

12-15-2021

Measuring Science Teachers' Emotional Experiences with Evolution Using Real World Scenarios

William Lee Romine

Wright State University - Main Campus, william.romine@wright.edu

Rutuja Mahajan

Amber Todd

Wright State University - Main Campus, amber.todd@wright.edu

Follow this and additional works at: <https://corescholar.libraries.wright.edu/biology>



Part of the [Biology Commons](#), [Medical Sciences Commons](#), and the [Systems Biology Commons](#)

Repository Citation

Romine, W. L., Mahajan, R., & Todd, A. (2021). Measuring Science Teachers' Emotional Experiences with Evolution Using Real World Scenarios. *Eurasian Journal of Science and Environmental Education*, 1 (1), 1-26.

<https://corescholar.libraries.wright.edu/biology/892>

This Article is brought to you for free and open access by the Biological Sciences at CORE Scholar. It has been accepted for inclusion in Biological Sciences Faculty Publications by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.



Eurasian Journal of Science and Environmental Education


Volume 10, Issue 1, 1 - 26.

<https://www.ejsee.com/>

Measuring Science Teachers' Emotional Experiences with Evolution using Real World Scenarios

William Romine * 
Wright State University, USA

Rutuja Mahajan 
Wright State University, USA

Amber Todd 
Wright State University, USA

Received: May 30, 2021 • Revised: October 16, 2021 • Accepted: December 15, 2021

Abstract: Low acceptance of evolution remains an obstacle to quality biology instruction. We develop and utilize a novel assessment which measures emotional experience in light of real-world evolution education scenarios. We presented 296 science teachers 4 pro-evolution and 8 anti-evolution scenarios and asked them to rate their levels of joy, anger, sadness, fear, disgust, shame, and guilt elicited by that scenario on an ordinal 5-point scale. We used exploratory factor analysis to extract the most important dimensions in the teachers' responses, Rasch analysis to explore the validity of the extracted subscales, and stepwise regression to find the most important factors driving emotional dispositions. We extracted 3 factors: (1) pro-evolution experience (positive emotions on pro-evolution and negative emotions on anti-evolution scenarios), (2) anti-evolution experience (negative emotions on pro-evolution and positive emotions on anti-evolution scenarios), and (3) feelings of regret over anti-evolution scenarios (shame and guilt on anti-evolution scenarios). Acceptance of evolution facts and a non-theistic religious orientation were positively related to pro-evolution experience. Anti-evolution experience was predicted by lack of microevolution acceptance and lack of teacher preparation. Feelings of regret around anti-evolution scenarios were driven by acceptance of evolution facts and lower levels of teacher preparation. This work advances our understanding of how teachers relate affectively to the theory of evolution and offers empirical insight into ways to improve dispositions about evolution.

Keywords: *Emotion, evolution acceptance, evolution education, measurement, psychometric.*

To cite this article: Romine, W., Mahajan, R., & Todd, A. (2021). Measuring science teachers' emotional experiences with evolution using real world scenarios. *Eurasian Journal of Science and Environmental Education*, 10(1), 1-27. <https://doi.org/10.30935/ejsee/11868>

Introduction

Despite the prevalence of anti-evolution views in the United States (US), US education policy emphasizes the importance of teaching evolution. Evolution is one of four Disciplinary Core Ideas in life sciences in the Next Generation Science Standards (NGSS) (NGSS Lead States, 2013) and both the National Association of Biology Teachers (NABT) and the American Association for the Advancement of Science (AAAS) have formal statements identifying evolution as a foundational component of scientific literacy (Brewer & Smith, 2011; NABT, 2011). Evolution is also one of 5 core content areas outlined in *Vision and Change in Undergraduate Biology Education* (Brewer & Smith, 2011). Although policy discussions related to science education have reached an overwhelming consensus that it is essential to help students understand evolution, biology teachers often do not rise to this task, especially in localities with persistent Christian (Bowman, 2008) and Islamic (Fouad, 2018) fundamentalism. Teachers in these localities who wish to teach evolution often do not feel prepared to face opposition from parents and school administrators, while others may not feel confident in their knowledge and ability to teach evolution in the first place (Romine et al., 2014). We see that the goal of understanding and improving the quality of evolution pedagogy that students receive is not a simple matter, but rather a problem rooted in complex emotional experiences rising from lack of community support (Cleaves & Toplis, 2007), religion (Trani, 2004), and misunderstandings about evolution (Berkman et al., 2008).

Helping science teachers develop acceptance of evolution as a valid theory is a paramount professional development goal (Sickel & Friedrichsen, 2013), but as of now, the conceptual scope for measuring evolution acceptance is defined narrowly. Up to this point, instrumentation for measuring evolution acceptance has focused on appraisal of statements related to the theory of evolution (Romine et al., 2017), but measurement of teachers' emotional experiences related to

* Corresponding author:

William Romine, Wright State University, Dayton, USA. ✉ romine.william@gmail.com

© 2021 by the authors(s); licensee EJSEE by Bastas, Cyprus. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>).

evolution education is more limited (Heddy & Sinatra, 2013; Hawley & Sinatra, 2019). We need more quantitative basis for understanding the emotional experiences harbored by science teachers in both the United States and abroad. The current dearth of research on quantifying teachers' emotional experiences makes it difficult to generalize on the actual feelings that teachers are experiencing and how these feelings may relate to school and community environment, personal education and religious background, and ultimately teaching practice.

Current validated tools for measuring evolution acceptance focus on giving the participant a statement about evolution and asking them to rate their acceptance of that statement on a Likert scale. Current instrumentation includes the Measure of Acceptance of the Theory of Evolution (MATE) (Rutledge & Warden, 1999), the Inventory of Student Evolution Acceptance (I-SEA) (Nadelson & Southerland, 2012), the Evolutionary Attitudes and Literacy Survey (EALS) (Hawley et al., 2011), and the Generalized Acceptance of Evolution Evaluation (GAENE) (Smith et al., 2016). The MATE, I-SEA, and GAENE are similar in that they provide a statement about evolution such as *Humans do not evolve; they only change their behavior* (from the I-SEA), *Simple organisms such as bacteria change over time* (from the GAENE), or *Evolution is not a scientifically valid theory* (from the MATE), and then ask the participant to rate the extent to which he/she agrees with the statement. The EALS is similarly written in a Likert format, but instead focuses measuring variables which may affect acceptance of evolution, some of which include religiosity, knowledge, misconceptions, and distrust of science. Teachers and students may understand what evolution states, and even accept that evolution is necessary for the study of biology, but this bears sharp contrast to the emotional acceptance, and subsequent personal belief, required to accept the theory as part of one's worldview (Cobern, 1994).

The current approach for measuring evolution acceptance focuses on cognitive appraisal, which is an initial trigger for emotion (Scherer, 2005). Although the emotional process may be initiated cognitively through data obtained by the senses, emotion goes beyond cognition in that it manifests as physical stimulus, action motivation, communication, and behavior (Ekman & Davidson, 1994; Scherer, 2005), which are related directly to the perceived relevance of a particular object or event of attention to one's well-being (Frijda, 1986). While it is interesting to understand appraisal related to the theory of evolution, quantifying emotional experience is critical for understanding the actual feelings teachers harbor towards evolution. Teachers' emotions may serve as a critical factor influencing how they ultimately decide to teach evolution, and given that affect is a key component of contemporary theoretical models of evolution acceptance (Deniz et al., 2008; Ha et al., 2012), measurement of emotion is important in understanding acceptance as it is currently framed.

In this study, we introduce new instrumentation to measure teachers' emotional experiences around evolution: the Evolution Emotion Assessment with Real-world Scenarios (E-EARS). We first describe the framework behind the E-EARS, and then explore its validity using the Rasch model. Finally, we use it to investigate factors influencing teachers' reported emotional experiences in response to the scenarios on the E-EARS. Through this process, we address three questions:

1. How do teachers' discrete emotions cluster together to form unique emotional experiences?
2. What is the efficacy of real-world scenarios in allowing us to draw valid inferences about teachers' distinct emotional experiences?
3. What role does cognitive appraisal play in predicting teachers' emotional experiences, and what other factors are important in driving the emotional experiences teachers harbor with respect to the scenarios we present in the E-EARS?

Literature Review

Work toward understanding thinking about evolution

Two prominent models have been developed to represent the process by which thinking about evolution occurs (Deniz et al., 2008; Ha et al., 2012). The model put forth by Deniz and colleagues considers evolution thinking in terms of cognitive, affective, and contextual domains. Their model conceptualizes evolution acceptance as a product of both cognition and affection around the theory of evolution. While the model of Deniz et al. (2008) focuses on evolution thinking in general, the model of Ha et al. (2012) presents a more constrained focus on evolution acceptance specifically, suggesting that this is mediated through both conscious and unconscious pathways. Foremost, people will make a conscious decision on whether or not they accept evolution through their understanding of evolution and their reasoning around what they perceive as evidence for or against evolution. Indeed, facilitation of conscious experiences is an important goal of evidence-based pedagogies focused on evolution (Beard et al., 2014). Ha et al. (2012) further propose that these conscious experiences will merge with past experiences, values, and social interactions to generate unconscious feelings which manifest as an underlying feeling of certainty. This indicates a bridge between cognition and emotional experience. Evolution acceptance as it is currently defined serves as that bridge, but emotional experience itself needs more exploration.

Validated measures of evolution acceptance

Both qualitative and quantitative methods have been utilized to understand evolution acceptance. Qualitative methods which have been utilized include open response (Nehm & Schonfeld, 2008; Robbins & Roy, 2007) and interview protocols

(Donnelly et al., 2009; Nehm & Reilly, 2007). In this study, we focus on quantitative measurement of emotional experience related to evolution which has been conceptualized as 'evolution acceptance' in current instrumentation. Measurement of evolution acceptance has been pursued with significant interest for over two decades. Johnson and Peeples (1987) undertook one of the first attempts to measure evolution acceptance using a survey with documented psychometric information. Just over a decade later, the Measure of Acceptance of the Theory of Evolution (MATE) was developed (Rutledge & Warden, 1999); it has been both utilized and criticized extensively since then (for a review, see Romine et al. (2017)). The MATE has been used as a single-dimensional measure, although Romine et al. (2017) found that it works better as a two-dimensional instrument measuring acceptance of facts and acceptance of credibility of statements related to evolution. The Inventory of Student Evolution Acceptance (I-SEA) and the Evolution Attitudes and Literacy Survey (EALS) are also constructed multi-dimensionally. The I-SEA was designed to improve upon the MATE by making its measures more fine-grained; it measures acceptance of statements related to microevolution, macroevolution, and human evolution (Nadelson & Southerland, 2012). The Generalized Acceptance of Evolution Evaluation (GAENE) is the most recent development in the line of traditional evolution acceptance measures. The authors argue that it measures evolution acceptance unidimensionally without the potential confound with knowledge of evolution (Smith et al., 2016). In contrast, the Evolution Attitudes and Literacy Survey (EALS) measures constructs peripheral to evolution acceptance including creationist reasoning, political/religious conservatism, knowledge, misconceptions, and exposure to evolution (Hawley et al., 2011). These instruments get at cognitive appraisal (Scherer, 2005), but are not able to elucidate the emotions experienced when encountering scenarios related to the theory of evolution.

Extending evolution acceptance to emotional experience

Emotion and rationality are often positioned as being antithetical in nature (Zembylas, 2002). However, present understanding of emotion suggests that emotional expression begins with appraisal, a rational process (Scherer, 2005). Evolution acceptance, as it is defined by current instrumentation, entails evaluation or appraisal of specific statements related to evolution. For example, asking for a person's agreement with statement about evolution such as *Humans do not evolve; they only change their behavior* (from the I-SEA), *Simple organisms such as bacteria change over time* (from the GAENE), or *Evolution is not a scientifically valid theory* (from the MATE) solicits the person's appraisal of a particular idea, but does little to elucidate how the person relates to the theory of evolution emotionally. While it is interesting to understand cognitive appraisal, going beyond this is critical for understanding the actual feelings harbored towards evolution and how these relate to the short-term decisions made regarding what and how to teach. Given that affect is a key component of contemporary theoretical models of evolution acceptance (Deniz et al., 2008; Ha et al., 2012), measurement of emotion is an important next step.

Study of Emotion in Evolution Education Contexts

The formal study of emotions was commenced by none other than Charles Darwin as reported in his book, *The Expression of the Emotions in Man and Animals* (Darwin, 1872). As his third book describing and applying evolutionary theory, Darwin made the case that the emotions occur as discrete modules, and subsequent actions or decisions stemming from particular emotions play a role in differential survival of animals. For this reason, Darwin argued that emotions tend to be experienced, expressed and addressed in relatively uniform ways across animal kingdom. Examples include, but are not limited to, the "fight or flight" reaction in response to fear, the tendency to avoid objects or environments that elicit disgust, or the desire to engage with objects or situations that elicit feelings of joy. Wundt (1896) was the first to propose conceptualizing emotion in terms of dimensions such as valence (positive to negative) and arousal (excited to calm), and tension (tense to relaxed), and viewed this as a way to better conceptualize psychological similarities and differences between the discrete emotions such that a person's complex in-the-moment experience can be more meaningfully characterized in a quantitative way.

Emotion in science education has been researched primarily in the context of constructs peripheral to emotion including personal (Romine & Sadler, 2016) and situational interest (Lamb et al., 2012), self-efficacy (Palmer, 2006; Schoon & Boone, 1998), motivation (Glynn et al., 2007), engagement (Thompson & Windschitl, 2005), and morality (Fowler et al., 2009). Identity formation (Czerniak & Schriver, 1994; Malone & Barabino, 2009), frustration or anxiety with curricular (Eastwell & MacKenzie, 2009) and classroom environment factors (Young, 1991), and emotional politics (Hargreaves, 1998; Lasky, 2000) have also received attention. Direct exploration of the construct of emotion specifically in science teachers has primarily entailed qualitative research on classroom practice within ethnographic (Zembylas, 2002) and case study methods (Zembylas, 2004) undertaken within a post-modernist framework (Arribas-Ayllon & Walkerdine, 2017).

