The Generalized Acceptance of EvolutioN Evaluation (GAENE) 3.0: Enhancement and Validation of a New Measure of Acceptance

Amanda L. Glaze
Georgia Southern University

Scott Snyder
University of Alabama at Birmingham

Randolph A. Devereaux
University of Louisiana Monroe

Mike U. Smith
Mercer University

This study builds upon existing analysis of validity and reliability wherein the instrument performed at statistically strong levels in high school and post-secondary applications (Smith, Snyder & Devereaux, 2016). In response to concerns with prior validation, the researchers added items to address extremes in the Rasch person-item continuum, removed Rasch model mis fitting items, collapsed two correlated items, and conducted further analysis of construct (convergent) validity through comparison to two existing measures of acceptance. Furthermore, this study explored the construct validity of the GAENE and conducted a comparison among secular and religious university populations. Analysis of the GAENE and related measures determined (1) that the added items on the GAENE successfully address the extremes in the Rasch person-item continuum; (2) principle-axis factoring following removal of one item supported unidimensionality of the GAENE; (3) summated raw and Rasch-based score intercorrelations suggests the measures (GAENE, MATE & ATET) are not coincident to one another evidencing convergent validity, however, at least half the variance in GAENE was not explained by MATE or ATET, demonstrating divergent validity; and (4) the GAENE fit to assumptions of Rasch modelling following removal of two items.

Keywords: GAENE, evolution acceptance, philosophy of science, Rasch model

INTRODUCTION

Evolution teaching and learning has sometimes been lamented as a major failure of science education (Smith, Snyder, & Devereaux, 2016). In response to shortcomings that are demonstrated in both classroom assessment and in public opinion polls, evolution education research has seen exponential growth in focus over the last two decades (Glaze & Goldston, 2015; Pobiner 2016). A key facet of this discussion has been the differentiation between content understanding and acceptance of evolution. Although discussion persists as to the nature of acceptance and whether it is an appropriate goal for evolution instruction at various levels, our work has been based on the premise that the two can be fruitfully distinguished. Philosophers generally distinguish between the two, recognizing that acceptance, but not necessarily understanding alone, engenders a commitment to use that understanding (Smith, Snyder & Devereaux, 2016). Separation of the two concepts has two further benefits. First, it facilitates learning by students who find acceptance of evolution, but not understanding of evolution, as challenging to their personal beliefs. Second, it also facilitates the study of the factors that influence the efficacy of evolution instruction, such as teasing out the separate effects of knowledge and acceptance on understanding.

This study is based on defining the acceptance of evolution as "the mental act or policy of deeming, positing, or postulating that the current theory of evolution is the best current available scientific explanation of the origin of new species from preexisting species" (Smith, Snyder, & Devereaux, 2016, p. 1290). Some studies of evolution acceptance have produced conflicting findings, particularly among studies of the correlation between evolution acceptance and understanding (e.g., Brem, Raney, & Schindel, 2003; Deniz, Donnelly, & Yilmaz, 2008; Southerland & Sinatra, 2004). Although there are various explanations for these conflicting findings (Southerland & Sinatra, 2004), one possible explanation is that the measures employed in these studies (particularly, the Measure of Acceptance of the Theory Evolution [MATE] (Rutledge and Warden 2000)) have shortcomings in need of address.

Exploration of the literature surrounding the development and use of the MATE and other instruments, demonstrates critical shortcomings that impact reliability and validity of the measure as a whole or in part based on how it was developed and how it is used (Mazur, 2004; Nehm & Schonfeld, 2007; see table in Smith, et als. 2016, p. 1290). Once identified, these shortcomings served as a useful template for the generation of a new measure of acceptance, one that specifically addressed the criticisms of existing measures as a tool for improvement. Thus, researchers developed and standardized a more psychometrically robust measure—the Generalized Acceptance of EvolutioN Evaluation (GAENE-pronounced "gene") (Smith, Snyder & Devereaux, 2016). The purpose of this paper is to report enhancements of the GAENE, including new validation studies, to develop and test items that should be easier to endorse (including for those respondents who largely reject evolution), to investigate the behavior of a troublesome GAENE item shown in prior validation, and to investigate the need for collapsing two response options in research-related use of the GAENE suggested in the validation of the measure (Smith, Snyder & Devereaux, 2016).

The Importance of Psychometrically Sound Instruments for Research

Research is the cornerstone of scholarship and effective practice. Rating scales are the frequently used for monitoring attitudes, beliefs, knowledge, and skills regarding science. Assuring the psychometric quality of new and extant rating scales is important to the advancement of knowledge and practice. At a basic level, the primary purposes of psychometric analyses are to ascertain: (a) whether the scores derived from the summation of rating items reflect a position along a single continuum, (b) the extent to which these scores are free from random error (reliability), and (c) the extent that the items reflect the attributes they are intended to measure (validity). It is therefore imperative that instruments that researchers employ must be psychometrically rigorous with appropriate levels of validity and reliability and that those measures should be subjected to periodic testing and revision as needed. However, science education researchers often report reliability data without including enough supporting details for the validity of their measures (Rutledge & Warden, 2000). When validity is mentioned at all, the discussion often goes no further than what has been called "face validity". Other researchers simply mention the results of prior validation studies, often conducted with samples that are far from isomorphic with their own sample (Smith et al, 2016; see below).

Validity, simply put, is a measure of how accurate an instrument is in its measurement of what it purports to measure. Whereas it would be easy to state that an instrument is either valid or not valid, validity is not a binary construct, existing more on a continuum with some instruments having greater validity than others based on rigorous testing. This paper therefore centers around a further validation of the GAENE instrument following enhancement and revision based on prior study of the measure's validity in assessing acceptance of evolution. Our model for assessing validity is based on that endorsed by the American Education Research Association, the American Psychological Association and the National Council on Measurement in Education (1999). This includes positioning on the non-binary nature of validity, that validity is evidenced in the internal structure of the instrument, and that validity is centered upon the inferences made as a result of the instrument rather than the instrument itself. The Rasch scaling model provides a specific framework for evaluating and improving the validity of rating scales. As previously delineated in the GAENE 1.0 and 2.0 validation study:

> … the researchers sought to develop a unidimensional additive measure that addresses the degree of an individual's acceptance of the essential domain range of the components of evolution acceptance. The instrument is valid if: (i) there is agreement from experts that the items are important or essential for measuring the domain of acceptance of evolution; (ii) the resulting scale yields a unidimensional factor structure; and (iii) the items and scale perform in a manner that is consistent with Rasch modeling assumptions (Smith et al, 2016, p. 1292). Continued data-based validation of the GAENE is the primary purpose for this study.

