indicated strong agreement for both *Facts* (3.53 ± 0.20) and *Credibility* (3.53 ± 0.25), placing the group between "agree" and "strongly agree" for both dimensions. There was also little variation in responses on the MATE among the students in this group. These students expressed the highest average acceptance on A1 and A7, indicating agreement with an evolutionary time scale over millions of years (A1) and agreement that the Earth is at least 4 billion years old (A7). Cluster 1 students had the greatest difficulty agreeing with item A6, the only item with an average rating below 3 (the "agree" level) for the group; this indicated some degree of doubt regarding whether there are valid scientific data to support evolution.

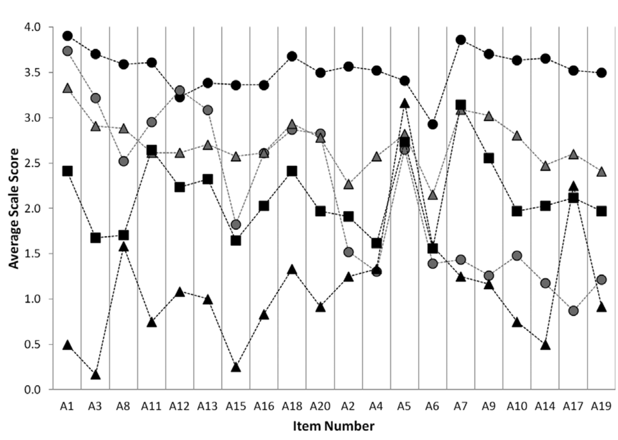

Figure 4. MATE scale score profiles for the five response clusters.

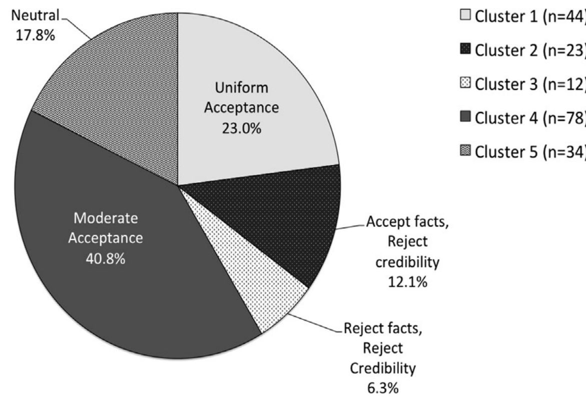

Figure 5. Features of each MATE acceptance profile by sample size and proportion of the overall study population.

Cluster 2: Accepting Facts, Rejecting Credibility

Cluster 2 (12.1% of the population; $n = 23$) is defined by its asymmetrical profile (Figure 4), exemplified by a significantly higher average agreement with *Facts* than *Credibility* ($t_{18} = 6.65$, $p << 0.001$). The average agreement with *Facts* about evolution was 2.90 ± 0.52 for this group, approximately at the "agree" level. On the other hand, average acceptance of *Credibility* was 1.43 ± 0.47, between "disagree" and "neutral." Among *Facts* items, Cluster 2 students expressed highest agreement with items A1 and A12. This indicates high acceptance of evolutionary change over millions of years (A1) and an appreciation of rigorous scientific methodology (A12).

A15 was the only item in the *Facts* dimension with an average rating below 2 ("neutral") for Cluster 2 students. Agreement with this item indicates that Cluster 2 students think that humans today exist in essentially the same form in which they always have. The contrast of item A15 with item A3 (humans are a product of evolutionary change), hints that essentialist biases could be sources of conflict for this group (e.g., species have essential identities that cannot change, Evans, 2008; Mayr, 1982).

Table 5
Key features of each cluster profile, including items with highest and lowest average agreement

Cluster Profile	Profile Features	Dimension 1: Facts	Dimension 2: Credibility	Items With Higher Avg. Acceptance	Items With Lower Avg. Acceptance
Cluster 1	Uniform acceptance	3.53 ± 0.20	3.53 ± 0.25	A1, A7	A6
Cluster 2	Accept facts, reject credibility	2.90 ± 0.52	1.43 ± 0.47	A1, A12	A9, A14, A17, A19
Cluster 3	Reject facts, reject credibility	0.84 ± 0.45	1.42 ± 0.78	A5, A8, A18	A3, A15
Cluster 4	Moderate acceptance	2.80 ± 0.23	2.62 ± 0.31	A1, A7, A9	A2, A6
Cluster 5	Dispersed neutrality	2.11 ± 0.35	2.16 ± 0.50	A5, A7, A9, A11	A3, A4, A6, A8, A15

Cluster 2 had persistent disagreement with items in the *Credibility* dimension. The group expressed the greatest disagreement with items A9, A14, A17, and A19. Low scores on these items indicate sympathy with the Biblical account of creation (A14), including support of the idea that all life appeared at the same time (A19) and organisms are in the same form they have always been (A9). Furthermore, Cluster 2 was the only group to agree with "much of the scientific community doubts if evolution occurs" (A17).

Cluster 3. Rejecting Facts and Credibility

Like Cluster 2, Cluster 3 is defined by an asymmetrical profile (Figure 4). However, the asymmetry goes in the opposite direction and is less pronounced. We also see slightly higher average agreement with *Credibility* and less agreement with *Facts* ($t_{18} = 2.03, p = 0.058$). Cluster 3 comprised the smallest proportion of students in our population (6.3% of the population; $n = 12$).

There is no statistical or practical difference in average acceptance of *Credibility* between Cluster 2 and 3 students. Scheffe difference tests indicate that Cluster 3 acceptance of individual items on the *Credibility* dimension are statistically equal to those in Cluster 2, with the exception of item A17, which was significantly higher in Cluster 3 ($d_{\text{scheffe}} = 1.40$, $\text{SE}_{\text{scheffe}} = 0.30$, $p << 0.001$). This indicates that while Cluster 3 students have similar concerns about the credibility of evolutionary science, they are more likely than individuals in Cluster 2 to acknowledge evolution acceptance within the scientific community.