Quantitative measurement of emotion in science education come about with the interest of connecting emotional experience related to evolution and conceptual change within an instructional context (Heddy & Sinatra, 2013). To this end, the Evolution Emotions Scale was developed to measure the emotions elicited about the specific topic to which the students were just introduced. This measure utilized the discrete emotions approach to measurement of emotions. Students were asked to rate their emotions on a 5-point Likert scale indicating agreement with the extent to which they expressed the emotions of enjoyment, hope, pride, anger, anxiety, shame, hopelessness, and boredom, after learning

about the topic (Heddy & Sinatra, 2013). A study by Hawley and Sinatra (2019) focused on measuring emotions about evolution in Christian science teachers in the context of a professional development program focused on reducing the perceived conflict between faith and science. They found that this program facilitated a reduction in negative emotions and an increase in positive emotions (Hawley & Sinatra, 2019). In this study, the authors took a dimensional measurement approach, putting positive and negative emotions together in order to generate respective scores based on valence. In our current work, we also seek to build on this work and bridge the discrete emotions and dimensional approaches in order to generate parsimonious quantitative models that yield meaningful qualitative interpretations.

Methodology

Sample

The MATE, I-SEA, and the E-EAR, along with questions about teachers' backgrounds and practices related to teaching evolution, were administered through Qualtrics survey software. Links to the survey were posted on the National Science Teachers Association (NSTA) listserv and on the Twitter page of a high school in the Midwestern United States, and targeted biology teachers. Forty teachers responded to the Twitter post, and 260 responded to the link on the NSTA listserv. Two hundred ninety-six teachers completed the survey. A majority of the respondents reported teaching high school (155, 52%). However, we also had 83 middle school teachers (28%), 32 elementary teachers (11%), and 25 college teachers (8%). A single kindergarten teacher also responded.

Of the teachers reporting their gender, 98 (33%) were male and 184 (62%) were female. A majority of these teachers were White (229, 77%). Black (24, 8%) and Hispanic (25, 8%) were also represented, along with 3 teachers who reported Asian ethnicity (1%). Fifteen teachers (5%) reported "other" ethnicities. Along with a variety of ethnicities, a variety of religions were represented in the study. One hundred seventy-three (58%) reported monotheistic religions including Christianity (129, 44%) and Islam (44, 15%). Fifty-nine (20%) reported atheistic and 35 (12%) reported agnostic faith. Three teachers (1%) reported Buddhist faith and no teachers reported Hindu faith. The remainder reported faiths other than these listed.

A majority of the teachers participating in this study indicated backgrounds, dispositions, and conditions favorable for teaching evolution in their classrooms. Two hundred-seventy (92%) of the teachers indicated that they personally accepted evolution as being valid and true. Consequently, 266 (90%) of the teachers indicated that they felt responsible for teaching evolution. One hundred twenty-two teachers (42%) indicated that they spent a moderate amount of time on evolution in their classes. One hundred (33.8%) indicated they spent a greater than moderate amount of time teaching evolution. The remainder spent little-to-no time on evolution. One hundred eighty-four (62%) of the teachers reported that they had taken an upper level evolutionary biology class, 125 (42%) felt highly confident in their knowledge about evolution, and 126 (43%) felt highly prepared to teach evolution. Most participating teachers reported relatively supportive environments: 279 (94%) reported that the school was supportive and 243 (82%) indicated that parents were supportive.

Treatment of Missing Data

We encountered multiple teachers (100 of the 296 total teachers) who had one or more missing survey responses. In total, 15% of the total response cells were missing. Excluding participants with missing data results in a loss of precision and power and can introduce non-response bias into the analyses (Sterne et al., 2009). Instead of excluding the participants with missing data, we utilized the method of multiple imputation using the Markov Chain Monte Carlo (MCMC) approach in SPSS 21.0 to impute the missing data based on a predictive model (Schafer, 1999). Multiple imputation addresses a key limitation of single imputation in that it accounts for variance between observations (Sterne et al., 2009). It carries the assumption that the data are missing at random (MAR), meaning that missingness is associated only with the variables observed in the data (Sterne et al., 2009). By using the MCMC approach to introduce an error based on variation in the variables across the imputation, also called between-imputation error, each missing observation is replaced with two or more acceptable values, providing a distribution of possibilities. To achieve this, the imputation model is applied to data where missing values are filled in a pre-defined number of times to generate a set number of complete data sets considering the variances between observations. Finally, these imputation datasets can be pooled to obtain a final dataset. In our study, linear regression was used to impute continuous responses, and multinomial logistic regression was used to impute categorical responses. These were specified as main effects models where a participant's missing response was predicted based on all of their other responses on the survey (Sterne et al., 2009). We generated 10 imputation datasets and then pooled these to obtain a final dataset.

Instrument Development and Validation

Structuring the Measure

Diverse methodologies have been utilized to measure human emotion, ranging from self-report (Izard, 1991) to observational (Yacoub et al., 2003) to physiological methods (Yuen et al., 2009). For data collection in this study, we focus

on the self-report method, and specifically the Discrete Emotions Approach (DEA) (Gruber et al., 2011; Izard, 1991). In this method, we provide the participant with a scenario, and then ask them to rate their emotional reaction to the statement on an ordinal scale with respect to specific emotion terms (Figures 1 and 2). In this study, we focus on seven emotions which Scherer (2004, 2005) labels as utilitarian—joy, anger, fear, disgust, sadness, shame, and guilt. These utilitarian emotions are particularly pertinent to evolution education in that they facilitate teachers' strategies for adaptation and success according to how they, themselves, define success (Scherer, 2004). In the case of evolution instruction, emotion dictates how teachers choose to adapt their classrooms and other learning environments to fit with their worldviews. While the DEA is not the only approach for qualifying emotion, we find it to provide a flexible and descriptive framework for data collection in that it allows us to analyze expressions of specific emotions as well as blends of different emotions which are expressed simultaneously. We later describe how we combine the DEA with the dimensional framework (Wundt, 1896) to facilitate descriptive and parsimonious quantitative measures.

Table 1. Scenarios on the E-EARS that were used to solicit emotional reactions to be reported by the teachers.

Pro-Evolution	Anti-Evolution
1: In 2008, Florida adopted new science curriculum standards that explicitly require teachers to teach the "scientific theory of evolution."	1: Prior to 2006, public school science textbooks in Georgia contained a sticker that said "this textbook contains material on evolution. Evolution is a theory, not a fact, regarding the origin of living things. This material should be approached with an open mind, studied carefully, and critically considered."
2: In 2007, Kansas approved new curriculum which removed any reference to intelligent design as a part of science. In 2013, Kansas adopted the Next Generation Science Standards which positions evolution as a fundamental principle of biology.	2: In 1999, Kansas changed their science education standards to remove any mention of "biological macroevolution, the age of the Earth, or the origin and early development of the universe." Evolutionary theory no longer appeared in state-wide standardized tests and it was left to local school districts to decide whether to teach evolution or not.
3: A kindergarten teacher reads her students "Grandmother Fish: A Child's First Book of Evolution," which explains that fish, reptiles, mammals, apes, and humans share common ancestors and all life on earth is related.	3: In 2005, Kansas drafted new standards that "require critical analysis of evolution – including scientific evidence refuting the theory," effectively stating that intelligent design should be taught. The new standards provided a definition of science that does not hinder supernatural explanations.
4: Bill Nye the Science Guy, a popularly known science educator and evolution supporter, posted a YouTube video in 2012 entitled "Creationism is Not Appropriate for Children."	4: In 1994, a school board in Louisiana required teachers to read a statement to students before evolution is taught, telling them that evolution "should be presented to inform students of the scientific concept and not intended to influence or dissuade the Biblical version of Creation or any other concept."
	5: Legislators pass a bill saying that intelligent design and evolution must have equal time in public K-12 science classrooms.
	6: Public schools are taking children to the Ark Encounter, a Christian theme park operated by Answers in Genesis, a Young Earth creationism group. Ark Encounter has models of some of the kinds of animals believed to be on the Ark, including dinosaurs.
	7: Dr. John D. Morris, President of the Institute for Creation Research, wrote in 2006 that "evolution tales are pseudo-scientific stories about an imaginary history. Evolution is best understood as an anti-God origins myth, attempting to explain man's existence without a Creator. We can do better."
	8: Mike Pence, Donald Trump's pick for Vice President, stated that for teaching evolution and creationism in schools: "I think in our schools we should teach all of the facts about all of these controversial areas, and let our students, let our children and our children's children decide based upon the facts and the science."

In order to explore in a holistic way how teachers react emotionally to issues related to evolution education, we presented the teachers with both pro-evolution (4 scenarios; Table 1) and anti-evolution (8 scenarios; Table 1) scenarios, and then asked each teacher to rate their level of joy, anger, sadness, fear, disgust, shame, and guilt on a 5-level ordinal scale: none at all (0), a little (1), a moderate amount (2), a lot (3), and a great deal (4). Table 1 contains a list of our pro- and anti-evolution scenarios, respectively. Ratings for seven emotions onto 12 respective scenarios adds up to 84 total item responses on the E-EARS. In light of Wundt's (1896) dimensional model for emotion, it is reasonable to expect that high levels of some of these emotions (like anger and disgust, for example) may be reported simultaneously; exploration of sub-constructs is necessary to partition the relatively independent elements of the teachers' reported emotional experience.

Exploration of Sub-constructs

As we utilize the DEA method as a framework for data collection and labeling of emotions, we draw upon the dimensional approach (Wundt, 1896) when considering how to analyze the data and create useful scales. The seven utilitarian emotions (joy, anger, fear, disgust, sadness, shame, and guilt) described by Scherer (2004) may, at first look, suggest a 7-dimensional instrument where each emotion is given a unique scale. However, this model would be dubious from the perspective of measurement since emotions tend to be expressed together collectively to form a unique in-the-moment experience (Scherer & Ceschi, 2000), and indeed this is a key motivation for researchers who follow Wundt's (1896) model, theorizing that emotion can be reduced to an orthogonal three-dimensional space in terms of valence (positive to negative) and arousal (excited to calm), and tension (tense to relaxed) (Figure 1). Later researchers have chosen to reduce this space further into valence and arousal since tension is difficult to separate from valence and arousal in terms of physiological expression (Lang, 1995; Russell, 2003). This dimensional theory of emotion has been used in studies where independent, relatively language-free, quantitative measures of emotion are needed (Lane et al., 1999).

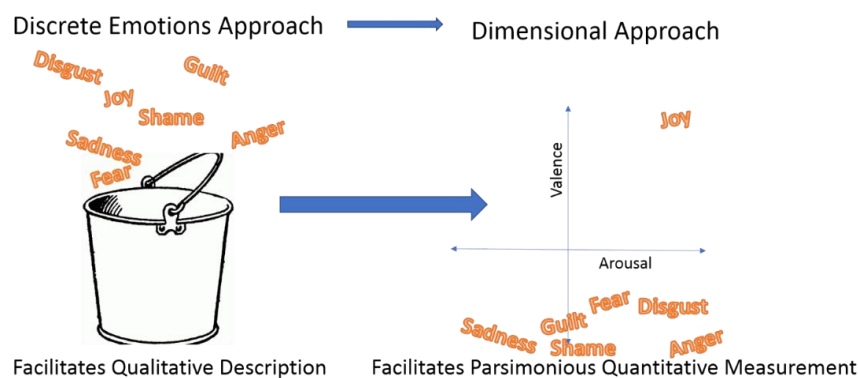


Figure 1. Quantitative description of emotion begins with identification of discrete emotions that likely coexist with each other, and then mapping them onto a lower-dimensional space.

Using the DEA approach for data collection, and in turn theorizing that the seven emotions may be expressed within a lower-dimensional space of emotional experience creates an important opportunity for understanding how teachers associate emotionally with scenarios related to evolution education. For example, in the context of this study, it is reasonable to hypothesize that certain submissive emotions like sadness, fear, shame, and guilt, or dominant emotions like anger and disgust, may be expressed together as a single experience (Bradley & Lang, 2000). Further, we may even hypothesize that joy and anger, which seem to be semantic opposites, are not true opposites of each other from a measurement perspective—while they are opposites in terms of valence, they are similar in that they both comprise high arousal (Russell, 1983). It is therefore not out of the question that an excited teacher may report high levels of both emotions while a relatively unexcited teacher may express neither emotion. It follows that understanding how these emotions cluster together in the data will yield important insight into how emotion related to evolution learning scenarios is actually expressed.

Further complicating study of the E-EARS's dimensionality is the fact that it contains 4 scenarios with pro-evolution orientation and 8 scenarios with anti-evolution orientation. It is generally assumed in survey design that negatively worded questions are true opposites of those which are positively worded, and that reverse coding the negatively worded questions will thereby put all questions on the same scale along the intended construct. We have found in previous research on evolution acceptance that this is not the case with current evolution acceptance instrumentation (Romine et al. 2017). While current acceptance measures elicit cognitive appraisal of statements about evolution, Romine et al. (2017, 2018) found that positively worded statements tend to elicit appraisal of the validity of evolution facts (or acceptance of the truth of evolution), whereas those which are negatively worded tend to elicit appraisal with regards to credibility of the theory of evolution (or rejection of incredible ideas about evolution). We have already made the case that emotions with positive (i.e. joy) and negative (i.e. anger) valence are not necessarily true opposites of each other. Further, how teachers express these emotions towards pro- and anti-evolution statements may be similarly difficult to predict and quantify without the help of data.

Factor Analysis

While the DEA operationalizes the utilitarian emotions semantically, the dimensionality construction of emotion views semantics as a descriptive approximation to what teachers are actually experiencing in the moment. This is important given that teachers' in-the-moment experience with a scenario comprises a blend of emotions, not a single discrete emotion (Scherer, 2004). We use exploratory factor analysis (EFA) in SPSS 21.0 to reconcile how emotional reactions are blended into unique single experiences which are thenceforth treated as instrumental subscales, and then use a confirmatory factor analysis model with ignorable cross-loadings to evaluate how well the factor structure reproduces

the data. EFA is a variable-centered clustering technique where latent factors among observed variables are extracted from the data (Collins & Lanza, 2010). EFA has been used extensively in science education research, particularly in the context of instrument validation (i.e. Romine et al., 2013; Corwin et al., 2015), and is a technique that is generally associated with classical test validation methods and other situations where the researcher wishes to reduce a larger feature set to relatively few latent constructs (Henson & Roberts, 2006).

