RESEARCH

Validity evidence for the translation and cultural adaptation of the measure of acceptance of the theory of evolution in Español (MATE-E)

Ángel Eduardo Pérez-Vega[1,5] , Riccardo Papa[2,3,4] and Michelle Borrero[2,5]*

Abstract

Background Evolution is the foundation for understanding life's diversity and interconnectedness. Acceptance of the theory of evolution is correlated with its effective teaching and learning. The Measure of Acceptance of the Theory of Evolution (MATE) is a widely used tool for assessing this acceptance; however, it requires adaptation and validity evidence for application in new linguistic and cultural contexts. This study aims to validate the Spanish adaptation of MATE (MATE-E) for Spanish-speaking high school students and biology teachers.

Results Evidence of content validity, response process, internal structure, relationship with other variables, and testing consequences supports the MATE-E's suitability for Spanish-speaking, Puerto Rican high school students and teachers. Analysis of the instrument's structure through exploratory factor analysis identified five factors. The instrument also shows strong internal consistency (Cronbach's α = 0.879). Additional evidence on the instrument's relationship with other translations or adaptations of the MATE supports the instrument validity for the intended construct. Additional data from teachers' pre- and post-test assessments following a professional development program affirms the MATE-E's cultural sensitivity and construct validity.

Conclusions The current study provides evidence for the adaptation, reliability, and validity of the MATE-E, supporting its use in research and evaluation among Puerto Rican Spanish-speaking secondary students and biology teachers.

Keywords Evolution, Evolution acceptance, Evolution education, Measure of acceptance of the theory of evolution, Instrument cultural adaptation, Instrument translation, Instrument validation

*Correspondence:
Michelle Borrero
michelle.borrero@upr.edu
[1] College of Education, University of Puerto Rico, Río Piedras Campus, San Juan, Puerto Rico
[2] Department of Biology, College of Natural Sciences, University of Puerto Rico, Río Piedras Campus, San Juan, Puerto Rico
[3] Molecular Science Research Center, University of Puerto Rico, San Juan, Puerto Rico
[4] Comprehensive Cancer Center, University of Puerto Rico, San Juan, Puerto Rico
[5] Center for Science and Math Education Research, University of Puerto Rico, Rio Piedras Campus, San Juan, Puerto Rico

Introduction

Evolution is a unifying principle in biology which provides a framework to understand the diversity of life (Dobzhansky 1973). Teaching evolution prepares students with the knowledge and critical thinking skills to understand biological processes and the connections that exist between all living things. Yet, despite its significance, the teaching of evolution often encounters resistance, misconceptions, and a lack of acceptance. These issues can impede students' comprehension and

Pérez-Vega *et al. Evolution: Education and Outreach*   (2025) 18:2

Page 2 of 20

educators' willingness to teach this critical topic thoroughly (Harms and Yarden 2023; Nehm and Schonfeld 2007).

One measure shown to influence the teaching and learning process of evolution is acceptance of the theory of evolution. There is ample evidence that the construct of evolution acceptance is influenced by culture, gender, education level, and religion, among other factors (Barnes et al. 2024; Romine et al. 2018). Due to the relevance of this construct in education, there is ample literature on its measurement, although it is not until recently that a consensus definition of evolution acceptance has been reached (Barnes et al. 2024). At present, evolution acceptance is defined as "agreeing that evolution is valid and the best explanation from science for the unity and diversity of life on Earth, which includes speciation, the common ancestry of life, and that humans evolved from non-human ancestors" (Barnes et al. 2024, p. 16). Teacher professional development (PD) programs that strengthen content knowledge and confidence to teach the subject effectively are instrumental to address and close the gaps present in the process of teaching and learning evolution (Friedrichsen et al. 2016). Measuring the effectiveness of PD activities depends on reliable and valid instruments that can accurately measure whether the PD objectives have been achieved in terms of content knowledge, skills, or attitudes of the participants. When working with participants who have different language and culture this task may involve the development of new valid and contextually relevant instruments, or that instruments that are already available are adapted to the participants' language and culture (Beniermann et al. 2023; Bravo 2003; Bravo-Vick et al. 2019; Peña, 2007). Both strategies require a strict methodology to gather validity evidence for the instrument. However, the adaptation process has the advantage that results can be compared with those available in literature.

As part of the education and outreach component of the project *Genomic Logic Underlying Morphological Adaptive Divergence* (NSF #1736026), we designed a PD program for high school biology teachers in Puerto Rico. The objective was to improve teachers' ability to teach evolution-related concepts effectively by providing workshops on evolution content knowledge, using evidence-based practices, and offering research experiences. One of the variables that we are interested in measuring is the impact of this PD on teachers' and students' acceptance of the theory of evolution. At the time of designing the PD experience, no instruments were identified in Spanish that measure this variable. As an alternative to creating new materials, the extant literature can be reviewed to determine whether there are content-appropriate instruments, notwithstanding any language barrier (Bravo-Vick et al. 2019). To assess acceptance of evolution, we considered three already available instruments: GAENE (Smith et al. 2016), I-SEA (Nadelson and Sutherland 2012), and MATE (Rutledge and Warden 1999). After careful evaluation, we selected the MATE to address the need for a culturally sensitive, valid instrument in participants' native language (Bravo 2003; Bravo-Vick et al. 2019) and undertook its translation and cultural adaptation into Spanish. It should be noted that a new instrument, the MATE 2.0, has been developed to align with the updated consensus definition of acceptance of evolution (Barnes et al. 2022), addressing many of the issues found in the original MATE (this updated version was not published at the time this research was conducted).

The Measure of Acceptance of the Theory of Evolution (MATE) was selected based on several criteria. First, it is the most widely used instrument for assessing acceptance of evolution, allowing results to be easily compared with those of previous studies (Barnes et al. 2024, 2022). Additionally, its intended use with high school biology teachers aligns with the original target population of the instrument (Rutledge and Warden 1999), and it has been applied in numerous international studies (Athanasiou et al. 2011; Lammert 2012; Tekkaya et al. 2012; Rutledge and Sadler 2007).

As we strive to be more inclusive, it is crucial to use instruments that are relevant and appropriate for the target population. For research purposes, having a linguistically and culturally adapted tool is essential. Therefore, the goal of this study is to gather validity evidence for the translation and adaptation of the MATE into Spanish, with consideration of Puerto Rican culture. The resulting instrument, developed through this work, will be referred to as MATE-E (i.e., MATE in Español).

Background

Most instruments published in the literature are available in English. Although developing a new instrument in a different language can provide valuable data, it limits the potential for comparison and generalization across studies. A more practical approach is to adapt existing, well-established instruments to the language and culture of the target population. However, this process presents challenges: direct translations can alter the meaning or intent of test items, and cultural differences can lead to varying interpretations or make certain items irrelevant. These factors must be carefully managed to maintain the instrument's validity and reliability in measuring the intended construct (Peña, 2007). As such, instrument adaptation requires a rigorous process beyond simple translation to prevent issues of validity, reliability, and

cultural bias (Peña, 2007; International Test Commission 2017). The goal is to create an adapted instrument that accurately and equivalently measures the target construct.

The International Test Commission (2017) has developed guidelines for test translation and adaptation to systematically gather validity evidence for an instrument. Validity refers to the extent to which evidence and theory support the intended interpretations of test scores for the proposed purpose (AERA, APA, NCME 2014; Messick 1989). The adaptation process collects validity evidence from several sources: content, response process, internal structure, instrument reliability, relationship with other variables, and results or consequences of testing (AERA, APA, NCME 2014; Beniermann et al. 2023; Creswell 2012).

Content validity or linguistic and cultural equivalence

Content validity refers to the extent to which a test accurately measures the intended construct. In an adaptation process it evaluates if the test items measure the same or significantly similar constructs in the population of interest. Thus, the process requires native translators as it considers the target language and culture. The recommended steps involved in the process are a forward translation, back- translation (translation back to the source language), and an expert panel review to resolve discrepancies and refine wording. The goal is to generate an instrument that is functionally equivalent to the original.

Response process

It refers to the cognitive actions of participants while completing an instrument. Data gathered through this process is used to adjust the instrument items or instructions to ensure comprehension and cultural relevance. Cognitive interviews or think-aloud protocols are used to ensure that test instructions and items have consistent meaning for the target population. This process provides evidence for the functional equivalence as it confirms that participants' responses correspond to the intended construct, rather than misunderstandings, ensuring the items function as expected.

Internal Structure and Instrument reliability

Internal structure validity evidence aims to evaluate the relationship between the items on the test and the theoretical structure of the intended construct. It assesses whether the items group together consistently with the expected dimensions or factors of the construct. Whereas instrument reliability examines the consistency of an instrument's scores across multiple measurements under consistent conditions. This evidence is gathered

by empirical analyses, collecting pilot data from participants that have similar characteristics to the targeted population. Data is analyzed through various statistical tests to determine the structure of the test (e.g., factor analysis) and its internal consistency (e.g., Cronbach's alpha). Table 1 summarizes data from reliability analyses conducted across several publications of the MATE.