To be useful in assessment, instruments must also be reliable. From a statistical standpoint, reliability is the consistency of a measure. Consistency for a rating scale concerns: (a) the degree to which items within a scale share common variance with other items and the additive total score (internal consistency from a traditional psychometric paradigm or person reliability from an item response psychometric paradigm), (b) the degree to which responses to a stable construct are consistent across repeated assessments (test-retest reliability), and (c) the degree to which the hierarchy of item difficulty is consistent across respondents (item reliability, from the item response psychometric paradigm. Like its counterpart, validity, reliability is also a non-binary descriptor, meaning that there exists a continuum of reliability. An instrument may be reliable to greater extent in some ways than in others (as opposed to claims that a measure "is reliable [or valid]" or not). In addition, a measure may demonstrate solid reliability but not be a robust measure of its intended construct, i.e., have low construct validity—a criticism we have lodged against the widely used Measure of the Acceptance of the Theory of Evolution (MATE) (Smith, Snyder, & Devereaux, 2016)).

A common error in academic research is the application of instruments that have been found to be robust with a given population to samples that are dissimilar to the test population in important ways. Whereas some populations are similarly situated in such a way as to make the application more reasonable, for example using a senior level secondary school measure on freshmen in a university setting, there are still inherent differences in those populations that have the potential to limit the reliability and validity of the measure in the second setting. One example of this error is the wide application of the MATE across samples including secondary students and college students, although the existing validity and reliability of the instrument were tested for use with in-service science teachers and with a small sample for university biology students (n=61) although not with sufficient rigor. Likewise, some researchers have continued to use the MATE without reference to the published shortcomings of the measure.

In response to criticisms of existing measures, researchers developed and published a rigorously validated evolution acceptance measure, called the Generalized Acceptance of EvolutioN Evaluation (GAENE—pronounced "gene"). Empirical testing of the measure produced a Lawshe content validity index (CVI) value of 0.76 and testing of the GAENE with large samples has shown high reliability among both high school and post-secondary groups with Cronbach's alpha values at 0.929 and 0.944, respectively. These standardization studies were conducted with the intention of application to both secondary and post-secondary students and included both individual statistical analysis of both separate and combined groups.

Calls for a Sound Measure of Evolution Acceptance in Evolution Education

Recent evolution education literature has focused on the importance of understanding how students and the public accept various statements about so-called "controversial" topics, as well as on the factors that impact evolution learning and understanding. Among the factors explored by science educators are the understanding of evolutionary concepts. Although the terms *acceptance* and *belief* are commonly used interchangeably in the public arena, our position which distinguishes between the two highlights important differences.

It is argued in the literature that beliefs are deeply-held personal feelings and endorsements that are based solely on the perceptions by the individual of the experiences they encounter. Belief requires no standard of evidence in that experiences and responses can be rational, irrational, or "extra-rational," based on how decisions about those beliefs are made by the individual. In short, there is no requirement for empirical, measurable evidence needed for a person to adopt a belief. This absence of a requirement for empirical support is most commonly applied to cultural and religious choices and is closely associated with the concept of faith, which is belief in something even in the absence of physical evidence.

Recently, researchers have increased their efforts to develop and employ psychometrically robust measures of acceptance. Previous research failed to demonstrate a strong and robust correlation between levels of evolution knowledge and acceptance that are typical in non-evolutionary topics such as photosynthesis (Sinatra, Southernland, McCanaughy, & Demastes, 2013), and there were major concerns that these findings related to shortcomings in the measures. Several articles in the last decade provide a strong summary of the study of acceptance of evolution (Glaze & Goldston, 2015; Pobiner, 2016; Smith & Seigel, 2016). These authors tend to share the view that acceptance of evolution should be the aim of teaching and learning because acceptance is more likely to lead to action than understanding alone (Smith & Seigel, 2016). This shift in goal comes largely from the distinction between knowledge and acceptance with the latter representing a deeper level of commitment. Studies across levels have also demonstrated conflicting results with some showing correlation between knowledge and acceptance (Sharmann et al, 2005; Shtulman & Calabi, 2012) and others showing little to no correlation (Glaze, Goldston & Dantzler, 2015; Sinatra et al, 2003; Deniz & Donnelly, 2011; Kahan, 2015). Furthermore, as we delve into the factors that have been shown to impact acceptance across populations (students in secondary and post-secondary settings as well as teachers), we find not only that the relationships are complex to the point that neither one is consistently found to be a requisite of the other, but also that there are internal and external factors that impact acceptance (Allmon, 2011; Brem & Sinatra, 2012; Glaze & Goldston, 2015; Nehm & Schonfeld, 2007; Pobiner, 2016; Rissler et al, 2014; Smith & Siegel, 2004).

Obstacles to Measuring Acceptance of Evolution

A primary obstacle to exploring and impacting acceptance of evolution lies in the availability of psychometrically robust measures that have been tested on large samples that are representative of the same populations that are being studied. Existing measures of acceptance of evolution represent a spectrum of lengths, target populations, and theoretical frames and each has not only strengths but also criticisms and limitations to their research applications that are often overlooked. Two such measures were included in this validation study in an effort to compare performance in measuring acceptance of evolution through convergent validity analysis. For this research, the Measure of Acceptance of Evolution (MATE) (Rutledge & Warden, 2000) was selected due to its frequency of use in the literature as well as the Attitudes Toward Evolutionary Theories (ATET) (Konnemann, Asshoff, & Hammann, 2016), a measure of acceptance that has been recently introduced. As noted above, the Measure of Acceptance of Theories of Evolution (MATE) instrument is one of the most widely used measures of acceptance of evolution in the existing literature (Glaze & Goldston, 2015; Pobiner, 2016; Smith 2010). The MATE is a 20 -item, five-response Likert-type survey that features "fundamental concepts of evolution and the nature of science as inquiry" (Rutledge & Warden, 1999, p. 14). It employs both positive and negative stems and includes "processes, evidence, the ability of evolutionary theory to explain phenomenon, human evolution, earth age, and modern evolution as studied by the scientific community" (Rutledge & Warden, 1999 p. 14). Despite its widespread use, there are a number of criticisms that have not been clearly addressed (Romine et al. 2016, p. 2; Smith et al. 2016, p. 1293; Wagler and Wagler, 2013).