EFA is exploratory in nature and is therefore a useful tool when the researcher does not have reliable prior information about the latent factor structure of an assessment and therefore intends to extract this from the data mathematically (Henson & Roberts, 2006; Osborne et al., 2008). We used the principal components method, which is useful toward finding the dimensions in a multivariate dataset with the greatest amount of variance (Jolliffe, 2011).

For diagnosis regarding the suitability of factor analysis for these data, we calculated the Kaiser-Meyer-Olkin (KMO) index of sampling adequacy (Kaiser, 1970) and performed Bartlett's test of sphericity (Bartlett, 1950). The KMO index indicates the proportion of variance in the responses that can potentially be factorized. A KMO index greater than 0.5 is indicative that factor analysis may be informative for the data (Hair et al., 1995). Along a similar line, Bartlett's test of sphericity tests the null hypothesis that the variables are unrelated. Rejecting this null hypothesis at the 0.05 alpha level indicates significant evidence in favor of the assumption that there is a latent structure in the data which can be modeled using a factor analytic procedure (Hair et al., 1995).

Number of factors was informed by parallel analysis (Horn, 1965) using SPSS 21. The goal of parallel analysis is to compare the eigenvalues generated from the raw variables to those generated from random ordered data (Williams et al., 2010). To perform parallel analysis, we generated a series of random datasets which were permutations of the original dataset so that the distributional properties of the original variables were preserved. This method is recommended for ordinal variables which tend to take non-normal distributions (O'Connor, 2000) and is robust to deviations from normality (Buja & Eyuboglu, 1992). This was accomplished using an SPSS macro developed by O'Connor (2000), in which we preserved the 99th percentile eigenvalue extracted from 1000 random permutations of the original data. If the random data eigenvalue is greater than the raw data eigenvalue, then we can conclude that the eigenvalue could be generated by chance and is therefore indicative of a non-important factor.

After factor extraction, we utilized the promax rotation method (Hendrickson & White, 1964) to diagonalize the factor solution allowing the factors to be correlated and evaluated the simple structure and interpretability of the solution. This is an important step given the tendency of parallel analysis to overestimate the number of factors when there are outliers in the data (Zwick & Velicer, 1986). As an additional step, in order to evaluate fit of the extracted factor structure to the data, we utilized a CFA approach hypothesizing that an individual response is a measure of a single latent variable with no cross-loading onto other latent variables. Estimation of the model was done using the diagonally weighted least squares (WLSMV) estimator in Mplus 7 (Muthén & Muthén, 2012). The probit link was used to account for the categorical nature of the responses. Fit with the data was evaluated using the Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and the Tucker-Lewis Index (TLI). An RMSEA close to 0.06 (Hu & Bentler, 1999), but not above 0.07 (Steiger, 2007) is indicative of acceptable fit. Acceptable fit is also indicated by a CFI and TLI above 0.9 (Bentler, 1990; Bentler & Bonett, 1980).

Rasch Analysis

After making a reflective and informed decision (Henson & Roberts, 2006) on the number of factors to retain, and which emotions loaded onto which factors, our attention turned to construct validity of the individual items towards measurement of each factor, which we call an *emotional sub-construct* from here on. Given that a teacher's emotional reaction to each scenario was rated on an ordinal five-level scale, and that there were seven emotions per scenario which may have unique scale structures, we utilized a constrained Rasch partial credit model (Masters, 1982), where responses associated with each emotion had a common rating scale, as a criterion for validity. Rasch frameworks differ in philosophy from other measurement frameworks like IRT and classical test theory (CTT) in that Rasch provides a philosophical criterion for validity of the data: the likelihood of a teacher selecting a particular response level is proportional only to the difference between the teacher's level of emotion and the difficulty threshold of that response (Wright & Stone, 1979). Other factors influencing this probability are considered detrimental to measurement, and are detected through misfit with the model (Bond & Fox, 2013). This practice of measuring the quality of the data through a standard model is different from the statistical practice of fitting the model to the data. We use the Rasch framework due to several epistemological advantages including alignment with the laboratory practice of calibrating machines based on a fixed standard before their use, better alignment with the scientific practice of falsification (Popper, 1968; Lakatos, 1976) than statistical model-fitting approaches, and generation of invariant linear estimates (Boone & Scantlebury, 2006).

We used the WINSTEPS software package (Linacre, 2006) to fit the unidimensional constrained Rasch partial credit model to the data within each subscale, which provided a standard for the validity of the items and subscales. Item validity was evaluated by their mean squares fit with the model. We used mean squares infit and outfit of the item response patterns with respect to what the model would predict based on the item's difficulty. These measures of fit have expected

values of 1, but Wright et al. (1994) suggest that values between 0.5 and 1.5 are indicative of potentially useful items. Values below 0.5 indicate that the item fits the model better than we would expect in real data, possibly indicating bias in the item that favors teachers with high intensity emotion (Masters, 1988). However, with regards to measurement validity, misfitting items (mean squares fit > 1.5) represent a significant validity concern as they indicate bias in favor of teachers with weak levels of emotional intensity which contradicts the intended scale directionality (Bond & Fox, 2013).

In addition to item construct validity, we were also interested in the usefulness of the scale itself in explaining and quantifying teachers' emotional experiences with respect to the scenarios presented. Of primary concern was unidimensionality, the efficacy of the Rasch model in capturing the systematic variance in the item responses within each subscale. As with item fit, we evaluated unidimensionality using a falsificationist perspective. We first invoked the assumption that the items are unidimensional, and then attempted to falsify this by inspecting the residuals in the item responses with respect to the Rasch model. If the items measure a single dimension, then the residuals should be random. We implemented principal components analysis (PCA) on the residuals, and inspected the first eigenvalue. If the residuals are random, this eigenvalue should fall below 2 (Linacre & Tennant, 2009). A first eigenvalue above 2 indicates some systematic departure from unidimensionality in the set of items with respect to the intended emotional sub-construct.

A related assumption of interest in this study is that of local independence of items. If items are locally independent, then they are unrelated to each other after accounting for variance due to the latent emotional sub-construct (Collins & Lanza, 2010). Local dependency is therefore quantified through correlation of item residuals with respect to the Rasch model; the residuals of locally independent items have zero correlation. Although the assumption of local independence is often taken for granted in Rasch studies, this is a potential issue in this assessment since scenarios may introduce dependency between the measures that is extraneous with respect to the emotional sub-construct.

Predicting Intensity of Emotional Experience

Upon extraction of Rasch measures for emotional factors, we were interested in which variables best predict these factors. We measured a variety of factors including religion which we coded as monotheistic [coded 1] or otherwise [coded 0], grade level taught, teaching practice (whether or not evolution is taught [coded 0 or 1]), how much time is spent teaching evolution [ordinal 0-4], and whether the teacher feels responsible for teaching evolution [0 or 1]) preparation to teach evolution (the extent to which the teacher feels prepared to teach evolution [ordinal 0-4]), whether he/she has taken an evolution class [coded 0 or 1], confidence in knowledge about evolution [ordinal 0-4]), external support for teaching evolution (parental [coded 0 or 1] and school support [coded 0 or 1]), and acceptance of evolution as it is traditionally measured (I-SEA measures for acceptance of micro-, macro-, and human evolution [continuous 1-5 scale], MATE measures of acceptance of facts about evolution and credibility of evolutionary theory [continuous 1-5 scale], and an explicit question asking whether or not the teacher accepts evolution as being true and valid [coded 0 or 1]). These variables were used to build multiple linear regression models using the forward stepwise algorithm in SPSS 21.0. In the forward component of this algorithm, the addition of each predictor is tried and the improvement in fit over the intercept-only model is assessed. We used a 0.01 alpha level for entry, meaning that if no variable improves fit at the 0.01 alpha level, then the algorithm stops. Otherwise, the variable that best improves the fit of the model is added. This is done sequentially until no variables improve fit of the model at the 0.01 alpha level. A backwards component operates between each forward component, and involves an attempt to remove each variable from the model in a leave-one-out process. If fit of the model did not reduce at the 0.05 alpha level upon removal of a variable, then it was removed from the set of variables included in the model before the next forward iteration continued. Since collinearity exists between these many of these variables, the purpose of this backward step was to reduce the effect of the order in which variables were entered on the resulting final model.

Results

RQ1: How do teachers' discrete emotions cluster together to form unique emotional experiences?

The KMO index for the collection of item responses in the factor analysis was measured at 0.92 which is well above the recommended threshold of 0.5 (Hair et al., 1995). Bartlett's test of sphericity was also significant ($\chi^2 = 25213$, $df = 3486$, $p < 0.001$), supporting the claim that there is a latent structure in the data that can be modeled using factor analysis.

From the parallel analysis, four factors emerged which exceeded those derived from random permutations of the data. The first factor had an eigenvalue of 25.1, which was much greater than the random data eigenvalue of 2.4. The second factor had an eigenvalue of 16.9, which was greater than the parallel random eigenvalue of 2.3. The third factor had an eigenvalue of 4.3, which was greater than the random eigenvalue of 2.2. The fourth factor had an eigenvalue of 2.2, which was only slightly above the randomly generated value of 2.1. The eigenvalue of 1.8 for the fifth factor was exceeded by the eigenvalue of 2.0 generated for the random data, indicating that including greater than 4 factors may not yield replicable conclusions. In consideration of the closeness between the raw data and random data eigenvalues for the fourth factor and the tendency of parallel analysis to extract minor components from the data (Zwick & Velicer, 1986), we inspected the pattern matrices for 3-factor and 4-factor solutions. We decided that the 3-factor solution's structure yielded a clearer and more parsimonious interpretation.

The pattern matrix (Table 2) for the solution with 3 emotional sub-constructs showed simple structure and a definitive loading pattern that was consistent across scenarios. Specifically, teachers scoring high on emotional sub-construct 1 tended to express Joy on the pro-evolution scenarios and Anger, Sadness, Fear, and Disgust on the anti-evolution scenarios. We labeled this emotional sub-construct *Pro-evolution experience*, which we will call "pro-evolution" throughout this report. Teachers scoring high on emotional sub-construct 2 tended to express the negative emotions of Anger, Sadness, Fear, Disgust, Shame, and Guilt on the pro-evolution scenarios and Joy on the anti-evolution scenarios. We therefore labeled this emotional sub-construct *Anti-evolution experience*, which we will call "anti-evolution" in this report. Teachers who scored high on emotional sub-construct 3 tended to express Shame and Guilt on the anti-evolution scenarios. We labeled this emotional sub-construct *Regret over anti-evolution scenarios*, which we will call "regret" in this report.

Table 2. Pattern matrix for the 3-factor EFA solution. Factor 1 comprises *pro-evolution experience*; Factor 2 comprises *anti-evolution experience*; and Factor 3 comprises *regret over anti-evolution scenarios*. Loadings less than 0.2 are omitted.

	Factor1	Factor2	Factor3
Pro1_Joy	0.627		
Pro1_Anger		0.881	
Pro1_Sadness		0.867	
Pro1_Fear		0.545	
Pro1_Disgust		0.769	
Pro1_Shame		0.793	
Pro1_Guilt		0.408	
Anti1_Joy	-0.284	0.417	
Anti1_Anger	0.791		
Anti1_Sadness	0.696		
Anti1_Fear	0.672		0.229
Anti1_Disgust	0.779		
Anti1_Shame			0.666
Anti1_Guilt			0.858
Anti2_Joy		0.416	
Anti2_Anger	0.713		
Anti2_Sadness	0.546	-0.31	
Anti2_Fear	0.663		
Anti2_Disgust	0.753		
Anti2_Shame			0.737
Anti2_Guilt		-0.2	0.912
Anti3_Joy	-0.243	0.411	
Anti3_Anger	0.808		
Anti3_Sadness	0.753		
Anti3_Fear	0.708		
Anti3_Disgust	0.815		
Anti3_Shame			0.747
Anti3_Guilt		0.25	0.788

Table 2. Continued

	Factor1	Factor2	Factor3
Pro2_Joy	0.478	-0.368	
Pro2_Anger		0.867	
Pro2_Sadness		0.807	
Pro2_Fear		0.87	
Pro2_Disgust		0.779	
Pro2_Shame		0.589	
Pro2_Guilt		0.756	
Anti4_Joy		0.42	
Anti4_Anger	0.775	0.22	
Anti4_Sadness	0.676		
Anti4_Fear	0.596		0.242
Anti4_Disgust	0.756	0.236	
Anti4_Shame			0.658
Anti4_Guilt			0.729
Anti5_Joy	-0.226	0.369	
Anti5_Anger	0.69		
Anti5_Sadness	0.709		
Anti5_Fear	0.691		
Anti5_Disgust	0.76		
Anti5_Shame			0.653
Anti5_Guilt			0.772
Pro3_Joy	0.628		
Pro3_Anger		0.762	
Pro3_Sadness		0.831	
Pro3_Fear		0.74	
Pro3_Disgust		0.887	
Pro3_Shame		0.885	
Pro3_Guilt		0.817	
Anti6_Joy		0.744	
Anti6_Anger	0.754		
Anti6_Sadness	0.718		
Anti6_Fear	0.738	0.202	
Anti6_Disgust	0.81		
Anti6_Shame			0.658
Anti6_Guilt			0.874
Pro4_Joy	0.767		
Pro4_Anger		0.616	
Pro4_Sadness	-0.302	0.361	
Pro4_Fear		0.594	
Pro4_Disgust		0.793	
Pro4_Shame		0.64	
Pro4_Guilt		0.515	
Anti7_Joy		0.76	
Anti7_Anger	0.721		
Anti7_Sadness	0.553		
Anti7_Fear	0.504		0.213
Anti7_Disgust	0.737		
Anti7_Shame			0.764
Anti7_Guilt			0.748
Anti8_Joy	-0.441	0.297	
Anti8_Anger	0.81		
Anti8_Sadness	0.776		
Anti8_Fear	0.737		
Anti8_Disgust	0.873	0.24	
Anti8_Shame	0.318		0.568
Anti8_Guilt			0.762

CFA suggested that 3 factors described the set of responses adequately. A 3-factor CFA model with ignorable cross-loading of responses between factors yielded an RMSEA of 0.061, a CFI of 0.90, and a TLI of 0.90. These are indicative of adequate fit, meaning that the 3-factor model with no cross-loading replicates the relationships between responses in the data acceptably.