Relationship with other variables

Validity evidence based on relationships with other variables assesses whether these relationships are consistent with the intended construct and interpretation of test scores. The process ensures that the instrument is measuring the intended construct in alignment with prior research. The evidence can be gathered from outcomes the test is expected to predict or from its relationship with other measures of similar or different constructs, that confirm that the instrument's results are meaningful and relevant to the research field.

Consequences of testing

Empirical evidence is gathered from instrument administration to the intended population. Data is analyzed to evaluate the appropriateness of the test score interpretations. It is an important source of evidence to determine intended and unintended claims from instrument administration.

The development of the MATE-E through this rigorous process will produce an instrument in Spanish that can be used to research the acceptance of the theory of evolution in high school teachers and students and compare results to what is already established in the literature. Our work also serves as a reference to others that aim to adapt and generate a culturally sensitive instrument for research.

Methods and materials

Design and data collection for the translation, adaptation, and validation of the MATE into Spanish.

The International Test Commission's (2017) guidelines for test translations and adaptations were followed. Pre-condition guidelines were used to evaluate which of the available tests was the most appropriate for the project's goals. Test development guidelines were followed by applying the recommendations of AERA, APA, NCME (2014), as well as Creswell (2012), to gather evidence for content validity, response processes, internal structure, relationships with other variables, and the consequences of testing. Confirmation guidelines were followed through sampling of a similar group to the target audience and using statistical testing and evidence supporting the norms, reliability, and validity of the MATE-E. Administration

Pérez-Vega *et al. Evolution: Education and Outreach*   (2025) 18:2

Page 4 of 20

Table 1 Comparison of the internal structure of several published versions of the MATE

Version	Country	Items	Cronbach's alpha	Summary of translation, adaptation, and validation process (Factors, % of variance, and construct naming whenever available)
Original MATE (Rutledge & Warden 1999)	United States of America	20	.98	One factor extracted (71.6%). The authors stated that items were developed following seven fundamental concepts stemming from the theory of evolution and categorized the acceptance of the evolutionary theory as the factor
MATE adaptation for undergraduate students (Rutledge & Sadler 2007)	United States of America	20	.94	No mention of factor analysis or variance explained. The authors classified items into six concepts and explained that a high Cronbach's alpha indicates that the items are highly related and that they measure the same psychological construct: acceptance of the evolutionary theory. Concepts and items: 1: Process of evolution (items 1, 9, 18, 19) 2: Scientific validity of evolutionary theory (items 2, 10, 12, 13, 14, 20) 3: Evolution of humans (items 3, 15) 4: Evidence of evolution (items 4, 6, 8, 16) 5: Scientific community's view of evolution (items 5, 17) 6: Age of the Earth (items 7, 11)
MATE of Greek education students (Athanasiou et al. 2011)	Greece	20	.79	The authors explained that the MATE was translated and adapted to Greek by Athanasiou. No additional description of this process was provided. Thus, there is no information on factor analysis or total variance explained. Authors used Rutledge and Sadler's (2007) concepts for the item categorization on the Greek translation and adaptation. Cronbach's alpha was reported for the entire instrument and for each subscale 1: Process of evolution (α = .71) 2: Scientific validity of evolutionary theory (α = .75) 3: Evolution of humans (α = .71) 4: Evidence of evolution (α = .69) 5: Scientific community's view of evolution (α = .83) 6: Age of the Earth (α = .65)
MATE translation and adaptation to German (Lammert 2012)	Germany	16	.83	Four factors were extracted (48%). The author eliminated three items due to low factor loadings and eliminated an additional item due to high factor loadings on two factors. Cronbach's alpha was reported for the entire instrument and for each subscale. Author categorized factors as follows: 1: Positive statements about the theory of evolution (α = .75) 2: Negative statements about the theory of evolution (α = .71) 3: Time dimensions (α = .70) 4: Development (α = .64)
MATE translation and adaptation to Turkish (Tekkaya et al. 2012)	Türkiye	18	.94	The authors summarized a translation, adaptation, and validation process previously presented via a poster in a conference. Authors eliminated two items corresponding with Rutledge and Sadler's (2007) concept of the age of the Earth due to low factor loadings. No additional information on factor analysis, total variance explained, or factor categorization is discussed

guidelines were followed by using a consistent administration process. Scoring and interpretation guidelines were followed through extant data to determine if original scoring and interpretation of scores was appropriate for the translated and adapted instrument. Lastly, documentation guidelines were followed to keep track of every investigative task that led to the completion of the translation, adaptation and validity evidence gathering process of this instrument. Figure 1 provides an overview of the design and data collection process that encompasses this entire work.

Sources of evidence for the validity of the MATE-E

In addition to following the International Test Commission's (2017) guidelines for the overall process of test translation and adaptation, a more specific process was carried out to gather the evidence that would serve as the backbone of the translation, adaptation, and validation process. The Measure of Acceptance of the Theory of Evolution (Rutledge and Warden 1999) was translated, validated, and adapted from English to Spanish (MATE-E) following guidelines of AERA, APA, NCME (2014), and Creswell's (2012) to gather evidence for content validity, response process, internal structure validity, instrument reliability, its relationship with other variables, and the consequences of testing with it.

Content validity evidence

Two methods were used to gather content validity evidence. The first one was through the translation/back-translation technique for semantic equivalence (Behling and Law 2000). The translation/back-translation process was carried out as follows: (1) a bilingual, Puerto Rican, native Spanish speaker and fluent English speaker with education expertise translated the instrument from English to Spanish; and (2) a bilingual, native English speaker and fluent Spanish speaker with biology expertise translated the instrument from Spanish back to English. All three versions of the instrument (original, translation, and back-translation) were submitted for evaluation to an expert panel.

The expert panel consisted of four bilingual experts (DeVellis 2012), all of them experts in biology. Two of them are also experts in education and one of them is an expert in instrument design. The experts received an orientation prior to the task of evaluating each item in terms of accuracy and clarity of its translation using an item evaluation sheet that contained all three versions of each item, a scale to evaluate accuracy, a scale to evaluate clarity, and a section to provide comments for each item. After receiving each expert's evaluation, an online meeting was scheduled to discuss the feedback and update the translated version for use during the response process evidence-gathering phase.

Response process evidence

Five high school students who had previously taken a biological sciences course and were enrolled in a high school biology course agreed to participate in this process. They answered the MATE-E instrument and then took part in a focus group in which they were asked about the test's content, as well as their thought process when answering each item. Students also provided recommendations to further improve items' accuracy and clarity. Recommendations from this process were reviewed with the expert panel to determine which to adopt in the final version of the instrument. This final version was then used in a pilot study to gather evidence of the instrument's internal structure.

Internal structure evidence

A non-experimental quantitative methodology study with a survey research design (McMillan 2016) was used to conduct a pilot study and gather evidence for the instrument's internal structure. The pilot study was conducted with secondary-level students because this test is intended to be administered to both high school teachers and students. The sampling used was non-probabilistic and by convenience. A total of 281 8th, 10th, and 11th grade students from the University of Puerto Rico, Río Piedras Campus' Laboratory Secondary School were invited to participate. MATE-E was self-administered online. The participants' responses were anonymous, and they did not receive any incentive for their participation.

Evidence based on MATE-E's relationships to other variables

For the purposes of this validity evidence gathering process, we wanted to verify if the outcomes of the MATE-E pilot study test remain consistent with our assumptions about how the test would discriminate between grades and between genders. To verify these assumptions, independent samples *t*-tests were conducted to compare the means for data sorted by gender as well as by grade, to determine whether the MATE-E discriminates between observed groups (Cronbach and Meehl 1955). Data was compared with other studies' (Athanasiou et al. 2011; Lammert 2012; Tekkaya et al. 2012; Rutledge and Sadler 2007) approaches on gathering this kind of evidence.