1. Lack of clear definition of the meaning of acceptance to drive the instrument.
2. Concerns about the content validity (reported as a single-round, open review of items by a five-member jury representing the "fields of evolutionary biology, science education, and the philosophy of science," and lack of constructive input or revision (Rutledge & Warden, 1999, p. 14).
3. Inclusions of evolution content questions and questions of a religious focus in the measure that are contraindicate because (i) researchers are often seeking to explore possible correlations between knowledge/understanding and acceptance and (ii) these items can unnecessarily conflate belief and acceptance (Smith, Snyder & Devereaux, 2016).
4. Absence of any quantitative exploration of the measure's validity or of the selection of cutoff scores indicating different "levels" of acceptance. (Smith, Snyder & Devereaux, 2016).
5. Widespread usage of the measure across a variety of populations (students at various levels, teachers, pre-service teachers) for which it has not been tested or has been tested with a less than adequate sample size (see Rutledge & Warden, 2007).

A new measure created in Germany sought specifically to address the conflation of knowledge and acceptance in the MATE and its validation only in populations in the United States. The Attitudes Toward Evolutionary Theory (ATET) instrument takes a novel social psychology approach to explaining evolutionary leaning by focusing on cognitive and affective evaluations and defines acceptance as a psychological attitude as opposed to the typical epistemological approach (Konnemann, Asshoff & Hammann, 2016; Maio & Haddock, 2010). This measure consists of 23 Likert-type items that focus on four attitude domains: "general attitude about evolution (5 items), belief in the promises of evolutionary theory (5), reservations about evolutionary theory (7) and emotions concerning evolutionary theory (7)" (Konnemann et al, 2016, p. 685). However, in the original study, reliability of the measure was only minimally established, and the validity of the measure was inadequately supported, as evidenced by failure to examine domain dimensionality, intercorrelations among the four domains, and intercorrelations with existing measures (Konnemann et al, 2016, p. 697).

Development and Shortcomings of the GAENE Instrument

The GAENE consists of statements about evolution which secondary and post-secondary students are asked to rate according to their agreement on a 5-option scale (1=Strongly Disagree, 2=Disagree, 3=I don't know/no opinion, 4=Agree, 5=Strongly Agree). Although the GAENE is not designed for individual performance comparison, it is intended for the study of evolution acceptance of groups of individuals in both classroom and research settings (Smith, Snyder & Devereaux, 2016). The 14-item GAENE 2.0 had an initial content validity index of 0.76 (Lawshe, 1975) and a Flesch-Kincaid reading level score of Grade 9.4--appropriate for use in both secondary and post-secondary students. The internal consistency/reliability was high, both for the high school and post-secondary groups (Cronbach alpha = 0.929 and 0.944, respectively), as well as for the combined samples (0.941).

While there are many strengths demonstrated in the statistical analysis of the GAENE, there were shortcomings that were also identified that are explored further in this study. First, although the measure was deemed to demonstrate construct validity, there is a need to demonstrate convergent validation as well in order to further support the construct validity of the instrument. Convergent validity is present when an instrument performs well in comparison to another measure of the same construct. The extent of shared variance between the GAENE, MATE and ATET provides evidence of convergent validity. A second issue identified in the earlier analysis of the GAENE is an absence of items at the extremes of the Rasch person-item continuum wherein either the sample did not contain representation from the extremes of the measure or the instrument was not suited to distinguish among individuals at the extreme high and low end of the spectrum. Third, one item in the original GAENE was identified for mis fitting the Rasch measurement model. The authors proposed that the misfit may have been affected by presentation ordering in the original study. For the purposes of this study, the item identified in that study was moved later in the measure. Finally, two items that were shown to have correlation in the prior validation study will be more closely examined to test and affirm the need to collapse the items into a single item to eliminate redundancy. The shortcomings previously mentioned are addressed through statistical analysis in this study.

METHODS

Research Questions

1. Do newly proposed (NGAENE) items extend the range of item difficulty of the GAENE, (i.e. the ability of the GAENE to discriminate among "strong acceptors" and among "strong rejectors")?
2. Do MATE and ATET items that are unidimensional with the GAENE (that load on the same factor as do the GAENE items) increase the spread of item difficulty for assessing acceptance of evolution?
3. What is the nature of intercorrelation between the GAENE, ATET and MATE (testing convergent and divergent validity)?
4. To what extent does the expanded GAENE measure perform in a psychometrically sound manner that is consistent with expectations of the Rasch model when administered to a collegiate sample?
5. Is there a practically significant difference between the levels of evolution acceptance of students at religiously affiliated and public colleges as measured by GAENE 3.0

Sample

Items from the expanded GAENE (v 2.0, 14 items + 10 proposed new items), the MATE (20 items), and the ATET (20 items) were completed by 952 subjects. The majority of subjects (n=848, 89.1%) were from secular institutions and were female (n=577, 60.6%). The majority of respondents were attending secular colleges (n=804) while 99 subjects were attending a religiously based college. A combination of purposive and convenience sampling was employed in soliciting participants for this study. The sample was drawn from introductory science courses whose instructors responded to an email invitation. This population was purposively selected because it is the likely target population for future research studies that might employ a measure of evolution acceptance. The sample is also one of convenience, i.e., the sample was not drawn so as to be statistically representative of any population.

Participants for the validation study were solicited through their individual courses via an email message that was shared with professors of sciences at 111 institutions in all 50 states and Puerto Rico as a part of a larger study on acceptance and understanding nationwide. The study email included a link to Survey Monkey where the student participants were provided consent information and the choice to participate or not participate. A waiver of documentation of consent was approved for this study. All study procedures were approved by the Institutional Review Boards of the first and last authors' host institutions.

Generation of Proposed New (NGAENE) Items

Unlike methods used to generate items for previous GAENE versions, proposed items were generated by the authors based entirely on their knowledge of prior research and their understanding of the range of student views of evolution. These items were generated in an explicit attempt to make them endorsable by students who hold either extremely positive or extremely negative views of evolution. The first set of proposed items were generated by the last author, who also interviewed five acquaintances known to hold extreme (positive or negative) views about evolution. Initial items were revised based on these interview responses. The first and last authors then reviewed, revised, and culled the item set, resulting in 10 additional items to be tested here.

Analysis Methods

Research Question One: *Extension of GAENE Range of Item Difficulty*

An important aim of this research was to generate and test additional GAENE items that would extend the range of item difficulties in order for the measure to better discriminate among individuals who hold particularly strong or particularly weak acceptance of evolution views as measured by the GAENE. Examination of differences between mean item difficulty and mean subject ability provide an indication of the overall congruence of the scale to the sample, and examination of the Rasch Wright Map provides information about the extent to which the NGAENE items extended the range of the acceptance of evolution construct.