RQ2: What is the efficacy of the scenarios in allowing us to draw valid inferences about these distinct emotional experiences?

We fit a unidimensional Rasch model to each of the respective emotional sub-constructs: (1) pro-evolution; (2) anti-evolution; and (3) regret. The scales showed adequate reliability for measurement of teachers' (0.95, 0.77, and 0.85, respectively) and items' (0.97, 0.92, and 0.96, respectively) positioning on each subscale. This said, we found some departure from unidimensionality in each of the subscales (1st eigenvalues = 3.77, 3.32, and 2.56, respectively). The largest item residual correlations for each subscale were 0.67, 0.58, and 0.29, respectively, indicating that while there is departure from unidimensionality, local dependency is not responsible for this, and does not present a significant measurement concern (Linacre, 2006). Positive loadings onto the first residual factor for the *pro-evolution* construct indicate that the multidimensionality is primarily caused by expression of fear on the anti-evolution scenarios (scenarios 1, 3, 4, and 5-8 have loadings greater than 0.3). It is interesting that fear around the anti-evolution scenarios constitutes higher positioning along the Rasch scale (Kruskal-Wallis test statistic = 15.6, df = 4, p = 0.004) than the other emotions aligning with pro-evolution experience (Figure 2), indicating that teachers with pro-evolution experience would have the tendency to express the other emotions along this subscale before expressing fear—only the teachers with the highest levels of reported pro-evolution experience expressed fear with respect to the anti-evolution scenarios. Anger and disgust on anti-evolution scenario 4 also loaded onto this residual dimension (loadings of 0.40 and 0.45, respectively). Expression of negative emotions on pro-evolution scenarios 1-3 were responsible for the multidimensionality in measures for the *anti-evolution* construct: anger on scenarios 1 and 2 (loadings of 0.44 and 0.51), sadness and shame on scenarios 1-3 (loadings between 0.32-0.39), guilt on scenario 3 (loading of 0.42), and fear on scenario 1 (loading of 0.55). This suggests some inconsistency between the pro-evolution and anti-evolution scenarios in measuring anti-evolution experience, but there were no significant differences between the different emotions with regards to positioning (Figure 3) along the Rasch scale (Kruskal-Wallis test statistic = 8.7, df = 6, p = 0.192). The small departure from unidimensionality in the *regret* construct is due to slight inconsistency between the emotions of shame and guilt. The responses of guilt on anti-evolution scenarios 1-6 have positive loadings (0.37-0.63) onto the residual factor, suggesting some commonality in these scenarios which elicits a guilt response in a way that is peripheral to the main factor of regret. Like the emotion of fear in the pro-evolution construct, the emotion of guilt sits higher on the Rasch scale (Figure 4) than shame (Kruskal-Wallis test statistic = 10.0, df = 1, p = 0.002), implying that only teachers with the highest levels of regret over the anti-evolution scenarios will feel the deeply personal emotion of guilt.

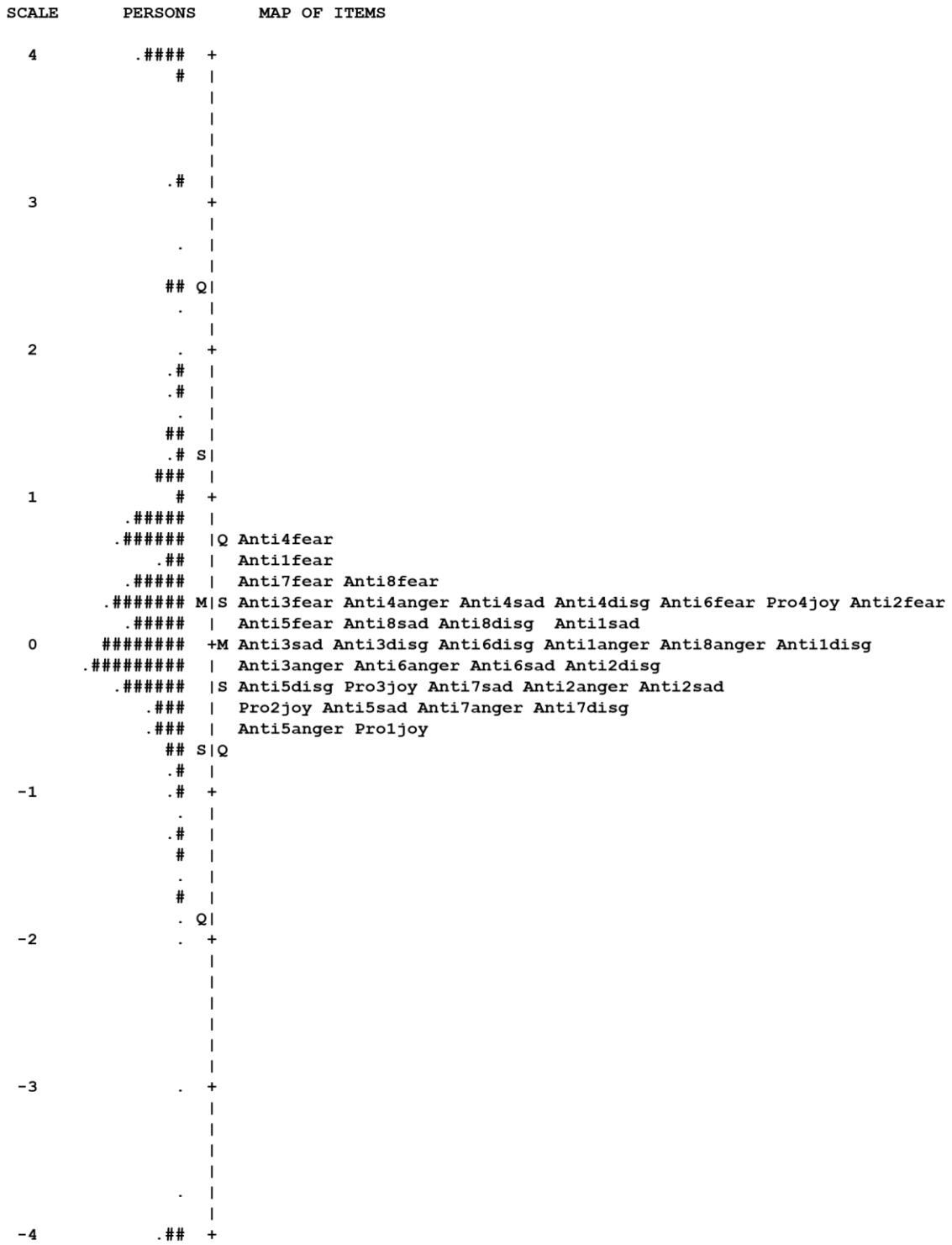


Figure 2: Person-item map for pro-evolution experience. Fear is significantly higher on the scale than the other emotions (Kruskal-Wallis test statistic = 15.6, $df = 4$, $p = 0.004$).

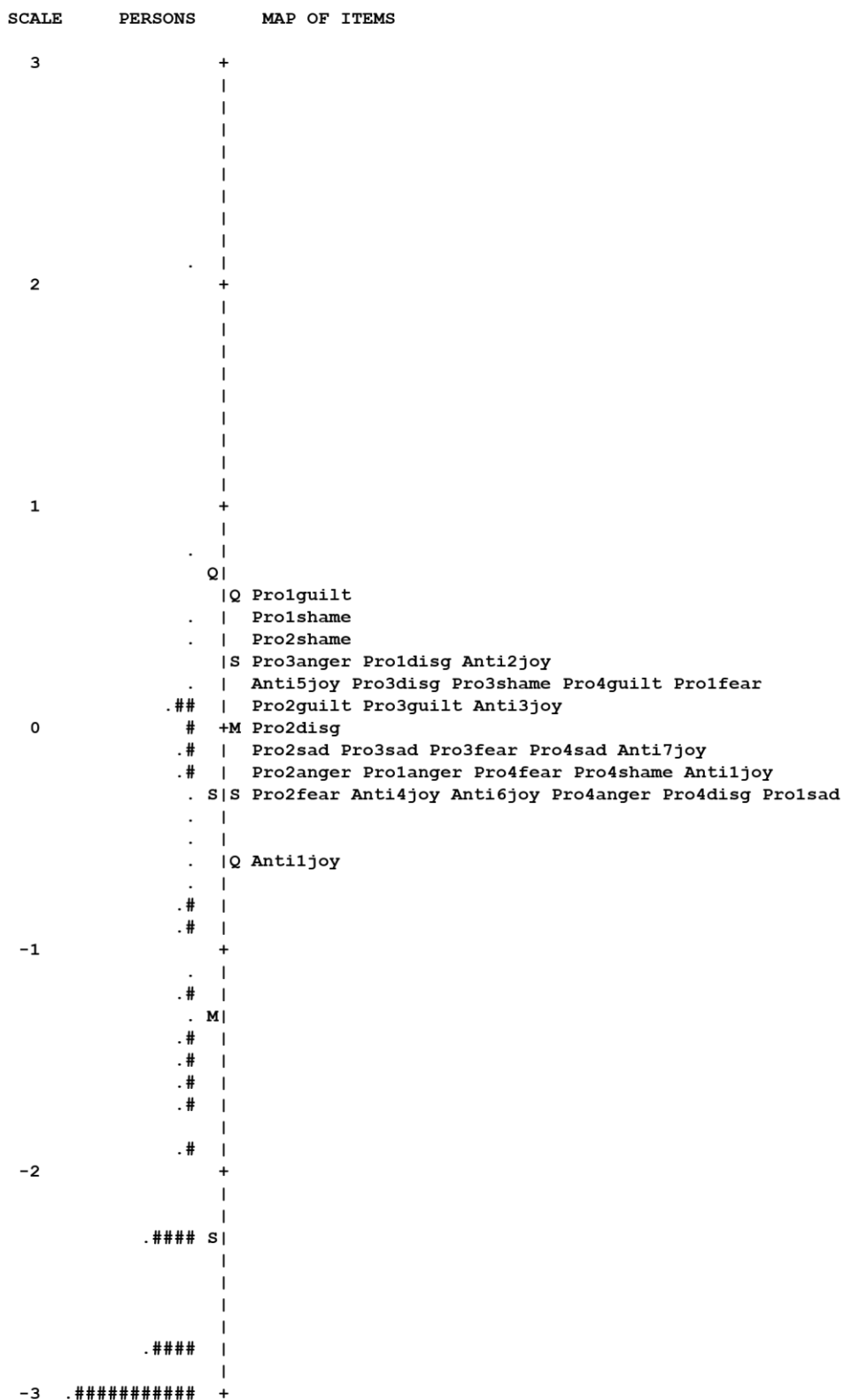


Figure 3: Person-item map for anti-evolution experience. No significant differences existed between the scale locations of emotions (Kruskal-Wallis test statistic = 8.7, df = 6, p = 0.192).

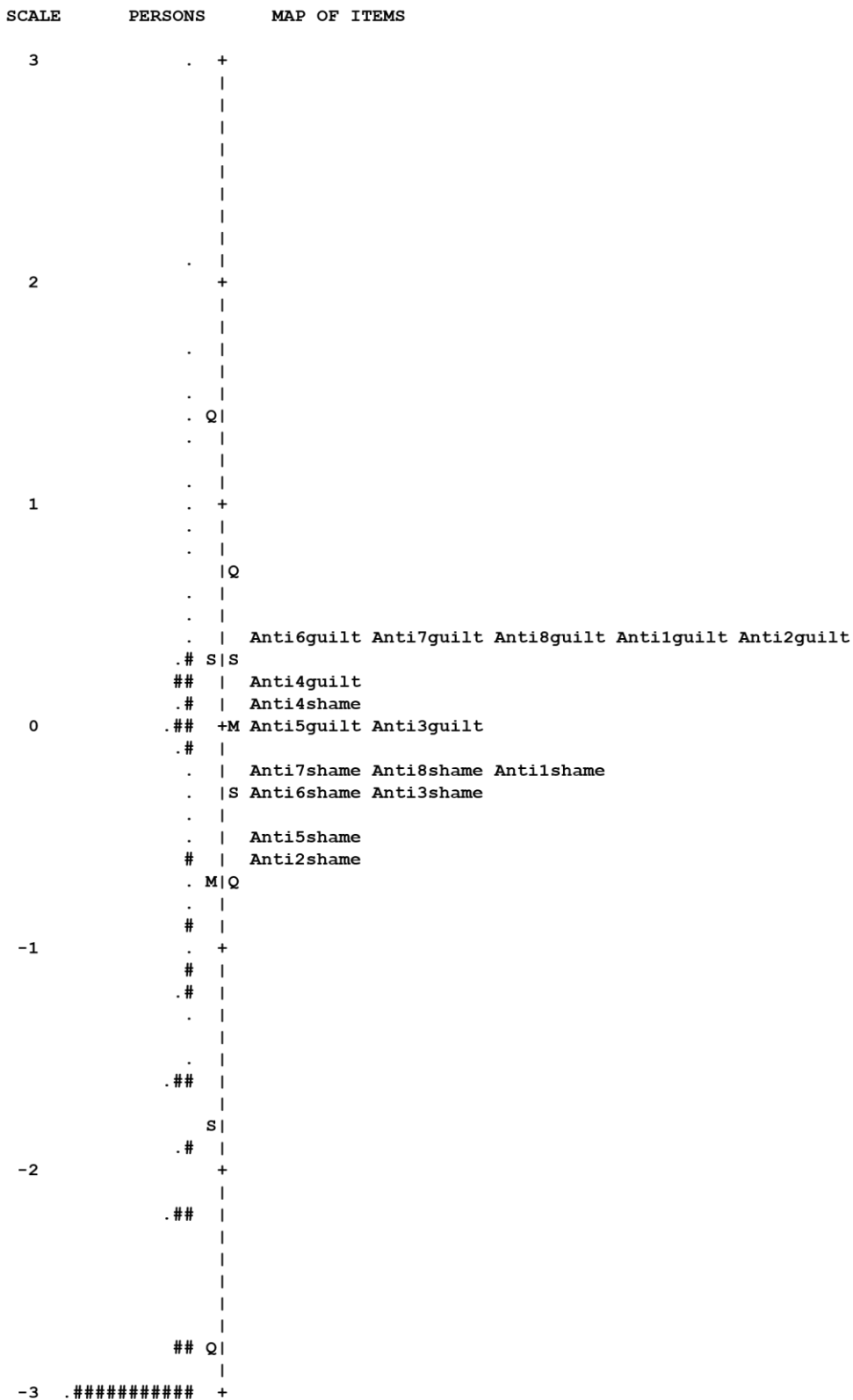


Figure 4: Person-item map for regret over anti-evolution scenarios. Guilt is significantly higher on the scale than shame (Kruskal-Wallis test statistic = 10.0, $df = 1$, $p = 0.002$).