Evidence based on consequences of testing and interpreting data

The MATE-E was administered to 11 biology teachers who participated in a pilot of a two-year professional development program that included workshops on natural selection, adaptation, evolution, heredity, and gene expression. Sampling of these teachers (n=11) was non-probabilistic
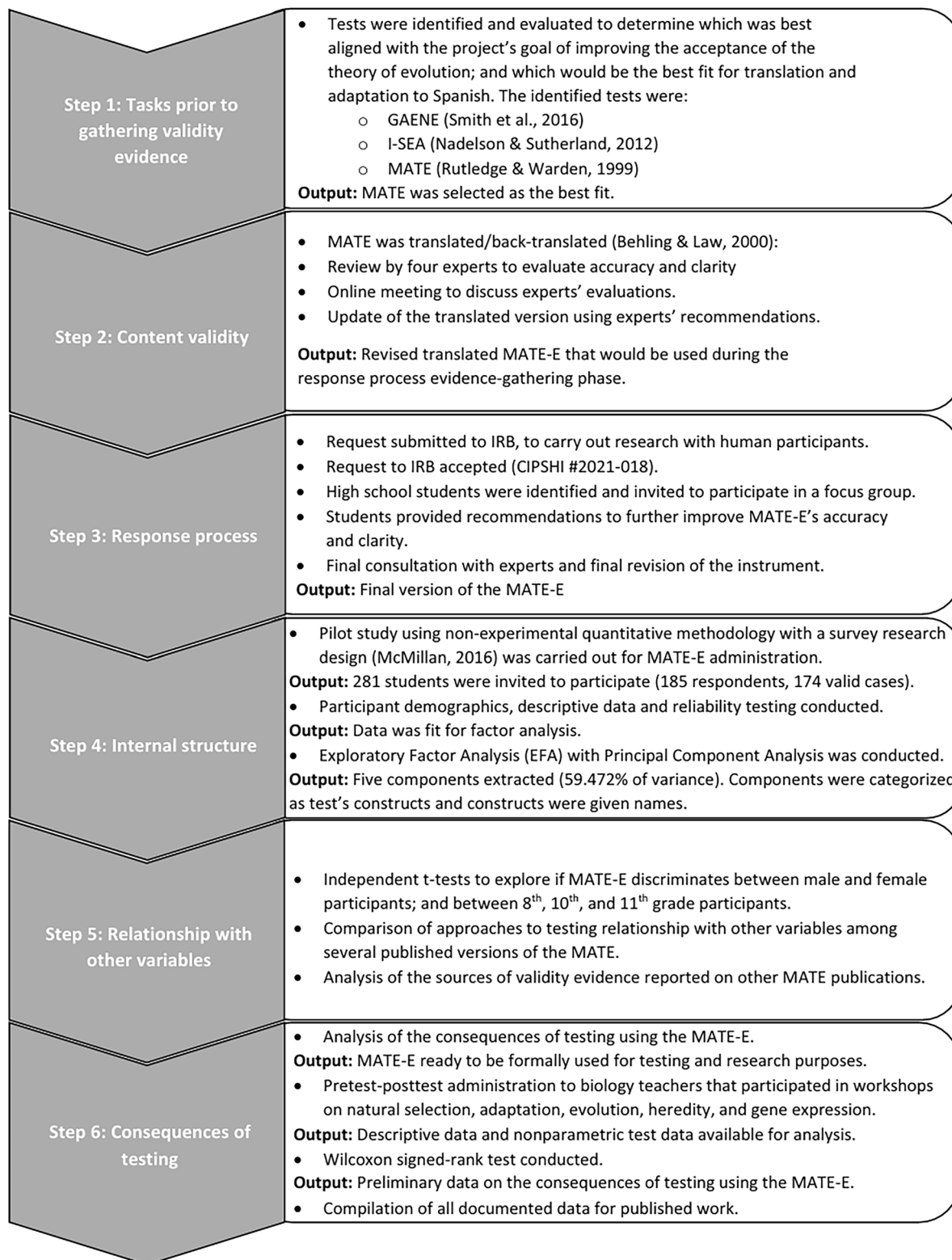

Step 1: Tasks prior to gathering validity evidence	• Tests were identified and evaluated to determine which was best aligned with the project's goal of improving the acceptance of the theory of evolution; and which would be the best fit for translation and adaptation to Spanish. The identified tests were:
	    ○ GAENE (Smith et al., 2016)
	    ○ I-SEA (Nadelson & Sutherland, 2012)
	    ○ MATE (Rutledge & Warden, 1999)
	Output: MATE was selected as the best fit.

Step 2: Content validity	• MATE was translated/back-translated (Behling & Law, 2000): • Review by four experts to evaluate accuracy and clarity • Online meeting to discuss experts' evaluations. • Update of the translated version using experts' recommendations. **Output:** Revised translated MATE-E that would be used during the response process evidence-gathering phase.
Step 3: Response process	• Request submitted to IRB, to carry out research with human participants. • Request to IRB accepted (CIPSHI #2021-018). • High school students were identified and invited to participate in a focus group. • Students provided recommendations to further improve MATE-E's accuracy and clarity. • Final consultation with experts and final revision of the instrument. **Output:** Final version of the MATE-E
Step 4: Internal structure	• Pilot study using non-experimental quantitative methodology with a survey research design (McMillan, 2016) was carried out for MATE-E administration. **Output:** 281 students were invited to participate (185 respondents, 174 valid cases). • Participant demographics, descriptive data and reliability testing conducted. **Output:** Data was fit for factor analysis. • Exploratory Factor Analysis (EFA) with Principal Component Analysis was conducted. **Output:** Five components extracted (59.472% of variance). Components were categorized as test's constructs and constructs were given names.
Step 5: Relationship with other variables	• Independent t-tests to explore if MATE-E discriminates between male and female participants; and between 8th, 10th, and 11th grade participants. • Comparison of approaches to testing relationship with other variables among several published versions of the MATE. • Analysis of the sources of validity evidence reported on other MATE publications.
Step 6: Consequences of testing	• Analysis of the consequences of testing using the MATE-E. **Output:** MATE-E ready to be formally used for testing and research purposes. • Pretest-posttest administration to biology teachers that participated in workshops on natural selection, adaptation, evolution, heredity, and gene expression. **Output:** Descriptive data and nonparametric test data available for analysis. • Wilcoxon signed-rank test conducted. **Output:** Preliminary data on the consequences of testing using the MATE-E. • Compilation of all documented data for published work.

Fig. 1 Summary of the design and data collection for the translation, adaptation, and validation of the MATE-E

and by convenience. The MATE-E was self-administered in person before and after these workshops.

Statistics

Data from the pilot study was analyzed using descriptive statistics and tested for normality using the Kolmogorov–Smirnov test. A reliability test was performed to determine Cronbach's alpha. Independent samples *t*-test were also conducted to report data according to variables of gender and grade. Prior to the Exploratory Factor Analysis (EFA), adequacy was tested using the Keiser-Meyer-Olkin Measure of Sampling Adequacy and Bartlett's Test of Sphericity. Exploratory Factor Analysis with Varimax rotation was used to reduce factors using Eigenvalue, Scree testing, factor loading, and cumulative percent of variance. Data from the 11 biology teachers was analyzed using descriptive statistics and a Wilcoxon signed-rank test was conducted to make inferences about the effect of the workshops on its participants. All statistical analyses were performed using IBM® Statistical Package for the Social Sciences (SPSS®) version 29.0.

Participants

Five high school students participated in the focus group phase to gather response process evidence. For the pilot study, 281 middle and high school students were invited to participate, of which 185 assented to their participation. All student participants attend the University of Puerto Rico, Río Piedras Campus' Laboratory Secondary School. Additionally, 11 biology teachers from secondary schools in Puerto Rico were invited to participate in pilot professional development (PD) workshops on topics related to the theory of evolution. These teachers provided data contributing to the instrument's validity evidence, specifically regarding the consequences of testing. This protocol was approved by the University of Puerto Rico Río Piedras Campus' Institutional Review Board (CIPSHI #2021-018).

Results

It is imperative to gather validity evidence for a translated and culturally sensible instrument to support its interpretation and to determine that it is being used for its intended purpose. Content, response process, and internal structure validity evidence was gathered to determine whether the MATE-E, was adequately translated and adapted for Puerto Rican Spanish-speakers. In addition, we gather evidence of the instrument's relationship with other variables and the consequences of testing.

Content validity

After the initial translation and back-translation process, the second step to gather evidence of content validity was an evaluation by experts (see Fig. 1). The experts received an evaluation checklist to determine the accuracy and clarity of the translations. Table 2 shows an example of how one of the items was evaluated by the expert panel.

After receiving all evaluations, the research team observed that there were three items (items 1, 11, and 19) with no revisions to be made and seven items (items 5, 7, 10, 12, 14, 18, and 20) with minor revisions or comments from one expert. Additionally, seven items (items 2, 3, 4, 6, 9, 13, and 15) had revisions or comments from two experts and the three remaining items (items 8, 16, and 17) had revisions and comments from three experts. No items had revisions from all four experts.

Although most items had revisions or comments from experts, most of these were related to wording and syntax issues. The experts' recommendations were mostly focused on the substitution of words or rearrangement of phrases so that the items would be clearer and more syntactically accurate. Some items (specifically, items 9, 15, and 16), however, had conceptual issues. The main conceptual issue stemmed from the use of more ambiguous terms ("de la misma forma" and "fácticos" which translate to "in the same way" and "factual").