Research Question Two: *Unidimensionality of MATE, ATET, and GAENE 2.0 and NGAENE Items*

We conducted principal axis factoring (a form of orthogonal factor analysis) with the combined set of all GAENE, MATE and ATET items to determine whether items from the MATE and ATET shared the underlying dimension addressed by the GAENE. Items from the ATET and MATE that loaded with GAENE items were submitted to Rasch analysis with WINSTEPS. Because the ATET uses a four-point rather than the five-point response scale used by the GAENE and the MATE, the "partial credit" scoring method was required in order to allow for the different item structures across the three scales. The resulting Wright Map was examined to determine the extent to which the inclusion of ATET and MATE items expanded the hierarchy of the acceptance of evolution construct. Items found load on the same factor as GAENE 2.0 items and are shown by the Wright map to extend the range of the GAENE were identified and considered for inclusion in the final version of the GAENE.

Research Question Three: *Intercorrelation of GAENE, MATE and ATET Scores for Participants*

As an indication of the convergent and divergent validity of the summated raw scores of the GAENE, 10 proposed items, MATE and ATET and the raw score and Rasch scores for the GAENE were submitted to correlational analysis for subjects who completed all three measures. It was expected that if the GAENE items measure a construct that is a similar to, but distinct from that measured by the ATET and the MATE, then intercorrelations between the GAENE and each of the other measures would be high (supporting convergent validity) but not high enough to suggest unity (supporting divergent validity).

Research Question Four: *Fit of GAENE to Assumptions of the Rasch Model*

To address the question of the fit of the GAENE to the assumptions of the Rasch model, subject responses to the 14-item GAENE 2.0 and the 10 NGAENE items were analyzed using the WINSTEPS (Linacre, 2017) program. The WINSTEPS program tests whether subjects and items perform in a manner consistent with the assumptions of the Rasch scaling model. Table 1 provides a listing of these assumptions along with the analytical procedures we employed to test each.

TABLE 1
RASCH MODEL ASSUMPTIONS

Assumption	Analytical Procedure
1. Scale measures a single dimension (unidimensionality)	Principal components analysis of Rasch residuals (PCAR)
2. Items perform in a manner consistent with the mathematical assumptions of the Rasch model	Examination of mean square infit and outfit statistics
3. Categories of the rating scale behave in the ordinal manner that was expected	Visual examination of category performance plots and analysis of expected threshold ordering
4. The measure reliably separates subjects into distinct ability levels	Examination of person separation reliability and separation index
5. The range of items is well-matched to the range of abilities of respondents (targeting)	Examination of Wright Map that sorts subjects and items on a common scale
6. The items on the scale perform consistently for different groups of subjects (invariance)	Differential Item Function (DIF) analysis

The ability to sum ratings across a set of ordinal-response items such as those of the GAENE in order to generate an interval-level score reflecting the amount of a specific trait requires each of these assumptions. Researchers have raised concerns that summing rating items that do not meet the assumptions of the Rasch measurement model generates scales that are not, in fact, interval-level and that may perform differently across groups. This failure therefore limits the scientific application of such scales for comparing groups and for monitoring changes across time. On the other hand, scales that meet the assumptions of the Rasch measurement model place subjects and items on the same interval-level continuum (measured in logits). In addition to improving comparative and longitudinal comparisons, this approach allows for a range of important applications including using different subsets of items to determine reliably the "person ability" (i.e., evolution acceptance) levels of subjects.

To test for unidimensionality, we conducted a Principal Components Analysis of (standardized) Residuals (PCAR) (Linacre, 2017). Unidimensionality is tested by determining whether the differences between the actual performance of items and the expected performance of items based on the Rasch model (residuals) represent more than one significant principal component. If less than 50% of the variance in scores is explained by items that load on the primary Rasch dimension and if the second dimension accounts for three or more eigenvalues (i.e., the amount of variance of accounted for by three or more items), there is evidence of multidimensionality (Linacre, 2007).

To determine the extent to which GAENE 3.0 items and subjects performed consistently with the Rasch measurement model, we calculated fit statistics, which are reported in standardized scores. Person ability scores and item difficulty scores are placed on the same continuum. The higher a person's ability (acceptance) on the continuum, the easier it will be for them to endorse any item. Likewise, the higher the item is on the person-ability (acceptance) continuum, the more difficult it will be for any subject to endorse that item. INFIT mean square statistics indicate the degree to which the model fits when items and subjects are at about the same position on the continuum. OUTFIT mean square statistics describe the degree to which the model fits the data when subject ability and item difficulty are far apart. In general, items should exhibit INFIT and OUTFIT statistics between .6 and 1.4 (Linacre, 2005).

Again, using WINSTEPS, we produced graphic category performance curves for each of the five response categories (Strongly Disagree through Strongly Agree). When using Likert-type scales such as the GAENE, it is important to assure that the responses to the categories are consistent with the specified order from lowest to highest and that there is separation between categories (e.g., Tennant & Conaghan, 2007). When this does not occur, the "disordered" categories compromise the Rasch model. Disordered categories are typically managed by collapsing (combining) categories. Graphic plots and examination of category performance at thresholds between graphed response categories provide evidence of whether categories are performing as expected or whether there is a need to combine categories.

Next, we evaluated the targeting of the combined set of 24 GAENE items. Ideally, the range of the endorsability of the items will match the range of the subjects' levels of acceptance of evolution. In that case, the items are "targeted" to the subjects (Dmitrov, 2012). Targeting assessment employs the WINSTEPS Wright Map which places the subjects and items along the same continuum as described above. This map enables the researcher to visualize the hierarchy of items (arranged in order of how difficult they were to endorse) as well as how those items map to the range of "abilities" of subjects (levels of evolution acceptance). We evaluated targeting in two ways. We first assessed how far the mean estimate of person ability deviated from the mean of the item difficulty continuum (Fisher, 2007). Under the Rasch framework, the mean of the item difficulty is first set to zero (Linacre, 2003) and differences between mean person ability and mean item difficulty are compared. Mean differences of less than 1 logit are considered good; differences between 1 and 2 logits are considered fair (Fisher, 2007). In addition, rating scales such as the GAENE are also considered to be well targeted when fewer than 15% of subjects are likely to fall outside of the "operational range" of responses to categories (Lo, Liang, Hang, Wu, Change, & Chang, 2015)

For a scale to be useful, the difficulty (endorsability) of any item in a rating scale should not change based on variables that are unrelated to performance on the measure (e.g., gender, date of data collection, ethnicity, etc.) (Englehard, 2013). Subgroups that have the same general level of a trait (e.g., acceptance of evolution) should perform equivalently (i.e., invariantly) on each item. To test this functioning, we performed Differential Item Functioning (DIF) analysis within WINSTEPS, which assesses the degree of such invariance by evaluating item difficulty across two subgroups after controlling for person ability. If differences of greater than .5 logits (Wang, 2008) are present or if statistically significant Rasch-based and Mantel-Haenszel DIF values are statistically significant, there is evidence of a violation of invariance that should be examined.