Finally, 4 responses in the *pro-evolution* construct (Table 3), 10 responses on the *anti-evolution* construct (Table 4), and 2 responses on the *regret* construct (Table 5) showed mean squares fit indices above 1.5. However, these response patterns also had positive point-biserial correlations (above 0.3). With respect to polytomous data, Smith (1996) describes these types of patterns as *informative-noisy*, meaning that despite some unpredictability in how the teachers respond, the responses may nonetheless provide useful measurement information.

Table 3. Table of Rasch difficulty and item fit for pro-evolution emotional experience.

Item	Measure	SE	Infit	Outfit	PtBis
Pro1Joy ^a	-0.55	0.06	1.02	1.93	0.56
Anti1Anger	0.02	0.05	0.83	0.95	0.70
Anti1Sadness	0.15	0.06	0.91	0.93	0.68
Anti1Fear	0.59	0.06	1.09	1.04	0.62
Anti1Disgust	-0.01	0.05	1.06	1.10	0.65
Anti2Anger	-0.30	0.06	0.93	0.86	0.69
Anti2Sadness ^a	-0.23	0.06	1.34	1.25	0.59
Anti2Fear	0.32	0.05	0.91	1.07	0.69
Anti2Disgust	-0.21	0.05	0.87	0.78	0.71
Anti3Anger	-0.08	0.05	0.76	0.75	0.74
Anti3Sadness	0.01	0.06	0.83	0.82	0.71
Anti3Fear	0.29	0.05	0.97	0.99	0.68
Anti3Disgust	0.00	0.05	0.80	0.84	0.74
Pro2Joy ^a	-0.50	0.06	1.32	1.54	0.55
Anti4Anger	0.23	0.05	1.04	0.94	0.64
Anti4Sadness	0.34	0.06	0.92	0.93	0.67
Anti4Fear	0.78	0.06	1.23	1.28	0.57
Anti4Disgust	0.26	0.05	1.14	1.00	0.62
Anti5Anger	-0.52	0.06	1.28	0.99	0.64
Anti5Sadness	-0.38	0.06	1.04	1.14	0.65
Anti5Fear	0.09	0.05	1.04	0.97	0.67
Anti5Disgust	-0.31	0.06	1.09	0.92	0.67
Pro3Joy	-0.21	0.06	1.12	1.27	0.62
Anti6Anger	-0.10	0.05	0.97	0.97	0.69
Anti6Sadness	-0.18	0.06	0.98	0.95	0.67
Anti6Fear	0.27	0.05	1.00	0.93	0.67
Anti6Disgust	-0.06	0.05	0.94	0.87	0.70
Pro4Joy	0.30	0.05	1.08	1.08	0.66
Anti7Anger	-0.46	0.06	1.05	0.87	0.65
Anti7Sadness ^a	-0.29	0.06	1.12	1.91	0.56
Anti7Fear ^a	0.43	0.06	1.33	1.42	0.57
Anti7Disgust	-0.39	0.06	0.81	0.91	0.68
Anti8Anger ^a	0.07	0.05	1.01	1.61	0.65
Anti8Sadness	0.16	0.06	1.14	1.05	0.65
Anti8Fear	0.39	0.06	0.91	0.87	0.70
Anti8Disgust	0.09	0.05	0.94	0.91	0.69

^aInformative-noisy pattern (Smith, 1996)

Table 4. Table of Rasch difficulty and item fit for anti-evolution emotional experience.

Item	Measure	SE	Infit	Outfit	PtBis
1=Pro1Anger	-0.16	0.07	0.84	0.69	0.77
2=Pro1Sadness	-0.27	0.07	0.80	0.62	0.80
3=Pro1Fear ^a	0.24	0.08	0.99	1.58	0.50
4=Pro1Disgust	0.33	0.08	0.93	0.60	0.67
5=Pro1Shame	0.54	0.10	0.40	0.39	0.72
6=Pro1Guilt ^a	0.65	0.11	1.37	2.42	0.36
7=Anti1Joy ^a	-0.18	0.07	1.46	2.28	0.50
8=Anti2Joy ^a	0.30	0.08	1.19	1.60	0.46
9=Anti3Joy ^a	0.06	0.07	1.51	1.30	0.47
10=Pro2Anger	-0.19	0.07	0.75	0.45	0.78
11=Pro2Sadness	-0.08	0.07	0.88	0.63	0.71
12=Pro2Fear	-0.34	0.07	0.88	0.66	0.76
13=Pro2Disgust	0.02	0.07	1.06	0.96	0.65
14=Pro2Shame ^a	0.41	0.09	1.55	1.01	0.48
15=Pro2Guilt	0.07	0.07	0.79	0.88	0.68
16=Anti4Joy ^a	-0.26	0.07	1.60	2.25	0.48
17=Anti5Joy ^a	0.19	0.08	1.42	1.61	0.43
18=Pro3Anger	0.34	0.09	0.57	0.44	0.71
19=Pro3Sadness	-0.14	0.07	0.95	0.60	0.75
20=Pro3Fear	-0.12	0.07	0.93	1.24	0.64
21=Pro3Disgust	0.20	0.08	0.48	0.35	0.81
22=Pro3Shame	0.22	0.08	0.48	0.38	0.78
23=Pro3Guilt	0.09	0.07	0.80	0.46	0.71
24=Anti6Joy	-0.33	0.07	0.79	0.84	0.80
25=Pro4Anger	-0.26	0.07	1.09	0.94	0.66
26=Pro4Sadness ^a	-0.14	0.07	1.21	1.56	0.45
27=Pro4Fear	-0.23	0.07	1.34	1.23	0.58
28=Pro4Disgust	-0.25	0.07	0.59	0.41	0.84
29=Pro4Shame	-0.15	0.07	1.07	0.83	0.65
30=Pro4Guilt	0.15	0.08	1.26	0.79	0.54
31=Anti7Joy	-0.07	0.07	0.85	0.50	0.77
32=Anti8Joy ^a	-0.63	0.07	1.75	2.24	0.46

^aInformative-noisy pattern (Smith, 1996)

Table 5. Table of Rasch difficulty and item fit for regret related to anti-evolution scenarios.

Item	Measure	SE	Infit	Outfit	PtBis
Anti1Shame	-0.22	0.07	1.08	1.06	0.67
Anti1Guilt	0.42	0.07	0.53	0.45	0.75
Anti2Shame	-0.58	0.07	0.83	0.75	0.75
Anti2Guilt	0.43	0.07	0.70	0.94	0.70
Anti3Shame	-0.28	0.07	0.88	0.91	0.70
Anti3Guilt	-0.01	0.06	0.88	0.79	0.74
Anti4Shame	0.11	0.07	1.25	1.11	0.60
Anti4Guilt	0.24	0.06	0.91	0.72	0.68
Anti5Shame ^a	-0.53	0.07	1.34	1.39	0.58
Anti5Guilt ^a	-0.03	0.06	1.31	1.84	0.62
Anti6Shame	-0.35	0.07	1.15	1.36	0.59
Anti6Guilt	0.43	0.07	0.62	0.45	0.74
Anti7Shame ^a	-0.16	0.07	1.17	1.48	0.61
Anti7Guilt ^a	0.37	0.07	1.32	1.38	0.59
Anti8Shame ^a	-0.24	0.07	1.18	1.33	0.60
Anti8Guilt ^a	0.42	0.07	0.76	1.52	0.66

^aInformative-noisy pattern (Smith, 1996)

Given the uniqueness of the contribution of each of these scenarios to the construct, we would caution against eliminating any of the scenarios without further research. However, looking more closely at these items can assist us in understanding how the respective scenarios function for measurement. Of the misfitting responses loading onto the *pro-evolution* factor, two comprised joy on Pro-evolution Scenarios 1 and 2, one comprised reported sadness on Anti-

Evolution Scenario 7, and one comprised reported anger on Anti-Evolution Scenario 8. It is interesting that the misfit of these items is indicated by the outfit index, which tends to be outlier-sensitive. The infit indices for these responses are close to the expected value of 1, with the exception of joy on Pro-evolution Scenario 2 (infit = 1.32). This discrepancy between infit and outfit indices indicates that outlying responses are present which attenuate expected fit with the Rasch model.

We see a more uniform match between infit and outfit indices in responses within the *anti-evolution* factor. These misfitting responses comprise expression of the emotion of joy on anti-evolution scenarios (1, 2, 3, 4, 5, and 8), indicating some inconsistency between the emotion of joy on anti-evolution scenarios and the negative emotions on the pro-evolution scenarios in defining the *anti-evolution* dispositional trait. This said, anti-evolution scenarios 6 and 7 did not show this pattern, warranting discussion of what anti-evolution scenarios 1-5 and 8 have in common, and how these differ from anti-evolution scenarios 6 and 7. The other misfitting responses entailed passive emotions in regard to the pro-evolution scenarios: fear and guilt on pro-evolution scenario 1, shame on pro-evolution scenario 2, and sadness on pro-evolution scenario 4. These misfits lacked the conceptual uniformity that we saw on the emotion of joy within this factor, meaning it may be more difficult to discern why these misfits occurred at the scenario level. Within the *regret* factor, the two misfitting responses comprised expression of guilt on anti-evolution scenarios 5 and 8. In scenario 8, we again see a large imbalance between the outfit index (1.52) and the infit index (0.76), suggesting that outlying responses from participants that do not fit the Rasch model may be responsible for this misfit.

The Wright maps (Figures 2-4) indicate that, of the three constructs, *pro-evolution experience* may be the most useful for measuring the emotional response of science teachers to the scenarios. Not only does this construct deliver the most reliable measure, but we also observed that the distribution of item locations matches well with the distribution of teacher measures. An item provides the most information about participants with ability matching the difficulty of that item (De Ayala, 2013), and so it makes sense that this construct may be the most informative for science teachers. Figures 3 and 4 show that the distributions of teacher measures for *anti-evolution experience* and *regret over anti-evolution scenarios* are similar in that a majority of the teachers had measures toward the bottom of the scale, meaning they tended to express both low levels of anti-evolution experience and low levels of regret over anti-evolution scenarios. That the items are positioned well above the measures for a majority of the participants indicates that they collectively solicit higher levels of anti-evolution experience and regret than a majority of the teachers hold. Scenarios which make it easier to express anti-evolution experience or regret would provide more useful information for these teachers. Pro- and anti-evolution experience measures have a moderate negative correlation ($r = -0.52$). A negative correlation is expected since they will tend to solicit different valence, but are not simply opposites along the same construct. However, regret has a positive correlation with both pro-evolution ($r = 0.40$) and anti-evolution experience ($r = 0.20$), meaning that there was some tendency for teachers with both high pro-evolution and anti-evolution experience to express shame and guilt, albeit this tendency was much higher for those teachers with pro-evolution experience.

RQ3: What are the most important factors in predicting teachers' emotional dispositions?

Extraction of three emotional sub-constructs begets three respective regression models for: (1) pro-evolution experience; (2) anti-evolution experience, and (3) shame and guilt around anti-evolution scenarios (Table 6). The final model for pro-evolution experience ($F_{2,293} = 136.6$, $p < 0.001$, $r^2_{adj} = 0.48$) contained two significant predictors: (1) acceptance of facts about evolution ($B = 1.39$, $SE_B = 0.099$, $t = 14.0$, $p < 0.001$, $r_{zero-order} = 0.67$, $r_{partial} = 0.63$) and (2) subscription to monotheistic faith ($B = -0.63$, $SE_B = 0.139$, $t = -4.53$, $p < 0.001$, $r_{zero-order} = -0.37$, $r_{partial} = -0.26$). The final model for anti-evolution experience ($F_{3,292} = 62.7$, $p < 0.001$, $r^2_{adj} = 0.39$) contained three predictors: (1) acceptance of microevolution ($B = -0.96$, $SE_B = 0.146$, $t = -6.57$, $p < 0.001$, $r_{zero-order} = -0.52$, $r_{partial} = -0.36$), (2) statement that one accepts evolution as true and valid ($B = -1.16$, $SE_B = 0.232$, $t = -5.01$, $p < 0.001$, $r_{zero-order} = -0.46$, $r_{partial} = -0.28$), and (3) preparation to teach evolution ($B = -0.28$, $SE_B = 0.059$, $t = -4.78$, $p < 0.001$, $r_{zero-order} = -0.42$, $r_{partial} = -0.27$). The final model for shame and guilt around anti-evolution scenarios ($F_{2,293} = 10.7$, $p < 0.001$, $r^2_{adj} = 0.062$) contained two significant predictors: (1) acceptance of facts about evolution ($B = 0.53$, $SE_B = 0.135$, $t = 3.93$, $p < 0.001$, $r_{zero-order} = 0.17$, $r_{partial} = 0.22$) and (2) preparation to teach evolution ($B = -0.30$, $SE_B = 0.085$, $t = -3.54$, $p < 0.001$, $r_{zero-order} = -0.14$, $r_{partial} = -0.20$).