Discrepancies such as this one were discussed by the experts in an online meeting. After the panel members evaluated the items, the meeting was scheduled to discuss their evaluations. The experts agreed on their comments on the items that had minor or moderate issues and provided revised versions for them. Additionally, they argued about how items 9, 15, and 16 needed further clarity. The experts proposed to include the word "fenotipo" (phenotype) in parentheses for items 9 and 15, and the word "verificable" (verifiable) in parentheses for item 16, to frame or contextualize the statements for potential participants.

After this meeting, an updated version of the MATE's Spanish translation (i.e., MATE-E) was generated that incorporated the panel's comments and evaluations. These items were then used for the focus groups to gather evidence of the response process. Table 3 presents an example of the changes on items after the experts' evaluation and deliberation.

Response process

A focus group was carried out to gather evidence for MATE-E's response process. Participants deemed the instrument's translated name and its instructions appropriate. They provided some recommendations to improve wording for some items and the instrument's scale. Most of these recommendations were minor and related to improving the wording on the items. Regarding the items that had the most issues during the content validity phase (items 9, 15, and 16), the focus group offered some

Pérez-Vega *et al. Evolution: Education and Outreach*   (2025) 18:2

Page 8 of 20

Table 2 Item evaluation sheet sample

Item #	Type	Statement	Translation		Clarity		Comments*
			Adequate	Inadequate	Clear	Confusing	
16	Original	Evolutionary theory is supported by factual, historical, and laboratory data			X (**E1**)		**E1**: La traducción no es 100% correcta, pero el mensaje es esencialmente el mismo. (*Translation is not 100% correct, but the message is essentially the same*)
	Translation	La teoría de la evolución está apoyada por datos históricos y de laboratorio fácticos		X (**E1**, **E2**, **E3**)		X (**E2**, **E3**)	**E2**: ¿Laboratorios fácticos? (*Factual laboratories?*)
	Back-translation	The theory of evolution is supported by historical data and laboratory results					**E3**: De la manera que se tradujo no mantiene el sentido de los tres tipos de datos incluidos en la premisa original. Sugiero: La teoría evolutiva es respaldada/apoyada por datos factuales, históricos y de laboratorio. (*The way it was translated does not maintain the meaning of the three types of data included in the original premise. I suggest: Evolutionary theory is supported by factual, historical, and laboratory data.*)

*For clarity purposes, expert (**E**) comments are translated into English in *italics*

Table 3 Example of expert evaluation of MATE-E items

Item #	Statement as evaluated by experts (Original statement in parenthesis)	Recommendations	Consensus for revised statement
1	Los organismos que existen hoy en día son el resultado de procesos evolutivos que han ocurrido durante millones de años (Organisms existing today are the result of evolutionary processes that have occurred over millions of years.)	None	Los organismos que existen hoy en día son el resultado de procesos evolutivos que han ocurrido durante millones de años
2	La teoría de la evolución no puede ser científicamente **comprobada** (The theory of evolution is incapable of being scientifically tested.)	Minor (wording issues)	La teoría de la evolución no puede ser científicamente **probada**
6	Los datos disponibles son **ambiguos (inciertos) sobre** si la evolución **ocurre en realidad** (The available data are ambiguous as to whether evolution actually occurs.)	Moderate (translation and back-translation issues)	Los datos disponibles son **ambiguos en cuanto a** si la evolución **realmente ocurre**
8	**Hay un cuerpo de datos significativos** que apoya la teoría de la evolución (There is a significant body of data which supports evolutionary theory.)	Major (translation, back-translation, and wording issues)	**Existe una cantidad significativa de datos** que apoya la teoría de la evolución
11	La edad de la tierra es de al menos 4000 millones de años (The age of the earth is at least 4 billion years.)	None	La edad de la tierra es de al menos 4000 millones de años
12	La teoría actual de la evolución es el resultado de una **sólida investigación y metodología** (Current evolutionary theory is the result of sound scientific research and methodology.)	Minor (wording issues)	La teoría actual de la evolución es el resultado de una **investigación y metodología científica sólida**
13	La teoría de la evolución genera predicciones **comprobables en relación con** las características de la vida (Evolutionary theory generates testable predictions with respect to the characteristics of life.)	Moderate (translation and wording issues)	La teoría de la evolución genera **predicciones que se pueden poner a prueba con respecto a** las características de la vida
16	La teoría de la evolución está apoyada por datos **históricos y de laboratorio fácticos** (Evolutionary theory is supported by factual, historical, and laboratory data.)	Major (translation, back-translation, conceptual, and wording issues)	La teoría evolutiva está apoyada por datos **fácticos (verificables) históricos y de laboratorio**

Issues on items and consensus revisions are written in **bold**

Pérez-Vega *et al. Evolution: Education and Outreach*   (2025) 18:2

Page 10 of 20

insights into how the items could be rewritten to provide further clarity. The focus group opted to leave the terms in parentheses for items 9 and 15. The opposite happened with item 16. The focus group proposed a revised version of the item that did not include the term in parentheses for further clarity. The recommendations given by the focus group participants for these three items can be found in Table 4.

A noteworthy recommendation from the focus group was to consider including Darwin whenever the theory of evolution is mentioned in the instrument (items 2, 4, 5, 6, 8, 10, 12, 13, 14, 16, 18, and 20). This, alongside all minor revisions and the recommendations for items 9, 15, and 16, were taken once more to the expert panel. A final online meeting with the expert panel was conducted to present the recommendations given by the focus group participants and confer whether to accept these suggestions. The expert panel accepted the minor recommendations but decided that it was not necessary to attribute the theory of evolution to Darwin in every item. Lastly, the experts accepted the focus group's recommendation for item 16 and deemed their revised versions of items 9 and 15 clear enough that the terms in parentheses could be eliminated. This last step led to the final MATE-E instrument that was used for the pilot study onwards (Table 5).

Pilot study and internal structure

A pilot study was done to gather evidence of the instrument's reliability of responses and internal consistency of the scores. The MATE-E instrument was administered to students (grades 8th, 10th, and 11th) from the University of Puerto Rico, Río Piedras Campus' Laboratory Secondary School. Since this translated and adapted instrument is intended to be administered to both high school biology teachers and students, the assumption in this case is that, if students understand the content of the instrument, teachers must also be able to understand it.

A total of 281 students were invited to participate in this pilot study, of which 185 assented to their participation, for a total response rate of 65.8%. The MATE-E was self-administered, and it was done online. The total response rate is consistent with Schonlau et al. (2001) estimates of total response rates for online self-administered surveys. Out of the 185 respondents, only 174 were valid for analysis. Eleven respondents submitted answers with missing data. The valid responses make up 94% of all responses and 61.9% of all invited participants. A description of the participants of this pilot study is summarized below.

Table 6 summarizes the descriptive statistics for the MATE-E pilot study. Data is shown for the entire dataset (overall mean score was 76.46, $SD = 13.147$)

as well as two subsets classified by gender and by grade. For gender, the mean score for males was 76.16 ($SD = 13.848$) and for females it was 76.73 ($SD = 12.560$). For grade levels, the mean score for 8th graders was 71.50 ($SD = 15.254$), while 10th graders mean score was 79.42 ($SD = 11.137$) and 11th graders had a mean score of 77.92 ($SD = 12.128$). A Kolmogorov–Smirnov test was performed to measure normality on the entire dataset. Test results show that data is normally distributed, $D(174), = 0.065$, $p = 0.066$.

In addition to the descriptive reports for the pilot study administration, a reliability analysis was performed to check for MATE-E's internal consistency. Since the test's items are scored as continuous variables, the appropriate test for reliability is Cronbach's coefficient alpha (1984). Results show that MATE-E's Cronbach's alpha based on standardized items is 0.879. Additionally, the Kaiser–Meyer–Olkin Measure of Sampling Adequacy index was calculated. Data shows that the sample used was adequate (KMO = 0.857). Similarly, a Bartlett's Test of Sphericity was performed, and results were significant ($p < 0.001$).

Exploratory Factor Analysis for MATE-E

An Exploratory Factor Analysis (EFA) with Principal Component Analysis was conducted to identify MATE-E's test factors. Five components with Eigenvalues greater than 1 were identified. These components represent 59.472% of the total variance explained. Table 7 summarizes the total variance explained by the extracted components.

Figure 2 shows a Scree plot with the Eigenvalues from each component as extracted from the Principal Component Analysis.

The Principal Component Analysis additionally informed how the MATE-E's items correlated with each of the five components. The rotation method used for this analysis was Varimax with Kaiser Normalization and the test was adjusted to show correlations greater than 0.40. Table 8 shows that eight items strongly correlate with the first component, five items with the second, four items with the third, two items with the fourth, and one item with the fifth.

Two of these items have strong correlations with more than one component. Item number 2 correlates strongly with the first and second component, while item number seven correlates strongly with the second and third component. However, item 2 has a stronger correlation to the first component, while item number 7 has a stronger correlation with the third component. Therefore, items 2 and 7 were grouped into the first and third component, respectively.