Next, we addressed the reliability of the measure by computing person separation reliability. Within the Rasch framework, person separation reliability refers to whether ability estimates (estimates of evolution acceptance) are replicable. Person separation reliability is similar to the Cronbach's Alpha statistic, but person separation reliability tends to be slightly lower due to the exclusion of extreme scores (e.g., individuals who endorse "strongly agree" on all items). Similarly, item separation is the degree to which the spread of items would be reproducible in similar samples of respondents (Bond and Fox, 2011). Reliabilities greater than .75 are considered acceptable. Additionally, the person separation index (PSI) (Wright and Masters, 1982) was used to identify the number of statistically distinct subgroups of respondents (strata) along the person-ability continuum (evolution acceptance). It is generally desirable that a scale can separate subjects into two or more strata and that the scale can separate items into three or more strata.

Research Question Five: *Differences in GAENE 3.0 Rasch scores between public and religiously affiliated colleges*

An independent samples t-test using Rasch-based scores derived from the GAENE 3.0 was used to compare average levels of evolution acceptance of students in our samples of public and religious colleges, followed by calculation of Cohen's *d*. We predicted that students who attend religiously affiliated colleges would be likely to have more negative views of evolution than students attending secular colleges and thus would have lower GAENE scores than their secular peers.

RESULTS

Research Question One: Extension of GAENE Range of Item Difficulty (Targeting—NGAENE)

When item averages for combined 1-item GAENE 2.0 and NGAENE items were set to zero, the average ability score of subjects was +1.11 logits, indicating that for this sample that the average level of acceptance of evolution was higher than the average difficulty of the items. Based on the 2007 criteria by Fisher, this level of targeting is considered "fair."

Examination of the operational range of the scale from the Wright map reveals that the difficulty levels of the items are appropriate for approximately 90% of respondents (see Appendix A). The set of respondents who score a 5 (strongly agree) with all items represented approximately 1% of respondents. These subjects do not contribute in meaningful ways to the analysis of the scale. Therefore, although the scale is more sensitive for the performance of subjects who fall below the mean of the population (i.e., those with moderate to low levels of acceptance of evolution), the range of items has been expanded by the added items, and there are now items that are more challenging for subjects with high levels of acceptance of evolution. These 10 items will be added to the final measure, pending the remaining Rasch analyses.

As can be seen in the Wright Map provided in Appendix A, the three most difficult items to endorse are new items (NGAENE4, 5, and 6: "Evolution is the most important theory devised by man," "I would bet my life on the claim that evolution is true," and "Understanding evolution has changed my life"). In addition, the second easiest item to endorse is a new item (NGAENE10: "Small changes can occur in a species over time"). This observation provides evidence that the additional items expanded the difficulty of the scale in terms of items that were most challenging to endorse, also supporting coverage of items that tended to be easiest to endorse. These additional items should improve the precision of scores for respondents at these two extremes of evolution acceptance.

Research Question Two: Unidimensionality of MATE, ATET, and GAENE 2.0 Items and Resulting Rasch-based Item Targeting

When all GAENE, ATET and MATE items were subjected to principal axis factoring, a two-factor solution resulted in all GAENE items falling in a single factor with a subset of MATE and ATET items also loading on that factor (see Appendix B). When the items loading together on this factor were submitted to PCAR (principal components analysis of residuals) in Winsteps, the PCAR contrast had an eigenvalue of 4, suggesting that there may be a second dimension evident in the combined scale. The analysis of this contrast suggested that some of the unexplained variance in the contrast was due to NGAENE7 (the most difficult item to endorse: "There are no gaps in the fossil record"). Elimination of this item resulted in an eigenvalue of less than 3 for this contrast, supporting unidimensionality. This item will therefore not be added to the final measure. Note, however, that NGAENE7 was the most difficult item to endorse. Thus, removing this item slightly narrows the aimed-for extension of the range of items appropriate for subjects who are most likely to accept evolution.

Employing the partial credit scoring algorithm in Winsteps to accommodate different scoring scale structures employed by the three measures (described above) for Rasch analysis, the resulting Wright map for ATET and MATE items that load on the primary factor is provided in Appendix C.

The Wright Map reveals that the ATET and MATE items that load with GAENE items do not fall near the extremes of the map and therefore do not add to the precision of estimates of evolution acceptance for subjects who are most likely or least likely to accept evolution.

Research Question Three: Intercorrelation of GAENE, MATE and ATET Scores for Participants

Intercorrelations of GAENE 2.0 and NGAENE items (summated raw scores and Rasch-based scores) with ATET and MATE summated scores and Rasch-based scores generated the Pearson correlations shown in Table 2.

TABLE 2
INTERCORRELATIONS OF GAENE, MATE, & ATET INSTRUMENTS

		GAENE-Rasch Score	GAENE	MATE Scale	ATET Scale
GAENE-Rasch Score	Pearson Correlation	1	.946**	.670**	.616**
	Sig. (2-tailed)		.000	.000	.000
	N	903	903	853	836
GAENE Scale	Pearson Correlation	.946**	1	.763**	.716**
	Sig. (2-tailed)	.000		.000	.000
	N	903	903	853	836
MATE Scale	Pearson Correlation	.670**	.763**	1	.850**
	Sig. (2-tailed)	.000	.000		.000
	N	853	853	853	836
ATET Scale	Pearson Correlation	.616**	.716**	.850**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	836	836	836	836

**. Correlation is significant at the 0.01 level (2-tailed).

Based on the correlations of summated raw scores, GAENE 3.0 and MATE share 57.8% of the variance, GAENE 3.0 and ATET share 51.3%, (and MATE and ATET scales share 72.3%). When the Rasch-based GAENE 3.0 scores are correlated, GAENE 3.0 and MATE shared 44.9% of the variance and GAENE 3.0 and ATET shared 37.9%. As predicted, these analyses demonstrate that there is considerable overlap in the meaning of the construct of evolution acceptance between GAENE 3.0 and both the MATE and the ATET (58% and 51%, respectively). The fact that such a large proportion of the variance is not shared by GAENE 3.0 and the MATE (or the ATET), however, strongly suggests that the constructs being measured by these scales are not coincident.

Research Question Four: Fit of GAENE 3.0 to Assumptions of the Rasch Model

Rasch Assumption 1: Unidimensionality

Principal components analysis of residuals (PCAR) of the combined (24-item) GAENE data revealed that, although 59% of the variance was accounted for by the primary Rasch dimension, a second component explained three eigenvalues. However, when the items that formed the second component were examined, the contrast included only items that were easiest to endorse. This suggests that the second component is not a separate construct, but rather an artifact associated with difficulty. Thus, the PCAR analysis demonstrates adequate support for unidimensionality.