Table 6. Significant predictors of emotional experience for pro-evolution, anti-evolution, and regret related to anti-evolution scenarios.

Construct	Predictor	B	SE _B	T ^a	r _{zero-order}	r _{partial}
Pro-Evolution	Intercept	-5.66	0.489	-11.58		
	Facts	1.39	0.099	14.03	0.67	0.63
	Monotheistic	-0.63	0.139	-4.53	-0.37	-0.26
Anti-Evolution	Intercept	4.52	0.615	7.35		
	Preparation	-0.28	0.059	-4.78	-0.42	-0.27
	Accept	-1.16	0.232	-5.01	-0.46	-0.28
	Micro	-0.96	0.146	-6.57	-0.52	-0.36
Regret	Intercept	-2.95	0.596	-4.94		
	Facts	0.53	0.135	3.93	0.17	0.22
	Preparation	-0.30	0.085	-3.54	-0.14	-0.20

^aSignificant at 0.01 alpha level

Discussion

The measurement framework behind the E-EARS provided a lens into understanding how teachers' emotions tend to blend together to form unique experiences associated with the specific scenarios. The utilitarian emotions we measured (joy, anger, sadness, fear, disgust, shame, and guilt) are conceptually distinct feelings, and so it might be expected that this would result in a 7-dimensional instrument. The 3-dimensional factor structure in the responses, however, suggests that the dimensional model (Hawley & Sinatra, 2019; Wundt, 1896) served as a more parsimonious and representative framework for how science teachers actually express their feelings. We emphasize here that this finding should not be taken to suggest that consideration of individual discrete emotions is not useful; without defining these at the outset, each respective factor would lose qualitative meaning. What it suggests is that certain emotions tend to be expressed together as a distinct experience, which is similar to what Hawley and Sinatra (2019) found in their study with Christian science teachers. The discrete emotions can be used to describe these experiences in language. We describe these in turn.

Pro-Evolution Experience

The strongest factor in the data constituted pro-evolution experience, or the tendency to express emotional states that would be expected from a teacher who supports the teaching of evolution in the classroom. The factor analysis indicated that this constituted expression of joy on the scenarios which described events favorable to teaching evolution, and anger, sadness, fear, and disgust on those scenarios which described unfavorable events. That the factor loading patterns were consistent across all scenarios, and that the scale displayed excellent reliability ($\rho_{\text{Rasch}} = 0.95$), indicates consistency in the way that the multiple scenarios solicit these emotions. However, the Wright map (Figure 2) helps us delve deeper toward prediction of the emotional experiences that particular teachers are likely to have.

Figure 2 indicates that the distribution of the emotions across the scale is relatively uniform with the exception of fear, which tends to be expressed only by those teachers with the highest levels of pro-evolution emotional experience. This suggests that joy on the positive scenarios and sadness, anger, and disgust on the negative scenarios tend to be expressed together regardless of a person's level of pro-evolution experience. Only those teachers with the highest levels of pro-evolution experience actually feel fearful of the resistance to evolution education (i.e. Griffith & Brem, 2004; Siani & Yarden, 2020).

Anti-Evolution Experience

Anti-evolution experience comprised the tendency to express joy after reading the scenarios adverse to teaching evolution and anger, sadness, fear, disgust, guilt, and shame after reading the scenarios describing events that were favorable to teaching evolution. Teachers' measures were skewed toward the lower end of the scale which is not surprising given our focus on science teachers. Unlike Hawley and Sinatra (2019) which focused on Christian educators, our study had 44% representation of Christian faith. Given the tendency for perceived conflict between religion and evolution to attenuate favorability toward the theory of evolution (Barnes et al., 2021), it is likely that this sample will be more pro-evolution than studies with samples that are more uniformly Christian. The respective emotions themselves were uniformly distributed across the Wright map (Figure 3), indicating that a teacher's level of anti-evolution experience did not dictate a tendency to express certain types of emotions over others.

Positive loading of the negative emotions on the pro-evolution scenarios was consistent across all scenarios within this factor. However, some inconsistency in the expression of joy on the anti-evolution scenarios was found which may be responsible for the slightly lower measurement reliability for this construct ($\rho_{\text{Rasch}} = 0.77$). The loadings for joy on the anti-evolution scenarios tended to be lower—below 0.5 on many of the scenarios—and accompanied by a negative loading onto the pro-evolution experience factor. This indicates that asking for a rating of joy on anti-evolution scenarios actually aligns negatively with pro-evolution experience in addition to aligning with anti-evolution experience. It is

noteworthy that the magnitude of the loading of joy on Scenario 8 is actually higher for the pro-evolution experience factor (-0.441) than for the anti-evolution experience factor (0.297). This unique pattern in Scenario 8 may be due to its political reference which may solicit a more emotionally loaded response than the other scenarios. A key point brought forth in our data is that despite the semantics, the constructs of “pro-evolution experience” and “anti-evolution experience” are not true opposites of each other; they are only moderately related ($r = -0.517$). This demonstrates that a teacher expressing pro-evolution experience will not necessarily express anti-evolution experience in a diametrically opposite way. This is particularly true for those who have relatively neutral levels of experience—we would expect more unpredictability in how these teachers react to the scenarios.

The Experience of Regret

The experience of regret, comprising the emotions of guilt and shame regarding events adverse to evolution education also showed remarkable consistency across the anti-evolution scenarios. Although this construct consisted of fewer responses than the others, the reliability was high ($\rho_{\text{Rasch}} = 0.85$). The distribution of teachers' measures for regret were similar to anti-evolution experience—a majority did not express regret even if they expressed pro-evolution experience. This accounts for the more moderate than expected correlation between pro-evolution experience and regret ($r = 0.40$).

An interesting aspect of this construct is the high separability between shame and guilt along the Wright map (Figure 4). Figure 4 shows that a majority of the teachers expressed neither shame nor guilt, indicating that they did not feel in an emotional way a sense of responsibility for the events hindering the teaching of evolution (Plutzer et al., 2020). However, those teachers with moderate measures expressed shame, and only those with the highest measures expressed guilt. This makes sense in that shame can be directed either inward towards oneself or outward towards others. Guilt, on the other hand, can only be directed inward, and therefore is expressed only by those teachers with the highest sense of personal responsibility for these negative events, which may tend to manifest as teachers gain more experience in the classroom (Noland, 2021). It is interesting that the only item where shame and guilt are positioned closely to each other is Anti 4, which refers to actions taken by a school board. This item is unique in that it refers to actions taken by a school while the other items refer to actions taken by states or individuals comparatively disconnected with the immediate school governance. The fact that this scenario asks teachers to reflect on actions taken by a school—thus hitting close to their own lives—may solicit inwardly-directed shame which relates more closely to the feeling of guilt than outwardly-directed shame.

Misfitting Responses and Next Steps for Improving the Measures

The data demonstrated that the E-EARS scales exhibited adequate reliability and that a majority of the items exhibited adequate construct validity evidence. Room for improvement exists nonetheless. PCA on residuals of the subscales indicated some departure from unidimensionality, and by taking a closer look at the items misfitting the Rasch model, we can work toward making the scales more unidimensional. The first significant pattern in Table 7 shows that the emotion of joy contains the largest number of misfitting responses. Specifically, joy on pro-evolution scenarios 1 and 2 misfit the Rasch model for the pro-evolution experience scale (Table 7). Figure 2 shows that these are the also easiest responses, meaning that respondents with relatively low levels of pro-evolution experience tended to express joy on these scenarios. From the misfit, it follows that even teachers at the bottom of the scale expressed joy on these constructs. It is interesting that both of these scenarios refer to curricular improvement which may have enhanced the tendency of teachers to express joy on these items despite relatively low measures for pro-evolution experience. We also see that the emotion of joy misfits with the anti-evolution experience scale for 5 of the 8 items. Here, we see that anti-evolution scenarios 3, 6, and 7 address the theory of creation more directly than the misfitting scenarios. It makes sense that those teachers with literalist monotheistic worldviews would express strong negative emotions regarding the pro-evolution scenarios coupled with strong feelings of joy on the scenarios specifically condoning supernatural explanations for the origins of life (Peñaloza et al., 2021). The influence of referencing curriculum and specific religious events on the solicitation of teachers' emotions is important to consider when planning future revisions and use of the E-EARS.

Two scenarios (Pro 1 and Anti 8) had three misfitting emotional responses, and an additional two scenarios (Pro 2 and Anti 5) had two misfitting responses (Table 7). It is instructive to look at these scenarios more closely to understand how they differ from the scenarios with comparatively well-fitting responses. It is interesting that the first two pro-evolution scenarios (Pro 1 and Pro 2) refer to action taken by state governments (Florida and Kansas, respectively) whereas the other two pro-evolution scenarios with comparatively well-fitting responses refer to a children's book (Grandmother Fish) and a television icon (Bill Nye the Science Guy). We observe that the misfitting negative emotions in Table 7 comprise submissive emotions (fear, shame, and guilt) (Bradley & Lang, 2000). This tells us that fear, shame, and guilt related to government actions measure anti-evolution experience somewhat inconsistently in comparison to when these emotions are expressed towards children's educational programming. It is difficult to speculate in an informative way on why this inconsistency exists; this may require further exploration with open-response or interview methods.

Table 7. Misfitting responses (Rasch model infit or outfit > 1.5) by scenario and emotion.

Emotion	Scenario											
	Pro1	Pro2	Pro3	Pro4	Anti1	Anti2	Anti3	Anti4	Anti5	Anti6	Anti7	Anti8
Joy	Pro	Pro			Anti	Anti		Anti	Anti			Anti
Anger												Pro
Sadness				Anti							Pro	
Fear	Anti											
Disgust												
Shame		Anti										
Guilt	Anti								Regret			Regret

"Pro" refers to the construct of pro-evolution experience; "Anti" to anti-evolution experience; "Regret" to regret over anti-evolution scenarios

Interestingly, anti-evolution scenarios 5 and 8 (Anti 5 and Anti 8) also address politics and government, Anti 5 referring to the legislatures passing a bill, and Anti 8 referring to a statement by Vice President Mike Pence, both condoning intelligent design as a valid theory for explaining life on Earth. The way that teachers with high measures for pro-evolution experience express anger related to the Mike Pence scenario is inconsistent with the way that they express anger on the other anti-evolution scenarios, suggesting a measurable bias due to the political reference in this scenario, consistent with Gough (2021). Similarly, the way that guilt is expressed on these two items is inconsistent with how it is expressed on the other anti-evolution scenarios, suggesting that these scenarios solicit guilt in a measurably different way than those scenarios which address implementation of or changes to educational resources by instructional entities.

We conclude that the consistency of the E-EARS may be improved by removing scenarios describing statements by government or political entities and instead using scenarios that address contexts related to schools, classrooms, or educational resources *or* creating separate assessments using scenarios describing educational resources or government/politics, respectively.

Another source of inconsistency was the inclusion of joy. Although factor and reliability analyses suggested that joy on pro-evolution scenarios fit with pro-evolution experience, and joy on the anti-evolution items fit with anti-evolution experience, the Rasch model suggests some inconsistency in the responses as compared with the negative emotions on the anti- and pro-evolution scenarios, respectively. In light of this, consistency could be improved by: (1) removing joy from the instrument, or (2) making separate constructs to account for joy. Despite the misfit, we would caution against these practices since: (1) joy is an important utilitarian emotion and has positive valence (Scherer, 2004), (2) separating joy from the other emotions is not supported by the factor analyses, and (3) inclusion of joy within the pro- and anti-evolution experiences makes conceptual sense. We would recommend that efforts to improve consistency of the E-EARS first concentrate on improving the uniformity of scenario contexts (described above) before attempting to remove or partition out joy from the measure.

External Factors Driving Levels of Emotional Experience

Acceptance of evolution (qualified as *appraisal* in this study) was the most important variable dictating a teacher's level of emotional experience. Acceptance of facts about evolution was the most important positive predictor of pro-evolution emotion and regret around anti-evolution scenarios. Acceptance of microevolution and statement that evolution is true and valid were important negative predictors for a teacher's level of anti-evolution emotion. A question arising from this result is: why is acceptance of microevolution such an important determinant of anti-evolution emotion? What this implies is that since microevolution is relatively easy to accept, teachers with low acceptance of microevolution tend to have high levels of anti-evolution emotion. It is quite common to accept microevolution without accepting macroevolution (Barnes et al., 2021; Nadelson & Southerland, 2012). It makes sense that teachers who accept that microevolution can occur (for example, accepting that events like antibiotic resistance and changes in viruses can occur within observable time frames) are going to agree that evolution is important to teach in the classroom despite not agreeing with all aspects of the theory. This may lead to more temperate expression of anti-evolution emotion than those teachers who reject evolution in its entirety. As a whole, we did not find the importance of these appraisal constructs surprising given the temporal causal link between cognitive appraisal and the experience of feeling (Scherer, 2005); however, this is the first study to confirm this link in the context of evolution education.

Perhaps more interesting is that there were only two additional variables accounting for teachers' emotional levels: subscription to monotheistic faith (negatively related to pro-evolution emotion) and preparation to teach evolution (negatively related to both anti-evolution emotion and regret around anti-evolution scenarios). The negative relationship between monotheism and pro-evolution emotional experience makes sense in light of the teleological battle between evolution and literalist accounts for creation associated with monotheistic belief systems (Barnes et al., 2021). However, these belief systems are generally associated with rejection of evolution (Hill, 2014), and so we were surprised that monotheistic faith did not arise as an important predictor for level of anti-evolution emotion.

Toward explaining this, it is important to note that monotheistic faith and level of anti-evolution emotion did exhibit a significant positive correlation ($r = 0.272$). However, the covariance between monotheism and level of anti-evolution emotion was subsumed by the more important variables: acceptance of microevolution, statement that evolution is true and valid, and preparation to teach evolution. After accounting for these variables, monotheistic faith became a non-significant predictor.