Pérez-Vega *et al. Evolution: Education and Outreach*   (2025) 18:2

Page 11 of 20

Table 4 Sample of focus group's recommendations

Item and original statement	Statement evaluated by focus group	Recommendations	Statement as recommended
9: Organisms exist today in essentially the same form in which they always have	Los organismos existen hoy en día **esencialmente de la misma forma** (fenotipo) en que siempre han existido	1. Eliminar "esencialmente" (*Eliminate "essentially"*) 2. Sustituir "de la misma forma" por "con la misma apariencia física (*Substitute "in the same way" for "with the same physical appearance"*)	Los organismos existen hoy en día **con la misma apariencia física** (fenotipo) en que siempre han existido *Organisms exist today with the same physical appearance (phenotype) as they have always existed*
15: Humans exist today in essentially the same form in which they always have	Los humanos existen hoy en día **esencialmente de la misma forma** (fenotipo) en que siempre han existido	1. Eliminar "esencialmente" (*Eliminate "essentially"*) 2. Sustituir "de la misma forma" por "con la misma apariencia física (*Substitute "in the same way" for "with the same physical appearance"*)	Los humanos existen hoy en día **con la misma apariencia física** (fenotipo) en que siempre han existido *Humans exist today with the same physical appearance (phenotype) as they have always existed*
16: Evolutionary theory is supported by factual, historical, and laboratory data	La **teoría evolutiva** está apoyada por **datos fácticos** (verificables) históricos y de laboratorio	1. Sustituir "teoría evolutiva" por "teoría de la evolución" (*Substitute "evolutionary theory" for "the theory of evolution"*) 2. Sustituir "datos fácticos (verificables)" por "evidencia." (*Substitute "factual (verifiable) data" for "evidence"*) 3. Sustituir "de laboratorio" por "científica." (*Substitute "from laboratory" for "scientific"*)	La **teoría de la evolución** está apoyada por **evidencia** histórica y científica *The theory of evolution is supported by historical and scientific evidence*

Issues on items and recommendations are written in **bold**

*For clarity purposes, literal translations of recommendations and statements as recommended were written in *italics*

Table 5 Pilot study participant demographics

Groups	Male	Female	Total
8th grade	27	25	52
10th grade	6	13	19
11th grade	53	61	114
Total groups: 3	Total male: 86	Total female: 99	Grand total: 185*

*There were only 174 valid cases among the 185 participants; 11 cases had missing data

It is also important to point out that only one item correlates strongly with component 5. Despite this, the team decided not to discard this component, nor this item. The item possesses the strongest correlation among all observed data.

Figure 3 shows a path diagram for MATE-E's Exploratory Factory Analysis. To further help visualize this diagram, the Eigenvalues, rotation sums of squared loadings and percent of variance explained are shown for each factor. Additionally, factors are named (names are also translated to English in *italics* for clarity) and the correlation for each item is also shown. Rutledge and Sadler's (2007) concepts were considered when naming these factors.

Upon conducting Exploratory Factor Analysis, items were grouped into five components. After reviewing how the items were grouped within the components, as well as reference literature using the MATE instrument, each factor was named as follows: Validez del proceso de evolución (*Validity of the evolutionary process*), Evidencia de la evolución (*Evidence of evolution*), Evolución de los humanos y otras especies (*Evolution of humans and other species*), Punto de vista de la comunidad científica en torno a la evolución (*Scientific community's view of evolution*), and Edad de la Tierra (*Age of the Earth*). These factors, its corresponding items, and some sample items are shown on Table 9. The final translated version of the MATE-E instrument is available on Additional File 1.

MATE-E's relationship with other variables

Independent *t*-tests were conducted on data sorted by gender and sorted by grade to identify if the MATE-E discriminated between any of the groups. Our working assumption for the data sorted by grade is that 10th and 11th grade participants will score better than 8th grade participants due to biology content knowledge increasing with grade progression. On the other hand, we do not expect any significant differences between male and female participants as we did not find information on how evolution acceptance differs between genders on high school students. Therefore, we do not assume that the MATE-E has the ability to distinguish the evolution acceptance of male and female respondents. On data sorted by gender, female participants ($M=76.73$, $SD=12.560$) had only a marginally higher score than male participants ($M=76.16$, $SD=13.848$). The means were not statistically different between males and

Table 6 Descriptive statistics for pilot study administration (whole dataset, by gender, and by grade)

Whole dataset			Classified by gender		Classified by grade		
			Male	**Female**	**8th**	**10th**	**11th**
N	Valid	174	82	92	44	19	111
	Missing	11	4	7	8	0	3
Mean		76.46	76.16	76.73	71.50	79.42	77.92
95% Confidence Interval for Mean	Lower Bound	74.49	73.12	74.13	66.86	74.05	75.64
	Upper Bound	78.43	79.20	79.33	76.14	84.79	80.20
5% Trimmed Mean		77.15	77.07	77.25	72.27	79.47	78.51
Std. Error of Mean		.997	1.529	1.309	2.300	2.555	1.151
Median		77.50	77.00	79.00	72.50	81.00	79.00
Std. Deviation		13.147	13.848	12.560	15.254	11.137	12.128
Variance		172.851	191.765	157.761	232.674	124.035	147.093
Skewness		-.785	-.947	-.587	-.759	-.155	-.665
Std. Error of Skewness		.184	.266	.251	.357	.524	.229
Kurtosis		.968	1.650	.082	1.160	-.721	.200
Std. Error of Kurtosis		.366	.526	.498	.702	1.014	.455
Range		71	71	59	71	40	58
Minimum		28	28	40	28	59	40
Maximum		99	99	99	99	99	98

Pérez-Vega *et al. Evolution: Education and Outreach*   (2025) 18:2

Page 13 of 20

Table 7 MATE-E total variance explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.282	31.408	31.408	6.282	31.408	31.408	3.925	19.624	19.624
2	2.024	10.120	41.528	2.024	10.120	41.528	2.934	14.670	34.294
3	1.482	7.410	48.938	1.482	7.410	48.938	2.462	12.311	46.605
4	1.099	5.495	54.433	1.099	5.495	54.433	1.467	7.336	53.941
5	1.008	5.039	59.472	1.008	5.039	59.472	1.106	5.532	59.472
6	.857	4.285	63.757						
7	.822	4.110	67.868						
8	.784	3.918	71.785						
9	.749	3.745	75.530						
10	.696	3.482	79.011						
11	.621	3.103	82.115						
12	.600	3.002	85.117						
13	.565	2.823	87.940						
14	.463	2.313	90.253						
15	.398	1.991	92.244						
16	.390	1.948	94.192						
17	.320	1.600	95.791						
18	.303	1.517	97.309						
19	.294	1.468	98.777						
20	.245	1.223	100.000						

Extraction Method: Principal Component Analysis
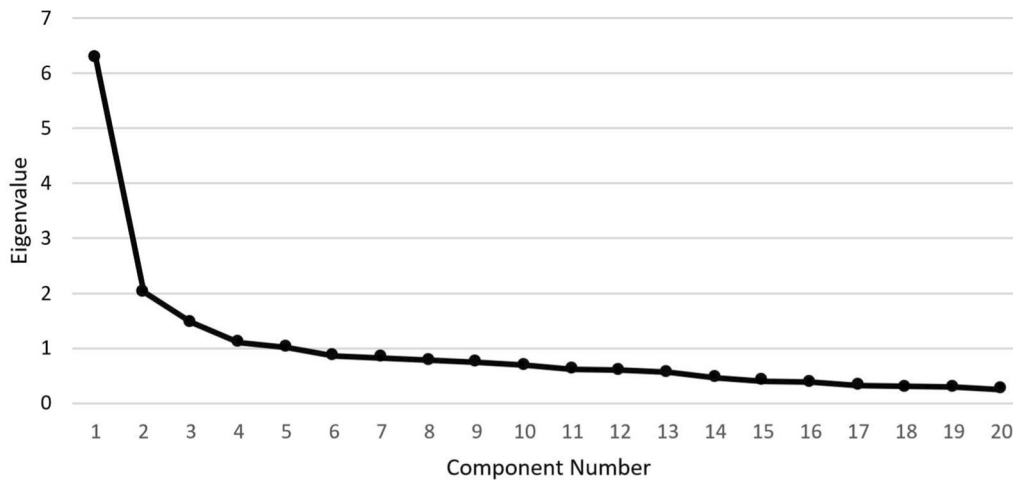
Fig. 2 Eigenvalues for each MATE-E component

females, $t(172) = -0.285$, $p = 0.776$. Data is summarized in Table 10.