The greatest defining feature of the Cluster 3 student is vehement rejection of the *Facts* dimension. The average *Facts* score for Cluster 3 students was 0.84 ± 0.45 (below the "disagree" level), a mean that is significantly lower than Cluster 2 students ($t_{18} = 9.50$, $p << 0.001$). Furthermore, Cluster 3 students have significantly lower acceptance levels than Cluster 2 students on all of the items in the *Facts* dimension. The only near exception is Item A8, which was similar to Cluster 2, but still significantly different by just under one level of agreement on the Likert scale ($d_{\text{scheffe}} = 0.94, \text{SE}_{\text{scheffe}} = 0.29, p = 0.038$).

Cluster 3 students had the lowest average scores on items A3 and A15, at 0.17 ± 0.58 and 0.25 ± 0.45, respectively. This suggests a particularly strong sentiment that humans are excluded from the process of evolution. The students in this group had the highest average *Facts* scores on items A8 and A18, at 1.58 ± 1.16 and 1.33 ± 0.89, respectively. While these scores are still below the "neutral" level, they indicate that some Cluster 3 students acknowledge the explanatory value of evolutionary theory (A18) and that data are used to support evolutionary theory (A8).

Cluster 4: Uniform Moderate Acceptance

As with Cluster 1, the profile of students in Cluster 4 (40.8% of the population; $n = 78$) is marked by uniformity across both dimensions (Figure 4). However, in comparison to Cluster 1, the average Cluster 4 levels of acceptance for both dimensions are moderate and significantly lower (*Facts:* $t_{18} = 7.54$, $p << 0.001$; *Credibility*: $t_{18} = 7.25$, $p << 0.001$). Cluster 4 students have scores above 3 (the "agree" level) on items A1, A7, and A9, indicating agreement that evolution occurs over millions of years (A1), disagreement that the Earth is less than 20,000 years old (A7), and disagreement that organisms (species) have always existed in their current forms (A9). Items with lower acceptance for Cluster 4 students include A2 and A6, with relatively neutral scores of 2.27 ± 0.77 and 2.15 ± 0.90, respectively. These scores indicate neutral views of evolution as scientifically testable (A2) and neutral views on the ambiguity of evolutionary data (A6).

Cluster 5: Dispersed Neutrality

Cluster 5 (17.8% of the population, $n = 34$) is demarcated by dispersed, but largely neutral views of both the *Facts* and *Credibility* dimensions. Means for both dimensions in Cluster 5 hover just above the "neutral" level of 2.0, with average acceptance of *Facts* at 2.11 ± 0.35 and acceptance of *Credibility* statistically equal at 2.16 ± 0.50. While relatively symmetric, this profile is more variable than the other symmetric profiles (Clusters 1 and 4). Within the *Facts* dimension, only item A11 had a score over 2.5 (closer to "agree" than "neutral"). A relatively high score above 3.0 on A7 (*Credibility* dimension) compliments A11, documenting that Cluster 5 students disagree with the idea that the Earth is less than 20,000 years old (A7) and agree that Earth is over 4 billion years old (A11). Items A5 and A9 in the *Credibility* dimension had average scores over 2.5 for Cluster 5 students, indicating agreement with the validity of evolution within the scientific community (A5) and agreement that organisms (species) do not change over time (A9).

Cluster 5 students had the most difficulty agreeing with items A3, A8, and A15 in the *Facts* dimension and items A4 and A6 in *Credibility* dimension. The low scores on A3, A8, and A15 suggest that Cluster 5 students do not agree that humans have evolved (A3, A15), and do not agree that there is a significant body of data supporting the theory of evolution (A8). This latter aspect is corroborated by agreement that evolution is based on speculation (A4) and that evolutionary data are ambiguous (A6).

RQ4: What Is the Role of Content Knowledge in Students' Acceptance of Evolution?

We find a Pearson correlation of 0.30 for acceptance of *Facts* with knowledge of macroevolution (Re-MUM, Walter & Romine, in review), and 0.43 between acceptance of *Credibility* and knowledge of macroevolution (Re-MUM, Authors, in review). When corrected for unreliability of the Rasch scales, these correlations become 0.39 and 0.56, respectively. This means that about 16% of the variance in the *Facts* scale and about 31% of the *Credibility* scale can be explained by knowledge of macroevolution. Although correlation does not equal causation, this may also imply that knowledge of macroevolution has twice the effect on students' acceptance of *Credibility* than *Facts*.

Although correlation analysis provided a quantitative first approximation on the extent to which knowledge and acceptance of evolution relate, analysis of differences in re-MUM measures between students in each cluster permitted a richer understanding of this relationship. Significant differences in average measures for knowledge of macroevolution were found between the five groups ($F_{(4,186)} = 11.81, p << 0.001, \omega^2 = 0.18$).

Given the moderate correlations described above, we were not surprised that the high evolution accepters (Cluster 1) had the highest average logit measure for knowledge of macroevolution ($M = 1.51$, $\mathrm{SD} = 0.95$). However, we found it more surprising that the students who generally accepted the facts, but not its credibility (Cluster 2) had the lowest average logit measure for macroevolution of all of the groups ($M = 0.051$, $\mathrm{SD} = 0.68$). Scheffe tests show that Cluster 2 students had average knowledge levels significantly lower than both the high acceptance ($d_{\mathrm{scheffe}} = 1.45$, $\mathrm{SE}_{\mathrm{scheffe}} = 0.22$, $p << 0.001$) and the uniform moderate acceptance (Cluster 4) students ($d_{\mathrm{scheffe}} = 0.76, \mathrm{SE}_{\mathrm{scheffe}} = 0.20, p = 0.009$).

The groups showing the lowest average acceptance (Cluster 3), uniform moderate acceptance (Cluster 4), and neutral acceptance (Cluster 5) had moderate, and relatively equal knowledge levels of macroevolution, with mean logit measures of 0.80 ($\mathrm{SD} = 1.36$), 0.81 ($\mathrm{SD} = 0.76$), and 0.68 ($\mathrm{SD} = 0.83$), respectively. These distinctions show that the relationship between knowledge and acceptance of evolution is not a simple linear relationship. Instead, this relationship is nuanced and reflected by a low-to-moderate effect size. The population estimate ω^2 for the ANOVA test was measured at 0.18, suggesting that the group distinctions for acceptance found in this study explain just under one-fifth of the variation in post-secondary students' knowledge of macroevolution.