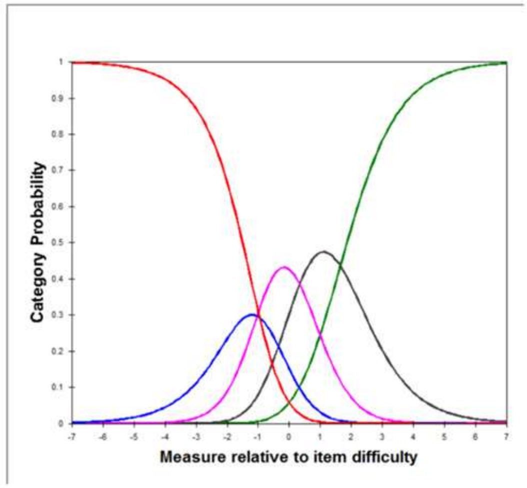
Rasch Assumption 2: Model Fit

Rasch analysis identified two items in the combined scale that misfit the assumptions of the Rasch model. Item 3 from the original scale ("Some parts of evolution theory could be true") yielded a MnSq outfit of 2.6, and Item 7 from the new items ("There are no gaps in the fossil record") yielded a MnSq infit of 1.7 (both > 1.4) and an outfit of 2.3. Elimination of these items brought the PCAR contrast eigenvalue down to an acceptable level of 2.2, and unidimensionality was further affirmed. Item GAENE3 will not be included in the final measure; Item 7 was already eliminated based on PCAR analysis above. (Note that GAENE3 is the same item that did not perform acceptably in our previous analysis.)

Rasch Assumption 3 Category Performance

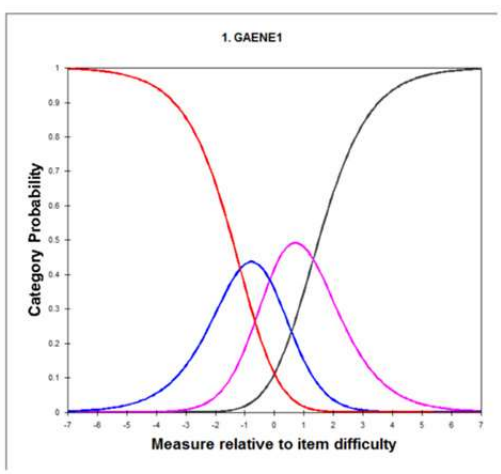
Given that a previously conducted Rasch analysis (Smith, Snyder & Devereaux, 2016) concluded that levels 1 and 2 of GAENE 2.0 responses should be collapsed (at least for research purposes), we began by examining category performance curves for the current dataset.

FIGURE 1
CATEGORY PERFORMANCE CURVES-CURRENT DATA SET



These curves indicated that levels 1 and 2 of the combined GAENE items (Strongly Disagree and Disagree) did not perform in an ordered and distinct manner. Specifically, the probability curves for Strongly Disagree and Disagree actually cross at a slightly higher point on the latent trait dimension than do the curves for Disagree and I don't know/No opinion. This is inconsistent with the expectation that these intersections (thresholds) will increase as evolution acceptance increases. If this expectation was met, the peaks in the category curves (modes of response in each category—1 to 5) would increase by category level along the latent dimension (e.g., acceptance of evolution). When Disagree and Strongly Disagree responses are merged resulting in four categories, the resulting category performance curves conform to Rasch expectations.

FIGURE 2
RASCH CONFORMITY OF CURVES



Furthermore, when thus modified, the resulting four-category model met Linacre's (2002) criteria for category performance, including appropriate infit and outfit MnSq values and appropriate intervals. As a result, subsequent analyses will be based on the merger of the Strongly Disagree and Disagree categories.

Rasch Assumption 4: Reliability

Item reliability for the combined GAENE is 1.0 and person reliability is .92 with person separation of 3.23, indicating that the scale reliably discriminates three "acceptance levels" of respondents from this population. The high item reliability indicates that differences in item difficulty values can be trusted for this sample and is likely facilitated by the large number of subjects in this sample. The high person reliability supports an argument that the score on the scale separates subjects based on ability and may be supported by the wide range of acceptance of evolution reported by the sample.

Rasch Assumption 5: Targeting

The appropriateness of the targeting of the expanded GAENE is addressed above.

Rasch Assumption 6: Invariance

One item (GAENE 11: "I would be willing to argue in favor of evolution in front of a small group of friends") was identified as exhibiting significant gender-based differential item functioning, as evidenced by both Mantel-Haenszel and Rasch DIF probabilities of below .01. This item was .38 logits more difficult for females to endorse than for males of equivalent ability. Given that this item functions well in our other analyses, we chose to retain it in the final measure, GAENE 3.0.

Research Question 5: Differences in GAENE 3.0 Rasch Scores Between Public and Religiously Affiliated Colleges

A statistically significant difference ($t=8.3$; sig $< .001$) was found in average Rasch scores between respondents from public colleges and universities (Mean=1.28, SD=1.82, n=804) and Christian colleges (Mean = -0.27, SD=1.11, n= 104), a difference off 1.53 logits (Cohen's d=0.88). These findings indicate that, as predicted, the average level of acceptance of evolution is substantially higher in public colleges than in religious colleges.

CONCLUSIONS

Summary

RQ1. Rasch analysis demonstrated that the 10 newly proposed items, especially NGAENE4, 5, 6 and 10 extended the range of the GAENE at both extremes of the continuum of evolution acceptance, particularly at high levels of evolution acceptance.

RQ2. In principal axis factoring of the combined data from all three measures, all GAENE items and a subset of MATE and ATET items loading on a single factor. PCAR analysis suggested that this factor was unidimensional once one item (NGAENE7) was eliminated. However, the Wright map revealed that none of the ATET or MATE items included in that factor fell near the extremes of evolution acceptance.

RQ3. Intercorrelations among the three measures were substantial (explaining 59% of the variance), supporting the expectation that the three measures are similar and evidencing convergent validity of the GAENE with these measures. On the other hand, almost half of the variance in GAENE responses is not explained by either the MATE or the ATET, evidencing expected divergent validity.

RQ4. GAENE 3.0 behaves in a psychometrically manner consistent with the assumptions of the Rasch model with one exception. Rasch misfit statistics identified for exclusion one previous GAENE2.0 item (GAENE3) and (again) NGAENE7. These items were removed to produce the final revised measure, the GAENE 3.0 (see full measure in Appendix D).