Independent *t*-tests were also performed to compare means between the following grade pairs: 8th and 10th, 8th and 11th, and 10th and 11th (see Table 11). The effect size of the difference between 8th grade and 10th grade participants, measured by Cohen's *d*, was $d = -0.559$, which indicates a medium effect. On average, 8th grade students scored eight points lower than 10th grade students. On the other hand, the effect size of the difference between 8th grade and 11th grade participants was $d = -0.491$, which indicates a small effect. When compared to each other, 8th grade participants scored, on average, close to 6.5 points lower than 11th grade participants. Differences between 10th and 11th grade were not statistically significant.

Table 8 MATE-E rotated component matrix

Item #	Component				
	1	2	3	4	5
3	.727				
20	.724				
16	.694				
18	.658				
1	.654				
8	.625				
12	.563				
2	.515	.440			
10		.790			
4		.727			
6		.657			
17		.515			
14		.448			
9			.828		
15			.753		
19			.682		
7		.453	.455		
13				.733	
5				.653	
11					.853

Extraction Method: Principal Component Analysis

Rotation Method: Varimax with Kaiser Normalization

Rotation converged in 9 iterations

Rutledge and Warden ([1999](#)) explained that they developed 20 items after establishing the following seven fundamental concepts of the theory of evolution: the processes of evolution, the available evidence of evolutionary change, the ability of evolutionary theory to explain phenomena, the evolution of humans, the age of the earth, the independent validity of science as a way of knowing, and the current status of evolutionary theory within the scientific community (p. 14). As shown in Table 12, on a subsequent study for its adaptation for university students, Rutledge and Sadler ([2007](#)) specified six concepts and detailed the items that addressed each of those concepts. In most cases of translation, adaptation, and validation, these concepts were taken into consideration to name the constructs for each version of the MATE. A similar analysis regarding validity evidence shows that all instruments differ in the data reported. Table 12 summarizes sources of validity evidence reported in each article.

Consequences of testing using the MATE-E

The MATE-E test was translated to Spanish, adapted, and validity evidence was gathered with Puerto Rican Spanish-speaking high school students. This version of the instrument is intended to be used for both students and biology teachers, based on the assumption that biology teachers will be at least as familiar with the language and content knowledge of the instrument. Thus, the process of gathering evidence for the MATE-E and intending its use for students and teachers has the advantage of allowing follow-up studies that can help measure the change both students' and biology teachers' acceptance of the theory of evolution. However, this instrument alone does not have a direct consequence on a participant's acceptance of the theory of evolution. The interpretation of scores should be subject only to the results of administration.

To this extent, 11 biology teachers participated in workshops on natural selection, adaptation, evolution, heredity, and gene expression. The MATE-E was administered before and after a pilot two-year professional development (PD) program to test the hypothesis that teachers' MATE-E scores would increase after completing the PD activities. Table 13 shows the descriptive data for both administrations.

A Wilcoxon signed-rank test was conducted to determine whether the data is statistically significant. This test is the most appropriate due to sample size. The test indicated that the biology teachers' acceptance of the theory of evolution was significantly higher after they participated in workshops related to topics of the theory of evolution, $z = -2.449$, $p = 0.014$, with a high effect size ($r = 0.74$). The median score among biology teachers prior to participating in these workshops was 84, which indicates a high acceptance according to Rutledge and Warden ([2007](#)). The median score increased to 90 after the workshops, which indicates a very high acceptance (Rutledge and Warden [2007](#)).

Discussion and conclusion

Research design requires the use of instruments that can unequivocally address a particular construct, in this case the acceptance of the theory of evolution. To measure the impact of a professional development program on evolution-related concepts, we developed an instrument to assess the acceptance of the theory of evolution among Puerto Rican high school teachers and students whose first language is Spanish. The development of instruments adapted to other languages and cultures facilitates cross-cultural research, which has relevance from theoretical and practical perspectives (Peña [2007](#)). According to a recent report, the Spanish language is the world's second language in terms of number of native speakers (about 493 million) and the second language of international communication (Fernández Vítores [2021](#)). Moreover, Spanish is the most common non-English language spoken in the US (62% of homes) corresponding to Hispanics being the largest minority group in the US
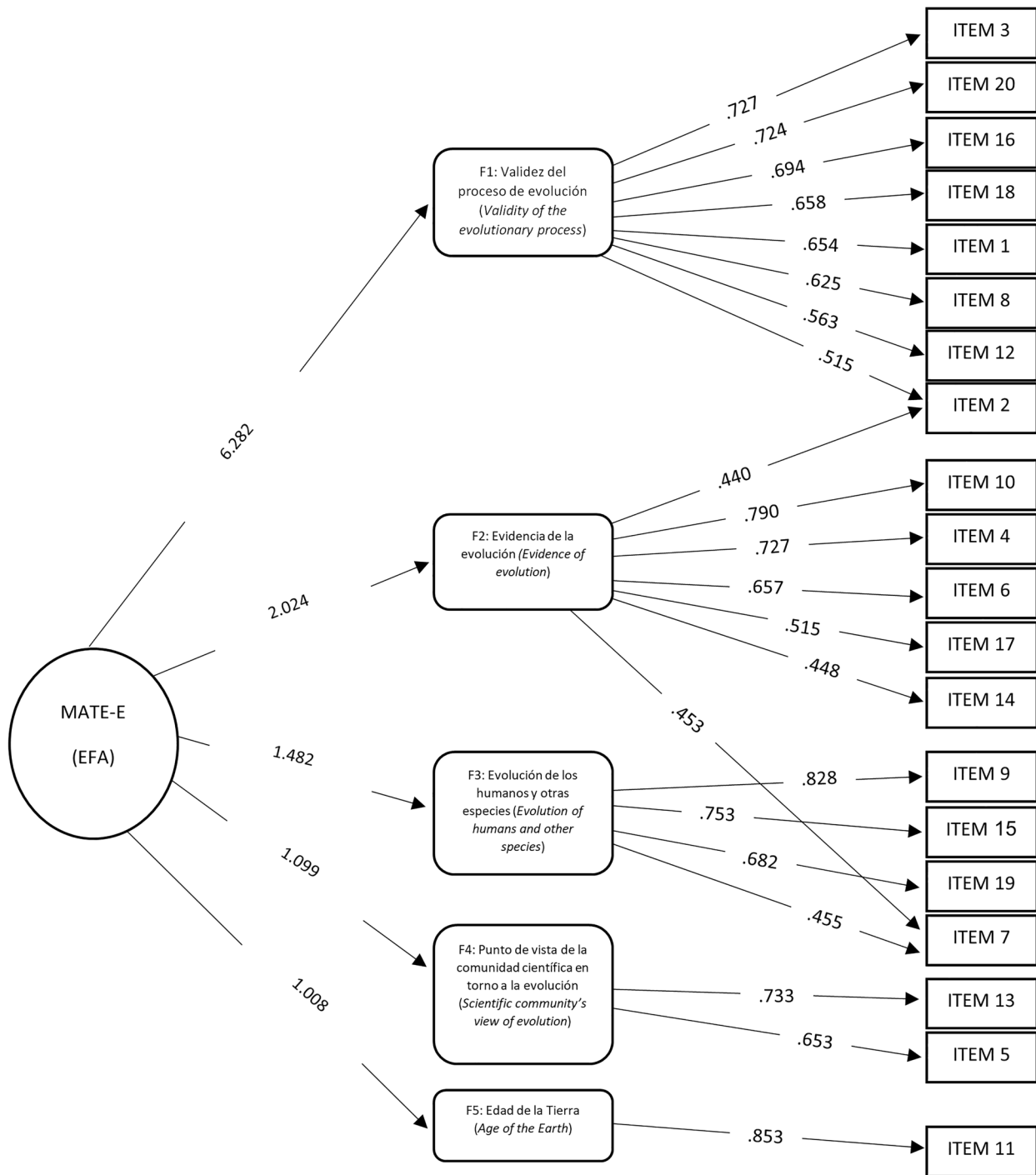
Fig. 3 Path diagram for MATE-E's Exploratory Factor Analysis

(Dietrich and Hernandez 2022). Hence, the development of a Spanish-language instrument has the potential to be adapted and used in different research scenarios, if cultural biases are addressed (Peña 2007).