Discussion

The MATE has been used frequently in evolution education research, but has not been validated on post-secondary students with Rasch analysis. The present lack of validity evidence for the MATE is troubling. As Wagler and Wagler (2013) suggest, it is unreasonable to assume that the MATE's behavior is invariant across different populations; indeed, they use CFA and Rasch methods to show the inadequacy of 1-, 2-, and 6-factor measurement structures for a Hispanic population of pre-service teachers. This said, evidence for lack of validity in certain populations should not be taken imply that the MATE is an invalid measure for all populations. We wish to note here that the substantial cluster classification agreement between the predominantly White sample which served as a focus in this study and the predominantly Hispanic sample used for model validation shows that the five-cluster solution we derived from the MATE is indeed generalizable and useful.

Efficacy and validity of any measurement tool, including the MATE, should be evaluated on a case-by-case basis. We discuss our results with the goal of providing new and continuing methodological support for the MATE as a measurement tool, write an operational definition of evolution acceptance for the MATE, and provide suggestions for how our cluster profiles inform and align with evolution education research.

The MATE as a Measurement Tool

The MATE is the most popular means to measure evolution acceptance in the published literature (Table 1). Rutledge and Warden (1999) and Rutledge and Sadler (2007) have already confirmed that the MATE exhibits high reliability and construct validity from the perspective of Classical Test Theory. They found that the MATE was largely unidimensional, citing that 71% of the variance in the data came from a single factor (Rutledge & Warden, 1999). Considering the high reliability, one would expect that a one-dimensional measurement model would account for students' responses on the MATE adequately. However, high reliability does not imply that the MATE is unidimensional (Schmitt, 1996). While the MATE has never been used two-dimensionally, we found that two separate unidimensional models better met the assumptions of both the Rasch and classical measurement frameworks, and that the new model uncovered facets of acceptance that were not previously gleaned from past work.

Exploring the "Facts" and "Credibility" Dimensions

Looking at items in each scale collectively, we found 10 items on the MATE measure acceptance of *facts and supporting data for evolution* (Scale 1, *Facts*), and the other 10 items measure *acceptance of the credibility of evolutionary science and rejection of non-scientific ideas* (Scale 2, *Credibility*). This supports two assertions. First, the two dimensions suggest that the MATE has not been used to its full capacity for providing detailed measures of students' acceptance of evolution. Second, the dimensions do not appear to be opposites of one another.

Readers who look at the MATE from an "expert perspective" may have difficulty accepting how "paired items" could be split between two factors. On the surface, negative wording would seem to imply that the two factors are measuring "acceptance" and "doubt," opposite manifestations of a single underlying construct. However, we suggest that negative wording on the MATE is more complex than this, and that in fact negatively worded items might elicit a different thinking pathway, largely independent from the positively worded dimension. While this may be difficult to accept, our theoretical frameworks support this assertion, documenting evolution acceptance as influenced by the affective domain (Deniz et al., 2008) and unconscious feelings of certainty (Ha et al., 2012). We also see these two dimensions as supporting a cognitive, knowledge-driven domain (*Facts and Data)* and an affective, emotion-driven domain of acceptance (*Credibility*), as documented in Deniz et al. (2008). The *Credibility* dimension may also be influenced by unconscious feelings of certainty as documented by Ha et al. (2012).

We also document support for the *Facts* and *Credibility* dimensions of acceptance in studies of evolution acceptance using the EALS (Hawley, Short, McCune, Osman, & Little, 2011) and I-SEA (Nadelson & Southerland, 2012). The items on our *Facts* dimension are similar to those on the *Knowledge/Relevance* higher order factor on the EALS (Hawley et al., 2011), a dimension that included items on participants' perceptions of evolutionary knowledge, knowledge of the scientific enterprise, genetic literacy, and relevance of evolutionary theory. Similarly, our *Credibility* dimension has items similar to those from the *Creationist Reasoning* higher order factor on the EALS, a dimension which included items on participants' distrust of the scientific enterprise, social objections, moral objections, young Earth creationist beliefs, and intelligent design fallacies. We also find support for the *Credibility* dimension in the work of Nadelson and Hardy (2015), who documented a significant correlation ($r = 0.43$; $p < 0.05$) between post-secondary students' trust in science and scientists (Nadelson et al., 2014) and evolution acceptance (I-SEA; Nadelson & Southerland, 2012).

How Are the "Facts" and "Credibility" Dimensions Different?

When examined at face value, items in the *Facts* and *Credibility* dimensions appear to be eliciting the same things. For example, A20, *evolution is a scientifically valid theory*, loads onto the *Facts* dimension, while its intended companion, A10, *evolution is not a scientifically valid theory*, loads onto the *Credibility* dimension. Although these are negative versions of the same statement, they appear to elicit different emotional reactions (an affective response as anticipated by the Deniz et al., 2008 model). Subsequently, although A10 and A20 are true reversals of one another semantically, if considered in terms of our conceptual frameworks, separating them is logical. We would expect A20 to load onto a cognitive dimension (since it elicits cognitive affirmation) and A10 to load onto an affective dimension (since it elicits affective denial).

Other item pairs like A7–A11 and A4–A14 could be also be considered antitheses of each other; however, unlike A10 and A20, the item pairs are not true reversals. In other words, the items do not differ by the inclusion of a "no" or "not," but rather by the inclusion of different adjectives, nouns, and phrases. Wording the items in this way causes some items to be scientifically correct and others to be non-scientific or conveying skepticism. For example, items A7 and A11 both refer to the age of the Earth. They differ by posing different geologic ages, one scientifically accurate (>4 billion years) and one scientifically inaccurate (<20,000 years). Hypothetically, if an individual thought the Earth was 5 million years old, they would disagree with both items.