The data indicate that preparation to teach evolution is among the most important predictors for a teacher's emotional reaction to these scenarios, negatively related to levels of anti-evolution emotion and regret. The directionality and effect sizes of these relationships are similar, but for different reasons. The negative relationship with level of anti-evolution emotion may be due to the fact that teachers who display high emotional levels resist the theory and therefore pursue less preparation (Silva et al., 2021). Conversely, this relationship could reflect that more preparation to teach evolution, including appropriate coursework and professional development, may help teachers become more comfortable with ideas associated with evolutionary theory (Silva et al., 2021). Since preparation to teach a subject is engendered through active pursuit of coursework, professional development, and reflective practice, it makes sense that teachers who report high preparation will feel that they have done their due diligence to help teach evolution, leading to lower levels of shame and guilt. The trend of teachers with lower levels of preparation reporting higher levels of shame and guilt may be reflective of the awareness of less prepared teachers that they need to be doing more to teach effectively and advocate for the theory of evolution in their science classrooms. This is supported by the observation that a majority of the teachers participating in this study reported emotional experiences reflective of a generally favorable attitude towards the theory of evolution.

Conclusions

The development of the MATE (Rutledge & Warden, 1999) was an important step commencing the journey toward understanding how both students and teachers relate to the theory of evolution on an emotional level. Although the development of the I-SEA and GAENE built upon the MATE, offering some key developments in the epistemological underpinning of how acceptance is qualified and measured, the wording and structure of the items is quite similar both in wording and quantitative functioning (Romine et al., 2018). These measures of 'evolution acceptance' have been critical to our efforts to understand how learners process new ideas about evolution (Deniz et al., 2008; Ha et al., 2012), but only encompass the first step toward understanding actual emotional responses, and ultimately teachers' decisions regarding evolution pedagogy in the classroom.

The construct of emotion is complex and multimodal. Development of the Evolution Emotions Scale (Heddy & Sinatra, 2013) laid groundwork for beginning to understand how to quantify emotions about evolution using a self-report methodology, and Hawley and Sinatra (2019) demonstrated the efficacy of a valence-based dimensional measurement approach with Christian science teachers. The present study extends this work by focusing on arousal in addition to valence. In addition, we found a positive link between acceptance and emotion. This is an important finding in the sense that emotion begins with appraisal of a situation or stimulus event which is relevant to the concerns of the teacher or learner, and then progresses to emotional experience which encompasses subjective feelings about a situation based upon the appraisal. These then lead to physiological outcomes including changes in breathing, blood pressure, and heart rate (Stemmler, 2002). Finally, a course of action is chosen; in our context, whether or not to implement or participate in a learning activity related to the theory of evolution. Current 'evolution acceptance' instrumentation measures the first step in this cascade, *appraisal*; the E-EARS was designed the measure the second step, *subjective feelings* (Scherer, 2005) building off of previous work (Hawley & Sinatra, 2019; Heddy & Sinatra, 2013). An important next step for future research involves understanding the physiological outcomes that are induced by a learner's subjective feelings, which could be explored using either observational (Gottman et al., 2005) or physiological sensor data (Romine et al., 2020, 2021).

Limitations and Recommendations

We discuss potential directions for future research in light of limitations of the E-EARS and our study design. First, it may make sense to consider the conceptual domain of the E-EARS a limitation in that only seven emotions were measured, and only one of these (joy) was a positive emotion. As indicated previously, we chose to focus on these seven emotions due to their designation of *utilitarian*. The utilitarian emotions are particularly important since they guide decisions that affect our well-being (including whether or not to engage in a learning activity) directly: "Such adaptive functions are the preparation of action tendencies (fight, flight), recovery and reorientation (grief, work), motivational enhancement (joy, pride), or the creation of social obligations (reparation)." (Scherer, 2005, p. 706). Future research, however, may instead focus on aesthetic emotions which tend to take a more positive orientation: "the feeling of awe, wonderment, admiration, bliss, ecstasy, fascination, harmony, rapture, and solemnity" (p. 706). Previous work with quantitative measurement of emotion with respect to the theory of evolution (Hawley & Sinatra, 2019; Heddy & Sinatra, 2013) included measurement of aesthetic emotions such as pride and boredom, but these were discussed and analyzed in terms of being positively or negatively valenced as opposed to being utilitarian versus aesthetic in nature. An ethnographic analysis by Zembylas (2002) identified primarily aesthetic emotions: excitement,

fascination, frustration, and disappointment, as among the most important emotions dictating an elementary science teacher's classroom experience. We predict that these emotions may be less manifest in the context of the specific pro- and anti-evolution scenarios that we presented given their focus on decisions and/or decisive statements; however, these could be evoked in an instrument containing scenarios focusing on experiences with teaching students, the beauty and usefulness of the theory of evolution in the study of life, the harmony and simplicity that the theory of evolution brings to the discipline of biology, fascination with evolutionary processes like natural selection and mutation, or admiration for the likes of Charles Darwin or Paul Berg who undertook key discoveries allowing us to understand and apply evolutionary theory to modern problems in biology and medicine.

A second limitation relates to our study design; that we invited science teachers at all academic levels and types of schools to participate. This decision was intentional, but nonetheless limits our ability to generalize our results to specific grades, types of schools, or academic contexts. To the end of exploring construct validity, we considered our approach of inviting all science teachers to participate to be conservative in that major differences in instrument functioning between different types of teachers would be expressed in attenuated reliability and construct validity measures. In addition, we note that this study focused on science teachers, not on students (similar to Hawley & Sinatra, 2019). A teacher's emotional orientation can dictate decisions made in the classroom, and the same goes with students. The way students relate to the specific contexts presented in the E-EARS may be quite different than the reactions expressed by teachers. We hypothesize that this is particularly true for the scenarios addressing school or curriculum reform—using these same scenarios for students may present face validity issues given that curriculum development is not an area of concern for many students. If using the E-EARS for studies with students, we recommend building on Heddy and Sinatra's (2013) approach of assessing students' reactions to particular learning experiences. Another possible approach may include revising the scenarios to focus on pop culture references or community opinions similar to scenarios Pro 3 and 4 (Reading Grandmother Fish and the Statement by Bill Nye the Science Guy) and Anti 6 and 7 (Field trip to Ark Encounter and the Statement by John D. Morris, president of the Institute of Creation Research).

Finally, it should be noted that this study takes a positivist approach, exploring emotion in terms of a testable outcome as opposed to a process. The stepwise regression analyses yield some insight into the types of experiences that may be responsible for the way that teachers react to the scenarios on the E-EARS, and provides evidence that evolution acceptance is a mediator for emotional experience. However, our data cannot illuminate processes by which factors like religious background, preparation to teach evolution, or evolution acceptance help a teacher construct and negotiate unique emotional experiences. In addition to offering a novel approach to measure emotional outcomes, we hope that this research will provide a useful starting point for scholars who wish to utilize constructivist and post-modern perspectives similar to Zembylas (2004) toward exploring the role of science teachers' experiences—both curricular and extracurricular—in dictating how they react to and negotiate the range of perspectives offered by their students, school, and community with respect to how best teach evolutionary theory.

Author Contribution Statement

Romine and Todd contributed to study design and data acquisition, Romine and Mahajan performed data analysis. Romine drafted and revised the manuscript. All authors participated in interpretation of the findings and in the final review and approval of the manuscript.

Acknowledgements

The authors thank Brian Bartel and Lab out Loud for assistance with data collection. The authors are grateful to Ryan Dunk for providing critical review and constructive feedback on the manuscript.

References

- Arribas-Ayllon, M., & Walkerdine, V. (2017). Foucauldian discourse analysis. In C. Willig & W. Rogers. *The Sage handbook of qualitative research in psychology* (2nd ed., pp. 91-108). Sage. <https://doi.org/10.4135/9781526405555.n7>
- Barnes, M. E., Roberts, J. A., Maas, S. A., & Brownell, S. E. (2021). Muslim undergraduate biology students' evolution acceptance in the United States. *Plos One*, *16*(8), 1-20. <https://doi.org/10.1371/journal.pone.0255588>
- Barnes, M. E., Supriya, K., Zheng, Y., Roberts, J. A., & Brownell, S. E. (2021). A new measure of students' perceived conflict between evolution and religion (PCoRE) is a stronger predictor of evolution acceptance than understanding or religiosity. *CBE—Life Sciences Education*, *20*(3), 1-16. <https://doi.org/10.1187/cbe.21-02-0024>
- Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*, *3*, 77-85. <https://doi.org/10.1111/j.2044-8317.1950.tb00285.x>
- Beard, J., Nelson, C., & Nickels, M. (2014). *Evolution and the nature of science institutes (ENSI)*. Indiana University. <https://ensiweb.bio.indiana.edu/index.html>

- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238. <https://doi.org/10.1037/0033-2909.107.2.238>
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588. <https://doi.org/10.1037/0033-2909.88.3.588>
- Berkman, M. B., Pacheco, J. S., & Plutzer, E. (2008). Evolution and creationism in America's classrooms: a national portrait. *PLOS Biology*, 6(5), 920-924. <https://doi.org/10.1371/journal.pbio.0060124>
- Bond, T., & Fox, C. M. (2013). *Applying the Rasch model: fundamental measurement in the human sciences* (2nd ed). Routledge. <https://doi.org/10.4324/9781410614575>
- 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. <https://doi.org/10.1002/sce.20106>
- Bowman, K. L. (2008). The evolution battles in high-school science classes: Who is teaching what?. *Frontiers in Ecology and the Environment*, 6(2), 69-74. <https://doi.org/10.1890/070013>
- Bradley, M. M., & Lang, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology. In R. Lane & L. Nadel (Eds.) *Cognitive neuroscience of emotion* (pp. 242-276). Oxford.
- Brewer, C. A., & Smith, D. (2011). *Vision and change in undergraduate biology education: A call to action*. American Association for the Advancement of Science. <https://visionandchange.org/finalreport/>
- Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate behavioral research*, 27(4), 509-540. https://doi.org/10.1207/s15327906mbr2704_2
- Cleaves, A., & Toplis, R. (2007). In the shadow of Intelligent Design: the teaching of evolution. *Journal of Biological Education*, 42(1), 30-35. <https://doi.org/10.1080/00219266.2007.9656104>
- Coburn, W. W. (1994). Comments and criticism. Point: Belief, understanding, and the teaching of evolution. *Journal of Research in Science Teaching*, 31(5), 583-590. <https://doi.org/10.1002/tea.3660310511>
- Collins, L. M., & Lanza S. T. (2010) *Latent class and latent transition analysis: with applications in the social, behavioral, and health sciences* (vol. 718). Wiley.
- Corwin, L. A., Runyon, C., Robinson, A., & Dolan, E. L. (2015). The laboratory course assessment survey: a tool to measure three dimensions of research-course design. *CBE—Life Sciences Education*, 14(4), 1-11. <https://doi.org/10.1187/cbe.15-03-0073>
- Czerniak, C. M., & Schriver, M. L. (1994). An examination of preservice science teachers' beliefs and behaviors as related to self-efficacy. *Journal of Science Teacher Education*, 5(3), 77-86. <https://doi.org/10.1007/BF02614577>
- Darwin, C. (1872). The expression of the emotions in man and animals. John Murray. <https://doi.org/10.7208/9780226220802>
- De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.
- Deniz, H., Donnelly, L. A., & Yilmaz, I. (2008). Exploring the factors related to acceptance of evolutionary theory among Turkish preservice biology teachers: Toward a more informative conceptual ecology for biological evolution. *Journal of Research in Science Teaching*, 45(4), 420-443. <https://doi.org/10.1002/tea.20223>
- Donnelly, L. A., Kazempour, M., & Amirshokohi, A. (2009). High school students' perceptions of evolution instruction: acceptance and evolution learning experiences. *Research in Science Education*, 39(5), 643-660. <https://doi.org/10.1007/s11165-008-9097-6>
- Eastwell, P., & MacKenzie, A. H. (2009). Inquiry learning: Elements of confusion and frustration. *The American Biology Teacher*, 71(5), 263-266. <https://doi.org/10.2307/27669426>
- Ekman, P. E., & Davidson, R. J. (1994). *The nature of emotion: Fundamental questions*. Oxford University Press.
- Fouad, K. E. (2018). Pedagogical implications of American Muslims' Views on Evolution. In H. Deniz & L. Borgerding (Eds). *Evolution education around the globe* (pp. 15-39). Springer. https://doi.org/10.1007/978-3-319-90939-4_2
- Fowler, S. R., Zeidler, D. L., & Sadler, T. D. (2009). Moral sensitivity in the context of socioscientific issues in high school science students. *International Journal of Science Education*, 31(2), 279-296. <https://doi.org/10.1080/09500690701787909>
- Frijda, N. H. (1986). *The Emotions*. Cambridge University Press.
- Glynn, S. M., Taasobshirazi, G., & Brickman, P. (2007). Nonscience majors learning science: A theoretical model of motivation. *Journal of Research in Science Teaching*, 44(8), 1088-1107. <https://doi.org/10.1002/tea.20181>
- Gottman, J. M., Murray, J. D., Swanson, C. C., Tyson, R. & Swanson, K. R. (2005). *The mathematics of marriage: Dynamic nonlinear models*. MIT Press.
- Gough, N. (2021). The Christian right's war on reality: Where do/should American science teachers stand? *Cultural Studies of Science Education*, 16, 1-10. <https://doi.org/10.1007/s11422-021-10058-4>
- Griffith, J. A., & Brem, S. K. (2004). Teaching evolutionary biology: Pressures, stress, and coping. *Journal of Research in Science Teaching*, 41(8), 791-809. <https://doi.org/10.1002/tea.20027>
- Gruber, J., Oveis, C., Keltner, D., & Johnson, S. L. (2011). A discrete emotions approach to positive emotion disturbance in depression. *Cognition and Emotion*, 25(1), 40-52. <https://doi.org/10.1080/02699931003615984>