The MATE has been widely used in evolution education research since it was developed in 1999 (Athanasiou et al. 2011; Athanasiou and Papadopoulou 2012; Barnes et al. 2024; Beniermann et al. 2023; Deniz et al. 2008; Lammert 2012; Metzger et al. 2018; Rutledge and Sadler

Table 9 Construct naming for MATE-E's factors

Factor	Construct names in Spanish (English translation in parenthesis)	Items	Sample items	Items in English
1	Validez del proceso de evolución (*Validity of the evolutionary process*)	1, 2, 3, 8, 12, 16, 18, 20	(20) La evolución es una una teoría científica válida	(20) Evolution is a scientifically valid theory
2	Evidencia de la evolución (*Evidence of evolution*)	4, 6, 10, 14, 17	(14) La teoría de la evolución no puede ser correcta ya que difiere del relato bíblico de la creación	(14) The theory of evolution cannot be correct since it disagrees with the Biblical account of creation
3	Evolución de los humanos y otras especies (*Evolution of humans and other species*)	7, 9, 15, 19	(15) Los humanos existen hoy en día esencialmente de la misma forma en que siempre lo han hecho	(15) Humans exist today in essentially the same form in which they always have
4	Punto de vista de la comunidad científica en torno a la evolución (*Scientific community's view of evolution*)	5, 13	(13) La teoría de la evolución genera predicciones comprobables en relación con las características de la vida	(13) Evolutionary theory generates testable predictions with respect to the characteristics of life
5	Edad de la Tierra (*Age of the Earth*)	11	(11) La edad de la tierra es de al menos 4000 millones de años	(11) The age of the earth is at least 4 billion years

Table 10 Independent samples *t*-test – data sorted by gender

Levene's Test for Equality of Variances		t	df	p value	Mean Difference	Std. Error Difference	95% Confidence Interval of Difference	
F	Sig						Lower	Upper
.075	.784	−.285	172	.776	−.570	2.002	−4.521	3.382

Table 11 Independent samples *t*-tests – data sorted by grade

Observed pairs (*t*-tests)	Levene's Test for Equality of Variances		t	df	p value	Mean Difference	Std. Error Difference	95% Confidence Interval of Difference	
	F	Sig						Lower	Upper
8th and 10th grade	.422	.518	−2.037	61	.046	−7.921	3.888	−15.696	−.146
8th and 11th grade	.534	.466	−2.754	153	.007	−6.419	2.331	−11.023	−1.815
10th and 11th grade	.120	.730	.504	128	.615	1.502	2.978	−4.390	7.394

Table 12 Summary of sources of validity evidence reported on several versions of the MATE

Article	Sources of validity evidence				
	Content	Response process	Internal structure	Relationship with other variables	Consequences of testing
Original MATE (Rutledge & Warden 1999)	Yes	No	Yes	No	No
MATE adaptation for undergraduate students (Rutledge & Sadler 2007)	No	No	Yes	Yes	Yes
MATE of Greek education students (Athanasiou et al. 2011)	*	*	Yes	Yes	No
MATE translation and adaptation to German (Lammert 2012)	Yes	No	Yes	Yes	Yes
MATE translation and adaptation to Turkish (Tekkaya et al. 2012)	*	*	Yes	Yes	No
MATE-E (Pérez-Vega et al. 2025) - current study	Yes	Yes	Yes	Yes	Yes

*It is mentioned in the article that the MATE was previously translated and adapted, but no evidence of content or response process is reported

2007; Rutledge and Warden 1999; Tekkaya et al. 2012). Moreover, it has been adapted, with various degrees of validation evidence, to other languages like German (Lammert 2012), Greek (Athanasiou et al. 2011), and Turkish (Tekkaya et al. 2012) (see Tables 1 and 12), making it difficult to reach generalizations (Kuschmierz et al. 2020). Nevertheless, it has been the instrument of choice, despite its limitations within the definition of the construct (Romine et al. 2018; Wagler and Wagler 2013), when comparing results to the existing literature.

In this work we present evidence on the translation, adaptation, and validation of the MATE in Español (MATE-E). A translation/back-translation framework, complemented by an evaluation done by an expert panel provided precise content validity for the instrument. Furthermore, a focus group was carried out to gather evidence for the response process. High school students were selected for this process because they are part of our target population for future studies. We assume that if the language is clear enough for high school students

to understand, their teachers would not have difficulties understanding the instrument. During the focus group, topics such as the test's content, and the participants' response process were discussed to address cultural equivalence. These results and their subsequent evaluation by the expert panel provided further validity evidence and insight into the changes needed to improve the instrument to reduce language and cultural biases.

To gather validity evidence for MATE-E's internal structure and reliability, a pilot study was carried out with secondary school students (n = 185). Reliability testing results show a Cronbach's alpha of 0.879, which is well within the range of a good ($0.80 \geq \alpha < 0.90$) internal consistency according to George and Mallery (2003).

The Cronbach's alpha for MATE-E, though, is lower than the one reported in the original instrument, which is 0.98 (Rutledge and Warden 1999). This may be due to factors such as a lower dataset than the original or due to cultural differences in translation and adaptation. However, it is noteworthy to point out that Cronbach's alpha

Pérez-Vega *et al. Evolution: Education and Outreach*   (2025) 18:2

Page 18 of 20

Table 13 Descriptive statistics for MATE-E administration to teachers

		Pretest	Posttest
N	Valid	11	11
	Missing	0	0
Mean		82.91	89.09
95% Confidence Interval for Mean	Lower Bound	75.66	83.43
	Upper Bound	90.15	94.76
5% Trimmed Mean		83.23	89.38
Std. Error of Mean		3.251	2.542
Median		84.00	90.00
Std. Deviation		10.784	8.432
Variance		116.291	71.091
Skewness		−.417	−.509
Std. Error of Skewness		.661	.661
Kurtosis		−.648	−.137
Std. Error of Kurtosis		1.279	1.279
Range		34	27
Minimum		63	73
Maximum		97	100

constantly remains in the range of good internal consistency even if other items are deleted. When compared to other versions of the MATE analyzed in this paper, the MATE-E shows good internal consistency and stands out as the second-highest reliability among all observed translations. These tests showed that the translated, adapted, and validated MATE-E possesses good internal consistency and additional tests for sampling adequacy and sphericity confirmed that the data could undergo an exploratory factor analysis.

An Exploratory Factor Analysis (EFA) was used to identify the underlying factor structure of the instrument after its translation and adaptation. As part of the EFA, a Principal Component Analysis extracted five components with Eigenvalues greater than 1. Despite Rutledge and Warden's (1999) initial reporting of a standalone factor of evolution acceptance for the original MATE test, more recent research, such as analyses done by Metzger et al. (2018) point to the instrument having more than one factor (Barnes et al. 2019).

The exploratory factor analysis identified five components that explain 59.472% of the instrument's variance, which closely approximates Pituch and Stevens' (2016) recommended 60% threshold of total variance explained for factor analysis. Due to each item's strong correlation to their respective components, no items from the MATE were excluded from the MATE-E instrument. This is an important difference between the MATE-E instrument

and other adaptations of the MATE in the literature, which could explain some of the differences researchers have found when using these adaptations (Athanasiou et al. 2011; Athanasiou and Papadopoulou 2012; Deniz et al. 2008; Lammert 2012; Rutledge and Sadler 2007; Rutledge and Warden 1999; Tekkaya et al. 2012). Factors were also found to behave in a comparable manner as Rutledge and Sadler's (2007) discussion of MATE's concepts. Some items aligned with the components similarly in the MATE-E. These concepts were considered when naming MATE-E's factors. All evidence considered; the MATE-E instrument is found to be fit for its use with Puerto Rican Spanish speakers.

Further analysis of the pilot study data to gather evidence based on the relationships with other variables showed, as expected, no statistically significant differences in acceptance of the theory of evolution between male and female students $t(172) = -0.285$, $p = 0.776$. We further explored if there were any potential differences by grade level, given that biology content knowledge tends to increase with grade progression. In Puerto Rico's high school curriculum, students typically take an introduction to biological sciences in 7th grade and a more comprehensive Biology course in 10th grade. As predicted, we found that the mean scores of 8th grade participants when compared to 10th and 11th grade participants were lower and statistically different.

The final version of the MATE-E was compared to other available versions and adaptations of the MATE. Rutledge and Warden (1999) recommended exploring the relationship between the acceptance of evolution with other variables (such as understanding of the theory of evolution or understanding of the nature of science) using other instruments. Some published adaptations of the MATE followed their suggestion (Athanasiou et al. 2012; Lammert 2012; Rutledge and Sadler 2007; Tekkaya et al. 2012). However, the MATE-E was analyzed using only the available variables (grade and gender) to keep the data strictly within the parameters of the translation, adaptation, and validation process.

Validity evidence for the instrument's use was gathered by measuring changes in acceptance of the theory of evolution among in-service high school biology teachers. The MATE-E was administered as a pre/post assessment along specific activities on concepts related to the theory of evolution as part of a two-year professional development (PD) program. We tested the instrument with eleven Puerto Rican, Spanish-speaking biology teachers. We hypothesized that teachers' MATE-E scores would increase after completing the PD activities. Indeed, the median score rose from 84 (high acceptance) to 90 (very

high acceptance), per Rutledge and Warden's (1999) criteria. A Wilcoxon test confirmed a significant increase in acceptance, supporting our hypothesis about the instrument's score interpretation. Moreover, the data suggests that the PD program effectively fostered greater acceptance of evolution among teacher participants.