Our data, therefore, suggest that MATE items should be considered autonomously rather than as item pairs. To put this idea in context of this study, negative phrasing and non-scientific statements are responsible for much of the MATE's two-dimensionality. Negatively worded and scientifically incorrect ideas elicit views rooted in distrust and denial (Scale 2. *Credibility*). Positively worded and scientifically correct ideas elicit confidence and comfort with facts and supporting data (Scale 1. *Facts*).

Negative and non-scientific ideas are documented in the *Credibility* dimension in two fashions. Foremost, items in the *Credibility* dimension include several negative terms and phrases, including "incapable" (A2), "not valid" (A4, A10), "ambiguous" (A6), "cannot be correct" (A14), and "doubt" (A17). Secondly, the *Credibility* dimension elicits agreement with non-scientific ideas, including the idea that Earth less than 20,000 years old (A7), organisms exist today in the essentially the same form they always have (A9), and organisms on Earth came into existence at about the same time (A19). In contrast, no items on the *Facts* dimension have negative terms and only one item is non-scientific (A15).

Operationally Defining Acceptance

One of the primary concerns surrounding the original MATE publications is the lack of an operational definition for "evolution acceptance" as supported by data (Hogan, 2000; Smith, 2010; Wagler & Wagler, 2013). The two acceptance dimensions identified in this study help us to work toward a data-driven operationalization of acceptance as measured by the MATE. Based on our findings and those of existing theoretical frameworks (Deniz et al., 2008; Ha et al., 2012) and related studies of factors influencing evolution acceptance (Nadelson & Hardy, 2015), we, therefore, define *acceptance* (as measured by the MATE) as including one or both of the following dimensions:

(1) Agreement with the fact that all life, including humans, has changed over extensive periods of time. This includes agreement with scientific data supporting evolution.
(2) Agreement that evolutionary science is credible. This includes disagreement with non-scientific ideas about the history of life.

Does the MATE Conflate Belief With Acceptance?

In the consideration of whether of the MATE could conflate acceptance with religious belief (per concerns of Hogan, 2000; Smith, 2010; Wagler & Wagler, 2013), it is key to discuss the items in the *Credibility* dimension with potential religious or teleological undertones (A7, A9, A14, A19). A young Earth view, as discussed in A7, is likely drawn from the Usher's chronology date of creation at 4004 BC (about 6,000 years ago). Items A9 and A19 state that species do not change and state that all life appeared at the same time (A19) and does not change (A9). We would expect these items to be in agreement with belief in a single creation event by a supernatural being (God). In contrast, A14 directly states that evolution is incorrect due to a particular religious dogma (the Biblical account of creation).

We cannot reject the idea that some items may directly relate to certain religious beliefs. However, most items do not. For example, a participant could agree that the Earth is less than 6,000 years old due to reasons other than creationist beliefs (such as lack of numeracy skills or unfamiliarity with the age of the Earth). In contrast, agreeing with A14 requires the participant to agree with the Biblical account of creation. A19 is certainly religiously charged. However, A19 does not specifically state that a creation event *by a supernatural force* is the reason that all living things came into existence at about the same time. For example, an individual could agree with A19 if they misunderstood an event like the Cambrian Explosion.

Moving beyond MATE items with religious undertones, we wonder if religious conflation matters. Does conflating acceptance damage the integrity of the MATE as a measure of evolution acceptance in post-secondary students from the Midwestern United States? The data suggest no, as we found that items with religious undertones (A7, A9, A14, and A19) had excellent fit with the Rasch model and were consistent with other items loading onto the *Credibility* dimension. While some may find this conflation bothersome, we know that religiosity has a significant relationship with evolution non-acceptance in the United States and worldwide (e.g., Hokayem & BouJaoude, 2008; Nadelson & Hardy, 2015; Rutledge & Mitchell, 2002; Trani, 2004). Perhaps elements of religious doctrine should be accounted for in an evolution acceptance instrument. Since we cannot untangle the two issues, it is imperative to address this directly in a measure of acceptance, and acknowledge that conflation with religion could be problematic in less religiously conservative populations.

Exploring Evolution Acceptance "Profiles"

We found that the MATE could separate students into one of five acceptance profiles, and that the model was generalizable to an external sample. Hence clustering expresses defined groups in the data on this sample that can be generalized to other samples to the same extent that the results of any quasi-experimental study can be generalized. Our data, therefore, suggest that these clusters may be common among many college students in a general education science course.

In our Midwest U.S. population, a majority of students fell into one of the two acceptance profiles (Clusters 1 and 4), and the remaining 36% were characterized by neutral views of evolution (Cluster 5, 17.8%), accepting facts but rejecting the credibility of evolution (Cluster 2, 12.1%), or rejecting both facts and credibility of evolution (Cluster 3, 6.3%). We did not have a profile that accepted the credibility of evolution but rejected the facts and supporting data for it in either the target or validation sample.

What is the qualitative distinction between a person who accepts the *facts*, but rejects the *credibility* of the theory (Cluster 2), and a person who rejects both the *facts* and the *credibility* of the theory (Cluster 3)? It is particularly interesting that Cluster 3 students recognize that scientists accept evolution, but Cluster 2 students do not. These phenomena can perhaps best be explained by the idea that most students enter college with dualistic epistemologies (Perry, 1970), and these beliefs are important to consider when teaching evolution. It is possible that Cluster 2 students were epistemological dualists, viewing the world in terms of black/white or right/wrong. Students with dualistic worldviews are less comfortable with data they perceive to be ambiguous (e.g., the *Credibility* dimension), and are often less knowledgeable and accepting of evolution (Sinatra, Southerland, McCounaughy, & Demastes, 2003). Dualistic students will also passively ignore information that is contradictory to their worldview, perhaps leading to low measures for knowledge of macroevolution in Cluster 2. In contrast, it is possible that Cluster 3 students had more of a multiplistic epistemology (Perry, 1970), that is, they thought both they and scientists were entitled to their own opinion. Unfortunately, views such as this could lead to suggestions that intelligent design, young Earth creationism, and evolution should hold equal status in schools, especially since Cluster 3 had adamant rejection of both evolution *facts* and *credibility*.