- 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(1), 95-121. <https://doi.org/10.1002/tea.20449>
- Hair, J., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis* (4th ed.). Prentice-Hall Inc.
- Hargreaves, A. (1998). The emotional politics of teaching and teacher development: With implications for educational leadership. *International Journal of Leadership in Education*, 1(4), 315-336. <https://doi.org/10.1080/1360312980010401>
- 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(1), 117-132. <https://doi.org/10.1007/s12052-010-0294-1>
- Hawley, P. H., & Sinatra, G. M. (2019). Declawing the dinosaurs in the science classroom: Reducing Christian teachers' anxiety and increasing their efficacy for teaching evolution. *Journal of Research in Science Teaching*, 56(4), 375-401. <https://doi.org/10.1002/tea.21479>
- Heddy, B. C., & Sinatra, G. M. (2013). Transforming misconceptions: Using transformative experience to promote positive affect and conceptual change in students learning about biological evolution. *Science Education*, 97(5), 723-744. <https://doi.org/10.1002/sce.21072>
- Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, 17(1), 65-70. <https://doi.org/10.1111/j.2044-8317.1964.tb00244.x>
- Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, 66(3), 393-416. <https://doi.org/10.1177/0013164405282485>
- Hill, J. P. (2014). Rejecting evolution: The role of religion, education, and social networks. *Journal for the Scientific Study of Religion*, 53(3), 575-594. <https://doi.org/10.1111/jssr.12127>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185. <https://doi.org/10.1007/BF02289447>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Izard, C. E. (1991). *The psychology of emotions*. Plenum Press. <https://psycnet.apa.org/doi/10.1007/978-1-4899-0615-1>
- Johnson, R. L., & Peeples, E. E. (1987). The role of scientific understanding in college: student acceptance of evolution. *The American Biology Teacher*, 49(2), 93-98. <https://doi.org/10.2307/4448445>
- Jolliffe, I. (2011). *Principal component analysis*. Springer.
- Kaiser, H. F. (1970). A Second-Generation Little Jiffy. *Psychometrika*, 35(4), 401-15. <https://doi.org/10.1007/BF02291817>
- Lakatos, I. (1976). *Falsification and the methodology of scientific research programmes: Can theories be refuted?* Springer. https://doi.org/10.1007/978-94-010-1863-0_14
- Lamb, R. L., Annetta, L., Meldrum, J., & Vallett, D. (2012). Measuring science interest: Rasch validation of the science interest survey. *International Journal of Science and Mathematics Education*, 10(3), 643-668. <https://doi.org/10.1007/s10763-011-9314-z>
- Lane, R. D., Chua, P. M., & Dolan, R. J. (1999). Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. *Neuropsychologia*, 37(9), 989-997. [https://doi.org/10.1016/S0028-3932\(99\)00017-2](https://doi.org/10.1016/S0028-3932(99)00017-2)
- Lang, P. J. (1995). The emotion probe: studies of motivation and attention. *American Psychologist*, 50(5), 372. <https://doi.org/10.1037/0003-066X.50.5.372>
- Lasky, S. (2000). The cultural and emotional politics of teacher-parent interactions. *Teaching and Teacher Education*, 16(8), 843-860. [https://doi.org/10.1016/S0742-051X\(00\)00030-5](https://doi.org/10.1016/S0742-051X(00)00030-5)
- Linacre J. M. (2006). *Winsteps*. Mesa Press. <https://www.winsteps.com/manuals.htm>
- Linacre, J. M., & Tennant, A. (2009). More about critical eigenvalue sizes in standardized-residual principal components analysis (PCA). *Rasch Measurement Transactions*, 23(3), 1228. <https://www.rasch.org/rmt/rmt233f.htm>
- Malone, K. R., & Barabino, G. (2009). Narrations of race in STEM research settings: Identity formation and its discontents. *Science Education*, 93(3), 485-510. <https://doi.org/10.1002/sce.20307>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174. <https://doi.org/10.1007/BF02296272>
- Masters, G. N. (1988). Item discrimination: When more is worse. *Journal of Educational Measurement*, 25(1), 15-29. <https://doi.org/10.1111/j.1745-3984.1988.tb00288.x>
- Muthén B. O., Muthén L. K. (2012). *Mplus 7 base program*. Muthén & Muthén, Inc. <https://bit.ly/3CGgog5>
- Nadelson, L. S., & Southerland, S. (2012). A more fine-grained measure of students' acceptance of evolution: development of the Inventory of Student Evolution Acceptance—I-SEA. *International Journal of Science Education*, 34(11), 1637-1666. <https://doi.org/10.1080/09500693.2012.702235>

- National Association of Biology Teachers. (2011). NABT's statement on teaching evolution. Retrieved February 6, 2019, from <https://bit.ly/3Bc6Ckm>
- Nehm, R. H., & Reilly, L. (2007). Biology majors' knowledge and misconceptions of natural selection. *BioScience*, 57(3), 263-272. <https://doi.org/10.1641/B570311>
- Nehm, R. H., & Schonfeld, I. S. (2008). Measuring knowledge of natural selection: A comparison of the CINS, an open-response instrument, and an oral interview. *Journal of Research in Science Teaching*, 45(10), 1131-1160. <https://doi.org/10.1002/tea.20251>
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. National Academies Press. <https://doi.org/10.17226/18290>
- Noland, T. (2021). *Teacher guilt: How can it inform instruction in foundational skills in reading* (Doctoral dissertation, St. John's University). Sr. John's Scholar. https://scholar.stjohns.edu/theses_dissertations/248/
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, 32(3), 396-402. <https://doi.org/10.3758/BF03200807>
- Osborne, J. W., Costello, A. B., & Kellow, J. T. (2008). Best practices in exploratory factor analysis. In A. B. Costello (Ed.) *Best Practices in Quantitative Methods* (pp. 6-99). Sage. <https://doi.org/10.4135/9781412995627>
- Palmer, D. H. (2006). Sources of self-efficacy in a science methods course for primary teacher education students. *Research in Science Education*, 36(4), 337-353. <https://doi.org/10.1007/s11165-005-9007-0>
- Peñaloza, G., El-Hani, C. N., & Mosquera-Suárez, C. J. (2021). Between Scientific Ideas and Christian Religious Beliefs. *Science & Education*, 30, 931-965. <https://doi.org/10.1007/s11191-021-00218-x>
- Plutzer, E., Branch, G., & Reid, A. (2020). Teaching evolution in US public schools: a continuing challenge. *Evolution: Education and Outreach*, 13(1), 1-15. <https://doi.org/10.1186/s12052-020-00126-8>
- Popper, K. (1968). Conjectures and Refutations. *British Journal for the Philosophy of Science*, 19(2), 159-168. <https://philpapers.org/rec/POPCAR-7>
- 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(8), 460-467. <https://doi.org/10.2307/4452205>
- Romine, W. L., Barrow, L. H., & Folk, W. R. (2013). Exploring secondary students' knowledge and misconceptions about influenza: Development, validation, and implementation of a multiple-choice influenza knowledge scale. *International Journal of Science Education*, 35(11), 1874-1901. <https://doi.org/10.1080/09500693.2013.778439>
- Romine, W. L., Barnett, E., Friedrichsen, P. J., & Sickel, A. J. (2014). Development and evaluation of a model for secondary evolution educators' professional development needs. *Evolution: Education and Outreach*, 7(1), 1-10. <https://doi.org/10.1186/s12052-014-0027-y>
- Romine, W. L., & Sadler, T. D. (2016). Measuring changes in interest in science and technology at the college level in response to two instructional interventions. *Research in Science Education*, 46(3), 309-327. <https://link.springer.com/article/10.1007%2Fs11165-014-9452-8>
- Romine, W. L., Todd, A. N., & Walter, E. M. (2018). A closer look at the items within three measures of evolution acceptance: analysis of the MATE, I-SEA, and GAENE as a single corpus of items. *Evolution: Education and Outreach*, 11(1), 1-20. <https://doi.org/10.1186/s12052-018-0093-7>
- Romine, W., Schroeder, N., Edwards, A., & Banerjee, T. (2021). Longitudinal classification of mental effort using electrodermal activity, heart rate, and skin temperature data from a wearable sensor. In A. Del Bimbo, R. Cucchiara, S. Sclaroff, G. Farinella, T. Mei, M. Bertini, H. Escalante, & R. Vezzani (Eds.), *Pattern Recognition: ICPR International Workshops and Challenges (Part II)* (pp. 86-95). Springer.
- Romine, W. L., Schroeder, N. L., Graft, J., Yang, F., Sadeghi, R., Zabihimayvan, M., D. Kaderiya, & Banerjee, T. (2020). Using machine learning to train a wearable device for measuring students' cognitive load during problem-solving activities based on electrodermal activity, body temperature, and heart rate: development of a cognitive load tracker for both personal and classroom use. *Sensors*, 20(17), 1-14. <https://doi.org/10.3390/s20174833>
- Russell, G. W. (1983). Psychological issues in sports aggression. In J. H. Goldstein (Ed.) *Sports violence* (pp. 157-181). Springer. https://doi.org/10.1007/978-1-4612-5530-7_10
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110, 145-172. <https://doi.org/10.1037/0033-295X.110.1.145>
- Rutledge, M. L., & Warden, M. A. (1999). The development and validation of the measure of acceptance of the theory of evolution instrument. *School Science and Mathematics*, 99(1), 13-18. <https://doi.org/10.1111/j.1949-8594.1999.tb17441.x>
- Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research*, 8(1), 3-15. <https://doi.org/10.1191/096228099671525676>
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695-729. <https://doi.org/10.1177/0539018405058216>

- Scherer, K. R., & Ceschi, G. (2000). Criteria for emotion recognition from verbal and nonverbal expression: Studying baggage loss in the airport. *Personality and Social Psychology Bulletin*, 26(3), 327-339. <https://doi.org/10.1177/0146167200265006>
- Scherer, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research*, 33(3), 239-251. <https://doi.org/10.1080/0929821042000317822>
- Schoon, K. J., & Boone, W. J. (1998). Self-efficacy and alternative conceptions of science of preservice elementary teachers. *Science Education*, 82(5), 553-568. [https://doi.org/10.1002/\(SICI\)1098-237X\(199809\)82:5%3C553::AID-SCE2%3E3.0.CO;2-8](https://doi.org/10.1002/(SICI)1098-237X(199809)82:5%3C553::AID-SCE2%3E3.0.CO;2-8)
- Siani, M., & Yarden, A. (2020). Evolution? I don't believe in it. *Science & Education*, 29(2), 411-441. <https://doi.org/10.1007/s11191-020-00109-7>
- Sickel, A. J., & Friedrichsen, P. (2013). Examining the evolution education literature with a focus on teachers: major findings, goals for teacher preparation, and directions for future research. *Evolution: Education and Outreach*, 6(1), 1-15. <https://doi.org/10.1186/1936-6434-6-23>
- Silva, H. M., Oliveira, A. W., Belloso, G. V., Díaz, M. A., & Carvalho, G. S. (2021). Biology teachers' conceptions of Humankind Origin across secular and religious countries: an international comparison. *Evolution: Education and Outreach*, 14(1), 1-12. <https://doi.org/10.1186/s12052-020-00141-9>
- Smith, M. U., Snyder, S. W., & Devereaux, R. S. (2016). The GAENE—generalized acceptance of evolution evaluation: development of a new measure of evolution acceptance. *Journal of Research in Science Teaching*, 53(9), 1289-1315. <https://doi.org/10.1037/t67588-000>
- Smith, R. M. (1996). Polytomous mean-square fit statistics. *Rasch Measurement Transactions*, 10(3), 516-517. <https://www.rasch.org/rmt/rmt103a.htm>
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, 42(5), 893-898. <https://doi.org/10.1016/j.paid.2006.09.017>
- Stemmler, G. (2002). Methodological considerations in the psychophysiological study of emotion. In R. J. Davidson, H. H. Goldsmith & K. R. Scherer (Eds.), *Handbook of affective science* (pp. 225-255). Oxford University Press.
- Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M. & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *British Medical Journal*, 338, 157-160. <https://doi.org/10.1136/bmj.b2393>
- Thompson, J. J., & Windschitl, M. (2005). "Failing girls": Understanding connections among identity negotiation, personal relevance, and engagement in science learning from underachieving girls. *Journal of Women and Minorities in Science and Engineering*, 11(1), 1-26. <https://doi.org/10.1615/JWomenMinorScienEng.v11.i1.10>
- Trani, R. (2004). I won't teach evolution; it's against my religion. And now for the rest of the story. *The American Biology Teacher*, 66(6), 419-427. <https://doi.org/10.2307/4451708>
- Williams, B., Onsmann, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, 8(3), 1-13. <https://doi.org/10.33151/ajp.8.3.93>
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. MESA Press. <https://research.acer.edu.au/measurement/1/>
- Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Loff, P. (1994). Reasonable mean square fit values. *Rasch Measurement Transactions*, 8(3), 370. <https://www.rasch.org/rmt/rmt83b.htm>
- Wundt, W. (1896). *Grundriss der psychologie* [Outline of Psychology]. Alfred Kröner Verlag.
- Yacoub, S., Simske, S., Lin, X., Burns, J. (2003) Recognition of emotions in interactive voice response systems. In H. Bourland (Ed.), *Proceeding of the 8th European Conference on Speech Communication and Technology (Eurospeech 2003)* (pp. 729-732). International Speech Communication Association.
- Young, D. J. (1991). Creating a low-anxiety classroom environment: What does language anxiety research suggest? *The Modern Language Journal*, 75(4), 426-437. <https://doi.org/10.1111/j.1540-4781.1991.tb05378.x>
- Yuen, C. T., San San, W., Seong, T. C., & Rizon, M. (2009). Classification of human emotions from EEG signals using statistical features and neural network. *International Journal of Integrated Engineering*, 1(3), 71-79. <https://bit.ly/3byxdOm>
- Zembylas, M. (2002). Constructing genealogies of teachers' emotions in science teaching. *Journal of Research in Science Teaching*, 39(1), 79-103. <https://doi.org/10.1002/tea.10010>
- Zembylas, M. (2004). Emotional issues in teaching science: A case study of a teacher's views. *Research in Science Education*, 34(4), 343-364. <https://doi.org/10.1007/s11165-004-0287-6>
- Zwick, R. W., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442. <https://doi.org/10.1037/0033-2909.99.3.432>