RQ5. As expected, differences between GAENE 3.0 scores of students at religiously affiliated and public colleges are both statistically and practically significant, if these convenience samples could be considered representative of the populations of American religiously affiliated and public colleges, with a large Cohen's *d* of more than one and a half standard deviations, which we judge to be practically significant based on similar study results.

IMPLICATIONS & FURTHER STUDY

GAENE 3.0 is recommended over prior versions when the target group is expected to have extreme views, especially if those views are likely to be positive. Otherwise, GAENE 2.0 continues to be acceptable for both classroom and research applications. Consistent with prior findings (Smith, Snyder & Devereaux, 2016), GAENE 2.0 Item #3 again mis performed, supporting a decision to remove this item from the final measure. Given that a primary application intended for this scale is in intervention studies in which pretreatment subjects are expected to have negative attitudes toward evolution, it may be worthwhile for researchers to seek additional items that can distinguish among subjects with low acceptance. On the other hand, the GAENE does distinguish well among positive and negative evolution attitude scores, therefore, the GAENE is appropriate for measuring treatment gains.

The intercorrelation data provide support for both the convergent and divergent validity of GAENE 3.0. These observations support the use of the measure by researchers and science educators. These data emphasize that the constructs measured by the GAENE, the MATE, and the ATET are similar, but not identical. As expected, the different definitions used to guide the development of each measure and the different development procedures employed have resulted in somewhat different measures. It is therefore incumbent upon future researchers interested in evolution acceptance to select the measure to use based on both the strength of its psychometric properties and the match between the researcher's concept of evolution acceptance and the theoretical framework under which the measure was designed. Furthermore, it will be necessary for researchers to provide an explicit description of why they chose the measure they employed.

Given that the MATE and the ATET are based on completely different understandings of evolution acceptance (as an educational construct or as a psychological construct, respectively), it is intriguing that they share more commonality (explained variance) than either one shares with the GAENE. Examination of the items that load on the second factor of the PCAR analysis suggests that this observation may be accounted for at least in part by the fact that both the ATET and the MATE include items related to religion that are not contained in the GAENE.

There may be a gender effect in one GAENE item (GAENE 11). The content of that item suggest that this differential effect may be related to gender differences in assertiveness within peer groups. Subsequent research with the scale should continue to examine this item for DIF, especially when the GAENE is used in studies where gender is a variable of interest.

REFERENCES

Allmon, W. D. (2011). Why don't people think evolution is true? Implications for teaching, in and out of the classroom. *Evolution Education and Outreach*, *4*, 648-665.

Alters, B. J., & Nelson, C. E. (2002). Perspective: Teaching evolution in higher education. *Evolution*, *56*(10), 1891-1901.

Brem, S. K., & Sinatra, G. M. (2012). Commentary on section II: bringing multiple levels of analysis to bear on evolution teaching and learning. In K. S. Rosengren, S. K. Brem, M. E. Evans, & G. M. Snatra (Eds.), *Evolution challenges: integrating research and practice in teaching and learning about evolution* (pp. 428-442). New York, NY: Oxford University Press.

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

Deniz, H., & Donnelly, L. A. (2011). Preservice secondary science teachers' acceptance of evolutionary theory and factors related to acceptance. *Reports of the National Centers for Science Education*, *31*(4), 2.1-2.8.

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

Glaze, A. L., & Goldston, M. J. (2015). Evolution and Science Teaching and Learning in the United States: A Critical Review of Literature 2000-2013. *Science Education*, *99*(3), 500-518.

Glaze, A. L., Goldston, M. J., & Dantzler, J. (2015). Evolution in the Southeastern United States: factors influencing acceptance and rejection in pre-service science teachers. *International Journal of Science and Mathematics Education*, *13*(6), 1189-1209. Doi: 10.1007/s10763-014-9541-1

Kahan, D. M. (2015). Climate-science communication and the measurement problem. *Advances in Political Psychology*, *36*, 1-43.

Konnemann, C., Asshoff, R. & Hammann, M. (2016), Insights into the diversity of attitudes concerning evolution and creation: a multidimensional approach. *Science Education*, *100*, 673-705.

Lawshe C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563-575.

Linacre, J. M. (2005). Rasch dichotomous model vs. one-parameter logistic model. *Rasch Measurement Transactions*, *19*(3), 1032.

Linacre, J. M. (2007). How to simulate Rasch data. *Rasch Measurement Transactions*, *21*(3), 1125.

Linacre, J. M. (2017). Teaching Rasch measurement. *Rasch Measurement Transactions*, *31*(2), 1630-1631.

Mazur, A. (2004). Believers and disbelievers in evolution. *Politics Life Science*, *23*(2), 55-61.

Nehm, R. H., & Schonfeld, I. S. (2007). Does increasing biology teacher knowledge of evolution and the nature of science lead to greater preference for the teaching of evolution in schools? *Journal of Science Teacher Education*, *18*(5), 699-723.

Pobiner, B. (2016). Accepting, understanding, teaching, and learning (human) evolution: obstacles and opportunities. *Yearbook of Physical Anthropology*, *159*, S232-S274.

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

Romine, W. L., Walter, E. M., Bosse, E., & Todd, A. N. (2016). Understanding patterns of evolution acceptance—a new implementation of the Measure of Acceptance of the theory of evolution (MATE) with Midwestern university students. *Journal of Research in Science Teaching*, *54*(5), 642-671.

Rutledge, M. L., & Warden, M. A. (2000). Evolutionary theory, the nature of science and high school biology teachers: critical relationships. *The American Biology Teacher*, *62*(1), 23- 31.

Shtulman, A., & Calabi, P. (2012). Cognitive constraints on understanding and acceptance of evolution. In K. S. Rosengren, S. K. Brem, M. E. Evans, & G. M. Sinatra (Eds.), *Evolution challenges: integrating research and practice in teaching and learning about evolution* (pp. 47-65). New York, NY: Oxford University Press.

Sinatra, G. M., Southerland, S. A., McConaughy, F., & Demastes, J. W. (2003). Intentions and beliefs in students' understanding and acceptance of biological evolution. *Journal of Research in Science Teaching*, *40*(5), 510-528.

Smith, M. U. (2010). Current status of research in teaching and learning evolution: II. Pedagogical issues. *Science & Education*, *19*(6–8), 539–571.

Smith, M. U., & Siegel, H. (2004). Knowing, believing, and understanding: What goals for science education? *Science & Education*, *13*, 553–582.

Smith, M.U., & Siegel, H. (2016). On the relationship between belief and acceptance of evolution as goals of evolution education. *Science & Education*, *25*(5-6), 473-496.