On a positive note, it is heartening that about one quarter (23%) of the students sampled expressed high acceptance of evolution, and that over half (~58%) of students expressed either neutral or moderate acceptance. Ingram and Nelson (2006) found that students displaying neutral levels of acceptance at the beginning of instruction tended to show among the greatest gains in response to pedagogy targeting evolution. This suggests that, with the help of quality instruction around evolution, it may be feasible to assist a majority of students in attainment of high levels of evolution acceptance.

Exploring Evolution Knowledge and Acceptance

Some authors note that rejection of evolution can serve as a barrier to developing knowledge about it (Coburn, 1994; Scharmann, 1990). Other research has shown that rejection of evolution does not affect the ability to learn about evolution; in this case, about natural selection (Bishop & Anderson, 1990; Demastes et al., 1995; Sinatra et al., 2003).

Our study supports an ongoing line of research that documents a significant positive relationship between knowledge of macroevolution and evolution acceptance (both acceptance of facts and acceptance of credibility). Walter (2013) and Nadelson and Southerland (2010a) also found that knowledge of macroevolution was significantly correlated with acceptance as measured by the MATE, with significant Pearson's r values of 0.47 and 0.49 (Nadelson & Southerland, 2010b; Walter, 2013).

We found significant correlations between knowledge of macroevolution and both MATE acceptance dimensions: *Facts* ($r = 0.39$) and *Credibility* ($r = 0.54$). Correlations of this nature support the need for teaching both macroevolution and natural selection concepts in college courses (Catley, 2006; Padian, 2010). In particular, we encourage instructors to teach macroevolution for its potential influence in improving students' perceptions of the credibility of evolutionary science and related rejection of non-scientific ideas.

However, it is important to consider whether the knowledge–acceptance correlation is due to the MATE conflating acceptance with knowledge of evolution (e.g., Smith, 2010). In our study, students with the highest acceptance of the *Facts* dimension (Clusters 1 and 4) also had the highest knowledge of macroevolution. However, this should not imply that the MATE is conflating knowledge with acceptance. The wording of items on the MATE clearly gets at *perceptions of scientific facts* rather than knowledge of the facts themselves. It is interesting that the correlation between knowledge and acceptance of facts ($r = 0.39$) is actually lower than the correlation of knowledge with acceptance of credibility ($r = 0.54$). If the *Facts*, dimension were indeed conflated with knowledge, the correlation between this dimension of acceptance and knowledge of macroevolution would be much higher. This conclusion could perhaps be criticized on grounds that the MUM has been so strongly criticized (Novick & Catley, 2012). However, Romine and Walter (2014) demonstrated the logical faults in many of these criticisms, and utilized many of Novick and Catley's valid criticisms to construct a revised version of the MUM (Romine & Walter, in Review) which was used to collect knowledge measures in this study.

Patterns within Cluster 2 (accept facts, reject credibility) provide additional support that the MATE does not conflate with knowledge with acceptance. Cluster 2 students had a low average *Credibility* score, despite high acceptance of evolution *Facts*. Furthermore, Cluster 2 students had significantly lower knowledge of macroevolution than the other clusters that accepted the *Facts* dimension (Clusters 1 and 4). It is, therefore, likely that the Cluster 2 students accept *Facts* due to feelings of truth or certainty about the facts rather than knowledge of those facts.

Ha et al. (2012) also discuss evolution acceptance as informed by both a cognitive domain (knowledge) and an affective domain (feelings of certainty). They found a strong correlation between MATE scores and feelings of certainty. Our findings support this correlation between acceptance and feelings of certainty, as feelings of correctness, conviction, or rightness could be elicited by terms like ambiguity and incorrectness in the *Credibility* dimension. Furthermore, considering scientific facts (or non-scientific ideas) on the MATE may also elicit feelings of this nature.

Implications for Future Research

Discussion of future research goals could perhaps best be accomplished through a discussion of this study's limitations and stimulating questions that have arisen out of the data, but are as of now left unanswered. We will first discuss limitations of the MATE itself. Using a 2-scale treatment from the Rasch perspective, we found that items A5, A11, and A15 did not fit well with the instrument. Item A15 (Humans exist today in essentially the same form in which they always have) appears to slightly misfit with the *Facts* dimension. This item is interesting in that it is the only negatively worded item on the *Facts* dimension. There is no obvious reason for this misfit. We suspect, however, that the word "always" may solicit differing views about the human timescale and the breadth of geologic time (Catley & Novick, 2009) that could introduce error into students' responses.

Item A5 (Most scientists accept evolution to be a scientifically valid theory) was the item with the lowest factor loading, 0.25 onto *Credibility* and no cross-loading onto *Facts*. It also misfit with the Rasch *Credibility* scale. This indicates that A5 did not fit well into either dimension. It is telling that level of agreement with A5 was similar across all five cluster profiles (Figures 3 and 4) indicating it is perhaps the least useful item on the MATE for drawing distinctions between levels of evolution acceptance. The wording of A5 gets more at feelings about scientists (e.g., Nadelson and Hardy, 2015) than feelings about evolution, and, therefore, can possibly be eliminated in future studies using the MATE.

A11 (The age of the Earth is at least 4 billion years) had a relatively decisive loading of 0.58 onto the *Facts* dimension. However, as discussed above, A11 displayed high outfit (1.52) with the Rasch model. Further analysis indicates that presence of a small number of anomalous responses at both extremes of the scale is to blame. To what can we attribute this? In light of our prior discussion, there is no reason to implicate conflation with knowledge or religion. Rather, we notice that this is the only question on the MATE that solicits students' acceptance of a large number. Given that misunderstanding of scale is endemic even at the college level (Libarkin, Kurdziel, & Anderson, 2007), we posit that responses on A11 may be getting confounded with understanding of numeracy in some students.

Although Wagler and Wagler (2013) explore the effect of adding dimensions to the MATE's measurement model on model fit, this study is the first to actually use the MATE as a multiple-scale instrument. This study is also the first to treat the MATE as a measure of a categorical latent variable. Our use of multiple measures (the Dunn Index and SSE) for cluster validity lends confidence that five distinct and unique cluster profiles exist in our sample of students. Further, we demonstrate the generalizability of these profiles for a large external sample in the Western United States. But to what extent are these results generalizable to other regions or academic contexts? While our data provide reason for optimism about the generalizability of our conclusions, we echo Wagler and Wagler in that generalization beyond our study should be exercised with a healthy amount of skepticism.

In addition to being generalizable, the five profiles derived from clustering have the benefit of making sense and resonating with our experience. Biology educators are certainly familiar with the high-, low-, and neutral-acceptance groups. It also makes sense that asymmetrical acceptance profiles would exist in our students. What is more likely to vary from place to place, and across academic levels, is the proportion of students in each profile. About a quarter of our students fell into the high acceptance profile, and about a fifth fell into one of the low acceptance profiles. We may expect to see this type of profile distribution in other Midwestern college communities, and inspection of Table 3 shows that this distribution is quite similar in our Western United States validation sample. These conclusions can be tested and strengthened in future studies through additional data from the MATE, other evolution acceptance instruments such as the GAENE and EALS, and interviews with students. For now, we leave detailed exploration of how this distribution changes in students in other regions of the United States and beyond, and how students' profiles change with academic level, to future studies.

Per the spirit of research, this study introduces more questions than it answers. We wish to conclude this article by elaborating on some of the more provoking questions that future research can explore and address more fully. First, we found that A5 is among the least useful items on the MATE. It does not fit well with either scale and its responses do not vary significantly between clusters. Can we conclude, then, that "scientists accept evolution" represents a different phenomenon "credibility of evolution" in general? What makes a person agree that scientists accept evolution? Is this rooted in students' understanding of the nature of science in general? Another lingering question is why the group with lowest average acceptance (Cluster 3) tended to acknowledge the explanatory value of evolutionary theory? Additionally, why did groups with the lowest average acceptance (Cluster 3), uniform moderate acceptance (Cluster 4), and neutral acceptance (Cluster 5) have moderate and relatively equal knowledge levels of macroevolution? This is an important result that needs explaining, but we leave this and other important questions to further inquiry. We plan to continue investigations of the MATE and the nature of evolution acceptance and encourage others to do comparable studies that continue to advance the field of measurement and theory in evolution education.

References

Athanasiou, K., & Papadopoulou, P. (2012). Conceptual ecology of the evolution acceptance among Greek education students: Knowledge, religious practices and social influences. International Journal of Science Education, 34, 903–924. doi: 10.1080/09500693.2011.586072

Barone, L. M., Petto, A. J., & Campbell, B. C. (2014). Predictors of evolution acceptance in a museum population. Evolution: Education and Outreach, 7, 23. doi: 10.1186/s12052-014-0023-2

Berkman, M. B., & Plutzer, E. (2011). Defending evolution in the courtroom, but not in the classroom. Science, 331, 404–405.

Bishop, B. A., & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. Journal of Research in Science Teaching, 27, 415–427. doi: 10.1002/tea.3660270503

Boone, W. J., Townsend, J. S., & Staver, J. (2011). Using Rasch theory to guide the practice of survey development and survey data analysis in science education and to inform science reform efforts: An exemplar utilizing STEBI self-efficacy data. Science Education, 95(2), 258–280.

Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. Science Education, 90(2), 253–269.

Brewer, C., & Smith, D. (2011). Vision and change in undergraduate biology education: A call to action. Washington, DC: AAAS.

Butler, W. (2009). Does the nature of science influence college students' learning of biological evolution (Unpublished doctoral dissertation). Tallahassee, FL: Florida State University.

Bybee, R. W. (1997). Achieving scientific literacy: From purpose to practices. Portsmouth, NH: Heinemann.

Catley, K. M., & Novick, L. R. (2009). Digging deep: Exploring college students' knowledge of macroevolutionary time. Journal of Research in Science Teaching, 46, 311–332.

Catley, K. M. (2006). Darwin's missing link—A novel paradigm for evolution education. Science Education, 90, 767–783.

Coburn, W. W. (1994). Belief, understanding, and the teaching of evolution. Journal of Research in Science Teaching, 31, 583–590.

Coleman, J., Stears, M., & Dempster, E. (2015). Student teachers' understanding and acceptance of evolution and the nature of science. South African Journal of Education, 35(2), 1–9. doi: 10.15700/saje.v35n2a1079

Collins, L. M., & Lanza, S. T. (2010). Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences. Hoboken, NJ: John Wiley & Sons, Inc.

Crivellaro, F., & Sperduti, A. (2014). Accepting and understanding evolution in Italy: A case study from a selected public attending a Darwin Day celebration. Evolution: Education and Outreach, 7, 13.

Dagher, Z. R., & BouJaoude, S. (1997). Scientific views and religious beliefs of college students: The case of biological evolution. Journal of Research in Science Teaching, 34, 429–445.

Demastes, S. S., Settlage, J., & Good, R. G. (1995). Students' conceptions of natural selection and its role in evolution: Cases of replication and comparison. Journal of Research in Science Teaching, 32, 535–550.

Deniz, H., & Sahin, E. A. (2016). Exploring the factors related to acceptance of evolutionary theory among Turkish preservice biology teachers and the relationship between acceptance and teaching preference. Electronic Journal of Science Education, 20(4), 21–43.

Deniz, H., Cetin, F., & Yilmaz, I. (2011). Examining the relationships among acceptance of evolution, religiosity, and teaching preference for evolution in Turkish preservice biology teachers. Reports of the National Center for Science Education, 31(4), 1.1–1.9.

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

Diamantidis, N. A., Karlis, D., & Giakoumakis, E. A. (2000). Unsupervised stratification of cross-validation for accuracy estimation. Artificial Intelligence, 116(1), 1–16.

Donnelly, L. A., Kazempour, M., & Amirshokoohi, A. (2009). High school students' perceptions of evolution instruction: Acceptance and evolution learning experiences. Research in Science Education, 39, 643–660.

Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. Journal of Cybernetics, 3, 32–57.