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*, 1289-1315.

Southerland, S. A., & Sinatra, G. M., (2005). The shifting roles of acceptance and dispositions in understanding biological evolution. In S. Alsop (Ed.), *Beyond Cartesian dualism: Encountering affect in the teaching and learning of science* (pp. 69-78). Dordrecht: Springer.
Tennant A., & Conaghan P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied and what one should look for in a Rasch paper. *Arthritis & Rheumatism*, *57*, 1358–1362.
Wagler, A., & Wagler, R. (2013). Addressing the lack of measurement invariance for the Measure of Acceptance of the Theory of Evolution. *International Journal of Science Education*, *35*(13), 2278–2298.
Wang W.C. (2008). Assessment of differential item functioning. *Journal of Applied Measurement*, *9*, 387–408.

APPENDIX A
WRIGHT MAP FOR ATET AND MATE ITEMS

```
 3      .##### +
           .# |
         .### |
         #### S|
           ## |
        .#### |
 2         .## +
        .#### | PropGAENE
         .### |
      .###### |
        .#### |T
      .###### | PropGAENE  PropGAENE
 1      .##### M+ PropGAENE
   .######### | ATET16  ATET18
   .######### |S ATET19  GAENE6
  ############ | GAENE9  PropGAENE  PropGAENE  PropGAENE
 ############# | GAENE1
    .######## | GAENE1  GAENE1  MATE16
 0    ######## +M ATET1   ATET3   MATE11  MATE12  MATE13  MATE20  MATE3
    .######## | ATET6   ATET7   ATET8   MATE1   MATE18
     ######## | GAENE1  GAENE2  MATE8   PropGAENE
    .######## S| GAENE1  MATE5   PropGAENE
        .#### |S GAENE1  GAENE4  PropGAENE
          .## | GAENE3  GAENE8
-1          ## + GAENE5
          ### | GAENE7
          .## |T PropGAENE
            . |
           ## |
            . |
-2           . T+
            . |
            . |
```

APPENDIX B
TWO-FACTOR PRINCIPAL AXIS FACTOR ANALYSIS ROTATED SOLUTION

	Factor	
	1	2
ATET1	.824	
MATE1	.807	
ATET3	.788	
MATE3	.779	
PropGAENE1	.779	
ATET8	.778	
MATE 20	.773	
ATET6	.773	
ATET7	.771	
PropGAENE5	.762	
GAENE14	.762	
GAENE12	.757	
GAENE11	.757	
GAENE10	.750	
ATET16	.743	
GAENE9.	.740	
MATE18	.739	
MATE8	.739	
MATE12	.733	
PropGAENE3	.729	
GAENE 4	.715	
MATE16	.713	
ATET18	.700	
GAENE6	.698	
GAENE13	.695	
PropGAENE4	.691	
PropGAENE2	.681	
MATE13	.681	
GAENE 2	.671	
ATET19	.670	
MATE11	.664	
PropGAENE11	.659	
PropGAENE9	.659	
PropGAENE6	.627	
PropGAENE8	.616	
GAENE1	.598	
MATE5	.578	
GAENE5	.537	
GAENE8	.533	
GAENE7	.528	
GAENE 3	.514	
PropGAENE10	.485	

PropGAENE7	.325	
ATET13		.779
ATET12		.773
ATET15		.759
MATE14		.747
MATE10		.736
ATET4		.728
ATET9		.713
ATET2		.712
ATET11		.706
MATE6		.689
MATE4		.681
ATET20		.664
MATE15		.662
MATE9		.646
ATET5		.628
MATE17		.623
MATE7		.622
MATE19		.582
MATE2		.574
ATET14		.573
ATET10		.511

APPENDIX C
WRIGHT MAP FOR COMBINED MEASURES

```
3       .#####  +
          .#  |
         .###  |
         #### S|
           ##  |
        .####  |
2          .##  +
        .####  |  PropGAENE
         .###  |
      .######  |
        .####  |T
      .######  |  PropGAENE  PropGAENE
1       .##### M+  PropGAENE
    .#########  |  ATET16  ATET18
    .#########  |S ATET19  GAENE6
  ############  |  GAENE9  PropGAENE  PropGAENE  PropGAENE
 #############  |  GAENE1
     .########  |  GAENE1  GAENE1  MATE16
0     ########  +M ATET1   ATET3   MATE11  MATE12  MATE13  MATE20  MATE3
     .########  |  ATET6   ATET7   ATET8   MATE1   MATE18
      ########  |  GAENE1  GAENE2  MATE8   PropGAENE
     .######## S|  GAENE1  MATE5   PropGAENE
        .####  |S GAENE1  GAENE4  PropGAENE
          .##  |  GAENE3  GAENE8
-1          ##  +  GAENE5
          ###  |  GAENE7
          .##  |T PropGAENE
            .  |
           ##  |
            .  |
-2            . T+
            .  |
            .  |
```

APPENDIX D
FINAL ADJUSTED GAENE MEASURE

Instructions: For the following items, please indicate your agreement/disagreement with the given statements using the following scale:
1=Strongly disagree, 2=disagree, 3=I don't know/no opinion, 4=Agree, 5=Strongly agree

Original GAENE 2.0 items

1. Most living things have some very basic similarities.
2. Everyone should understand evolution.
3. It is important to let people know about how strong the evidence that supports evolution is.
4. ~~Some parts of evolution theory could be true~~. (Item removed following analysis)
5. Evolutionary theory applies to all plants and animals, including humans.
6. People who plan to become biologists need to understand evolution.
7. I would be willing to argue in favor of evolutionary in a public forum such as a school club, church group, or meeting of public school parents.
8. Simple organisms such as bacteria change over time.
9. Nothing in biology makes sense without evolution.
10. Understanding evolution helps me understand the other parts of biology.
11. Understanding evolution helps me understand the other parts of biology.
12. Evolution is a good explanation of how humans first emerged on the earth.
13. Evolution is a scientific fact.
14. Evolution is a good explanation of how new species arise.

Added Items

1. All evidence supports the claim that evolution is true.
2. All species can be traced back to a single ancestor.
3. Evolution is a fact.
4. Evolution is the most important theory devised by man
5. I would bet my life on the claim that evolution is true
6. Understanding evolution has changed my life
7. Evolution explains how bacteria that are resistant to an antibiotic can arise in a population exposed to that antibiotic.
8. Evolution explains how careful breeding can produce members of a species that look different from their ancestors.
9. Small changes can occur in a species over time.

Note: No order of presentation is implied. The GAENE is designed such that the items are presented to participants in randomized order.