Evans, E. M. (2008). Conceptual change and evolutionary biology: A developmental analysis. In S. Vosniadou (Ed.), International handbook of research on conceptual change (pp. 263–295). New York, NY: Routledge.

Flanagan, J. C., & Roseman, J. E. (2011). *Assessing middle and high school students' understanding of evolution with standards-based items*. Paper presented at the Annual Conference of the National Association for Research in Science Teaching, Orlando, FL.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. Psychological Bulletin, 76(5), 378–382.

Gasson, S. (2004). Rigor in grounded theory research: An interpretive perspective on generating theory from qualitative field studies. In A. Woszczynski (Ed.), The handbook of information systems research (pp. 79–102). Hershey, PA: Idea Group.

Glaser, B. G. (1978). Theoretical sensitivity: Advances in the methodology of grounded theory. Mill Valley, CA: Sociology Press.

Glaser, B. G. (1992). Basics of grounded theory analysis: Emergence vs. forcing. Mill Valley, CA: Sociology Press.

Glaser, B. G., & Strauss, A. L. (1967). The discovery of grounded theory: Strategies for qualitative research. New York, NY: Aldine De Gruyter.

Glaser, B. G., & Strauss, A. L. (2009). The discovery of grounded theory: Strategies for qualitative research. New York, NY: Aldine Transaction Publishers.

Glaze, A. L., Goldston, M. J., & Dantzler, J. (2014). Evolution in the southeastern USA: Factors influencing acceptance and rejection in pre-service science teachers. International Journal of Science and Mathematics Education, 13, 1189–1209. doi: 10.1007/s10763-014-9541-1

Großschedl, J., Konnemann, C., & Basel, N. (2014). Pre-service biology teachers' acceptance of evolutionary theory and their preference for its teaching. Evolution: Education and Outreach, 7, 1–16.

Ha, M., Haury, D. L., & Nehm, R. H. (2012). Feeling of certainty: Uncovering a missing link between knowledge and acceptance of evolution. Journal of Research in Science Teaching, 49, 95–121.

Ha, M., Baldwin, B. C., & Nehm, R. H. (2015). The long-term impacts of short-term professional development: Science teachers and evolution. Evolution: Education and Outreach, 8, 11. doi: 10.1186/s12052-015-0040-9

Hawley, P. H., Short, S. D., McCune, L. A., Osman, M. R., & Little, T. D. (2011). What's the matter with Kansas? The development and confirmation of the Evolutionary Attitudes and Literacy Survey (EALS). Evolution: Education and Outreach, 4, 117–132. doi: 10.1007/s12052-010-0294-1

Hogan, K. (2000). Exploring a process view of students' knowledge about the nature of science. Science Education, 84, 51–70.

Hokayem, H., & BouJaoude, S. (2008). College students' perceptions of the theory of evolution. Journal of Research in Science Teaching, 45, 395–419. doi: 10.1002/tea.20233

Ingram, E. L., & Nelson, C. E. (2006). Relationship between achievement and students' acceptance of evolution or creation in an upper-level evolution course. Journal of Research in Science Teaching, 43, 7–24.

Johnson, R. L., & Peeples, E. E. (1987). The role of scientific understanding in college: Student acceptance of evolution. The American Biology Teacher, 49, 93–96+98.

Korte, S. E. (2003). The acceptance and understanding of evolutionary theory among Ohio secondary life science teachers (Unpublished masters thesis). Athens, OH: Ohio University.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159–174.

Lawson, A. E. (1983). Predicting science achievement: The role of developmental level, disembedding ability, mental capacity, prior knowledge and beliefs. Journal of Research in Science Teaching, 20, 117–129.

Lawson, A. E., & Worsnop, W. A. (1992). Learning about evolution and rejecting a belief in special creation: Effects of reflective reasoning skill, prior knowledge, prior belief and religious commitment. Journal of Research in Science Teaching, 29, 143–166.

Libarkin, J. C., Kurdziel, J. P., & Anderson, S. W. (2007). College student conceptions of geological time and the disconnect between ordering and scale. Journal of Geoscience Education, 55, 413–422.

Linacre, J. M. (2005). A user's guide to WINSTEPS MINISTEP Rasch-model computer programs. Chicago, IL: Winsteps.

Losh, S. C., & Nzekwe, B. (2011). Creatures in the classroom: Preservice teacher beliefs about fantastic beasts, magic, extraterrestrials, evolution and creationism. Science & Education, 20, 473–489.

Lovely, E. C., & Kondrick, L. C. (2008). Teaching evolution: Challenging religious preconceptions. Integrative and Comparative Biology, 48, 164–174.

Manwaring, K. F., Jensen, J. L., Gill, R. A., & Bybee, S. M. (2015). Influencing highly religious undergraduate perceptions of evolution: Mormons as a case study. Evolution: Education and Outreach, 8, 23. doi: 10.1186/s12052-015-0051-6

Masters, G. N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149–174.

Masters, G. N. (1988). Item discrimination: When more is worse. Journal of Educational Measurement, 25(1), 15–29.

Mayr, E. (1982). The growth of biological thought. Cambridge, MA: Harvard University Press.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. Biochemia Medica, 22(3), 276–282.

McKeachie, W. J., Lin, Y. G., & Strayer, J. (2002). Creationist *vs.* evolutionary beliefs: Effects on learning biology. The American Biology Teacher, 64, 189–192.

Miller, J. D., Scott, E. C., & Okamoto, D. (2006). Public acceptance of evolution. Science, 313, 765–766. doi: 10.2307/3846903

Moore, R., & Cotner, S. (2009). Educational malpractice: The impact of including creationism in high school biology courses. Evolution: Education and Outreach, 2, 95–100.

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

Nadelson, L. S., & Sinatra, G. M. (2009). Educational professionals' knowledge and acceptance of evolution. Evolutionary Psychology, 7, 490–516.

Nadelson, L. S., & Southerland, S. A. (2010a). Development and evaluation for a measuring understanding of macroevolutionary concepts: Introducing the MUM. Journal of Experimental Education, 78, 151–190. doi: 10.1080/00220970903292983

Nadelson, L. S., & Southerland, S. A. (2010b). Examining the interaction of acceptance and understanding: How does the relationship change with a focus on macroevolution. Evolution: Education and Outreach, 4, 82–88.

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, 1637–1666.

Nadelson, L. S., Jorcyk, C., Yang, D., Jarratt Smith, M, Matson, S., Cornell, K., & Husting, V. (2014). I just don't trust them: The development and validation of an assessment instrument to measure trust in science and scientists. School Science and Mathematics, 114(2), 76–86.

National Association of Biology Teachers. (2002). Statement on teaching evolution. Retrieved August 8, 2015 from http://www.nabt.org

Newport, F. (2004, November 19). Third of Americans say evidence has supported Darwin's evolution theory. Gallup News Service. Retrieved from www.galluppoll.com.

Nehm, R. H., & Reilly, L. (2007). Biology majors' knowledge and misconceptions of natural selection. Bioscience, 57, 263–272.

Nehm, R. H., & Schonfeld, I. S. (2007). Does increasing biology teacher knowledge of evolution and the nature of science lead to greater preference for the teaching of evolution in schools? Journal of Science Teacher Education, 18, 699–723.

NGSS Lead States. (2013). Next generation science standards: For states, by states. Washington, D.C.: National Research Council.

Novick, L. R., & Catley, K. M. (2012). Assessing students' understanding of macroevolution: Concerns regarding the validity of the MUM. International Journal of Science Education, 34, 2679–2703. doi: 10.1080/09500693.2012.727496

Nunnally, J. C. (1978). Psychometric theory (2nd edition). New York, NY: McGraw-Hill.

Padian, K. (2010). How to win the evolution war: Teach macroevolution! Evolution: Education and Outreach, 3, 206–214. doi: 10.1007/s12052-010-0213-5

Perry, W. G. (1970). Forms of intellectual and ethical development in the college years. New York: Holt, Rinehart and Winston.

Pew Research Center. (2015). Public and scientists' views on science and society. Retrieved from: http://www.pewinternet.org/2015/01/29/public-and-scientists-views-on-science-and-society/

Popham, W. J., & Popham, J. W. (2005). Classroom assessment: What teachers need to know. Washington, DC: Pearson/Allyn and Bacon.

Raîche, G. (2005). Critical eigenvalue sizes in standardized residual principal components analysis. Rasch Measurement Transactions, 19, 1012.

Rice, J. W., Olson, J. K., & Colbert, J. T. (2011). University evolution education: The effect of evolution instruction on biology majors' content knowledge, attitude toward evolution, and theistic position. Evolution: Education and Outreach, 8, 8. doi: 10.1186/s12052-015-0036-5

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, 1–17. doi: 10.1186/s12052-014-0024-1

Robbins, J. R., & Roy, P. (2007). The natural selection: Identifying & correcting non-science student preconceptions through an inquiry-based, critical approach to evolution. The American Biology Teacher, 69, 460–466.

Romine, W. L., & Walter, E. M. (2014). Assessing the efficacy of the measure of understanding of macroevolution as a valid tool for undergraduate non-science majors. International Journal of Science Education, 36, 2872–2891. doi: 10.1080/09500693.2014.938376

Rutledge, M. L., & Sadler, K. C. (2007). Reliability of the measure of acceptance of the theory of evolution (MATE) instrument with university students. The American Biology Teacher, 69, 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.

Rutledge, M. L., & Mitchell, M. A. (2002). High school biology teachers' knowledge structure, acceptance and teaching of evolution. The American Biology Teacher, 64, 21–28.

Scharmann, J. C. (1990). Enhancing an understanding of the premises of evolutionary theory: The influence of a diversified instructional strategy. School Science and Mathematics, 90, 91–100.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. Psychological Assessment, 8, 350.

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

Smith, M. U. (2010). Current status of research in teaching and learning evolution: I. Philosophical/epistemological issues. Science & Education, 19(6–8), 523–538.

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

Strauss, A. L. (1987). Qualitative analysis for social scientists. New York, NY: Cambridge University Press.

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. Duschl & R. Hamilton (Eds.), Philosophy of science, cognitive psychology, and educational theory and practice (pp. 147–176). Albany, NY: SUNY.

Tennant, A., & Connaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Care and Research, 57, 1358–1362.

Theodoridis, S., & Koutroumbas, K. (2001). Machine learning and its applications. Berlin Heidelberg: Springer.

Trani, R. (2004). I won't teach evolution; it's against my religion. And now for the rest of the story... The American Biology Teacher, 66, 419–427.

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, 2278–2298.

Walter, E. M. (2013). The influence of pedagogical content knowledge (PCK) for teaching macroevolution on student outcomes in a general education biology course (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (#3576041).

Walter, E. M., & Romine, W. L. (in preparation). Exploring knowledge of macroevolution through the Revised Measure of Understanding of Macroevolution (Re-MUM).

Walter, E. M., Halverson, K. L., & Boyce, C.-J. (2013). Investigating the relationship between college students' acceptance of evolution and tree thinking understanding. Evolution: Education and Outreach, 6, 26.

Wiles, J. R., & Alters, B. (2011). Effects of an educational experience incorporating an inventory of factors potentially influencing student acceptance of biological evolution. International Journal of Science Education, 33, 2559–2585. doi: 10.1080/09500693.2011.565522

Wright, B. D., & Linacre, J. M. (1996). Reasonable mean-square fit values, part 2. In J. M. Linacre (Ed.), Rasch measurement transactions (p. 370). Chicago, IL: Mesa Press.

Wright, B. D., & Stone, M. A. (1979). Best test design. Chicago, IL: Mesa Press.

Wiley, C. H. (2003). Reflection on the difficulties of teaching evolution in high school biology (Unpublished master's thesis). Athens, GA: University of Georgia.

Yousuf, A., bin Daud, M. A., & Nadeem, A. (2011). Awareness and acceptance of evolution and evolutionary medicine among medical students in Pakistan. Evolution: Education and Outreach, 4, 580–588. doi: 10.1007/s12052-011-0376-8

Supporting Information

Additional Supporting Information may be found in the online version of this article.