The present study provides evidence for the adaptation, reliability, and validity of the MATE-E, a Spanish translation of the Measure of Acceptance of the Theory of Evolution, and supports its use in research and evaluation with Puerto Rican, Spanish-speaking secondary-level students and in-service biology teachers.

Limitations

As expected in these types of studies, there are limitations associated with the generalization of results. First, all research activities were conducted using online tools and platforms as they took place during the COVID-19 pandemic lockdown period. Hence, the expert panel meetings, focus group, and the pilot study, were all performed online. Other researchers who plan to replicate this study may opt to carry out research activities in person. Second, there was potential sampling bias during the pilot study phase because all participants were from the University of Puerto Rico, Río Piedras Campus' Laboratory Secondary School. Researchers who may seek to perform a confirmatory factor analysis to test whether data fits the measurement model are recommended to further expand the population to include students from different schools in their sample. Lastly, regarding the data obtained from teacher participants, as there were few participants (n = 11) these results cannot be generalized because they do not meet sample size and sample representation criteria. A larger, more representative sample size should be considered for future studies, to confirm if results can be generalized.

It is noteworthy that before this report, Barnes et al. (2022) published a revised version of the MATE instrument, the MATE 2.0. The revised instrument contains nine items aligned to the consensus definition of acceptance of evolution and has a different scoring system (Barnes et al. 2022). Following Barnes et al. (2024) recommendation, we are considering the use of the MATE 2.0 for further studies; however, it should first be translated to Spanish and culturally adapted for our purposes. Evidence for its validation should also be gathered to determine if the instrument is appropriate for use with our target audience.

Abbreviations
EFA   Exploratory factor analysis
MATE   Measure of acceptance of the theory of evolution
MATE-E   Measure of acceptance of the theory of evolution-español
PD   Professional development

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12052-025-00217-4.

Additional file 1.

References
American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 2014.
Athanasiou K, Papadopoulou P. Conceptual ecology of the evolution acceptance among Greek education students: knowledge, religious practices

and social influences. Int J Sci Educ. 2012;34(6):903–24. https://doi.org/10.1080/09500693.2011.586072.

Athanasiou K, Katakos E, Papadopoulou P. Conceptual ecology of evolution acceptance among Greek education students: the contribution of knowledge increase. J Bio Educ. 2011;46(4):234–41. https://doi.org/10.1080/00219266.2012.716780.

Barnes ME, Dunlop HM, Holt EA, Zheng Y, Brownell SE. Different evolution acceptance instruments lead to different research findings. Evo Edu Outreach. 2019;12(4):1–17. https://doi.org/10.1186/s12052-019-0096-z.

Barnes ME, Misheva T, Supriya K, Rutledge M, Brownell SE. A revised Measure of Acceptance of the Theory of Evolution: introducing the MATE 2.0. Life Scie Ed. 2022;21(1):1–14. https://doi.org/10.1187/cbe.21-05-0127.

Barnes ME, Aini RQ, Collins JP, Dunk RDP, Holt EA, Jensen J, Klein JR, Misheva T, Nadelson LS, Reiss MJ, Romine WL, Shtulman A, Townley AL, Wiles JR, Zheng Y, Brownell SE. Evaluating the current state of evolution acceptance instruments: a research coordination network meeting report. Evo Edu Outreach. 2024;17:1. https://doi.org/10.1186/s12052-024-00194-0.

Behling O, Law K. Translating questionnaires and other research instruments. Problems and solutions. Thousand Oaks: Sage Publications, Inc.; 2000.

Beniermann A, Moormann A, Fiedler D. Validity aspects in measuring evolution acceptance: evidence of preservice biology teachers and creationists. J Res Sci Teach. 2023;60:1223–65. https://doi.org/10.1002/tea.21830.

Bravo M. Instrument development: Cultural adaptations for ethnic minority research. In: Bernal G, Trimble JE, Burley AK, Leong FTL, editors. Handbook of racial and ethnic minority psychology. Thousand Oaks: Sage; 2003. p. 220–38.

Bravo-Vick M, Padro-Collazo P, Borrero M. Mathematics and science attitude inventory: validation for use in a Spanish speaking context. Revista De Estudios e Investigación En Psicología y Educación. 2019;6(2):167–78. https://doi.org/10.17979/reipe.2019.6.2.5687.

Creswell JW. Educational research: Planning, conducting, and evaluating quantitative and qualitative research. Boston: Pearson; 2012.

Cronbach LJ. Essentials of psychological testing. New York: Harper & Row; 1984.

Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull. 1955;52(4):281–302. https://doi.org/10.1037/h0040957.

Deniz H, Donnelly LA, Yilmaz I. Exploring the factors related to acceptance of evolutionary theory among Turkish preservice biology teachers: toward a more informative conceptual ecology for biological evolution. J Res Sci Teach. 2008;45(4):420–43. https://doi.org/10.1002/tea.20223.

DeVellis RF. Scale development: theory and applications. Thousand Oaks: SAGE; 2012.

Dietrich S, Hernandez E. Nearly 68 Million People Spoke a Language Other Than English at Home in 2019. United States Census Bureau; 2022. https://www.census.gov/library/stories/2022/12/languages-we-speak-in-united-states.html.

Dobzhansky T. Nothing in biology makes sense except in the light of evolution. Am Biol Teach. 1973;35(3):125–9. https://doi.org/10.2307/4444260.

Fernández Vítores D. El español: Una lengua viva. Informe 2021. Instituto Cervantes; 2021.

Friedrichsen PJ, Linke N, Barnett E. Biology teachers' professional development needs for teaching evolution. Sci Educator. 2016;25(1):51–61.

George D, Mallery P. SPSS for Windows step by step: A simple guide and reference. 11.0 update. Boston: Allyn & Bacon; 2003.

Harms U, Yarden A. Teacher preparation and evolution education: challenges and opportunities—introduction to the special issue. Evo Edu Outreach. 2023;16:14. https://doi.org/10.1186/s12052-023-00193-7.

International Test Commission. The ITC guidelines for translating and adapting tests (second edition). International Test Commission; 2017.

Kuschmierz P, Meneganzin A, Pinxten R, Pievani T, Cvetković P, Mavrikaki E, Graf D, Beniermann A. Towards common ground in measuring acceptance of evolution and knowledge about evolution across Europe: a systematic review of the state of research. Evo Edu Outreach. 2020;13:18. https://doi.org/10.1186/s12052-020-00132-w.

Lammert N. Akzeptanz, Vorstellungen und Wissen von Schülerinnen und Schülern der Sekundarstufe I zu Evolution und Wissenschaft [Acceptance, conceptions, and knowledge of lower secondary education stu-dents about evolution and science] [dissertation]. Dortmund, Germany: Technical University of Dortmund; 2012. http://129.217.131.68:8080/handle/2003/29476.

McMillan JH. Fundamentals for educational research. Boston: Pearson; 2016.

Messick S. Meaning and values in test validation: The science and ethics of assessment. Educ Res. 1989;18:5–11. https://doi.org/10.3102/0013189X018002005.

Metzger KJ, Montplaisir D, Haines D, Nickodem K. Investigating undergraduate health sciences students' acceptance of evolution using MATE and GAENE. Evolution. 2018;11(1):10. https://doi.org/10.1186/s12052-018-0084-8.

Nadelson LS, Southerland S. A more fine-grained measure of students' acceptance of evolution: development of the inventory of student evolution acceptance—I-SEA. Int J Sci Educ. 2012;34(11):1637–66. https://doi.org/10.1080/09500693.2012.702235.

Nehm RH, Schonfeld IS. Does increasing biology teacher knowledge of evolution and the nature of science lead to greater preference for the teaching of evolution in schools? J Sci Teach Educ. 2007;18:699–723. https://doi.org/10.1007/s10972-007-9062-7.

Peña ED. Lost in translation: methodological considerations in cross-cultural research. Child Dev. 2007;78(4):1255–64. https://doi.org/10.1111/j.1467-8624.2007.01064.x.

Pituch KA, Stevens JP. Applied multivariate statistics for social sciences. New York: Rutledge; 2016.

Romine WL, Todd AN, Walter EM. A closer look at the items within three measures of evolution acceptance: analysis of the MATE, I-SEA, and GAENE as a single corpus of items. Evo Edu Outreach. 2018;11(17):1–20. https://doi.org/10.1186/s12052-018-0093-7.

Rutledge ML, Sadler KC. Reliability of the measure of acceptance of the theory of evolution (MATE) instrument with university students. Am Biol Teach. 2007;69(6):332–5. https://doi.org/10.1662/0002-7685(2007)69[332:ROTMOA]2.0.CO;2.

Rutledge ML, Warden MA. The development and validation of the measure of acceptance of the theory of evolution instrument. Sch Sci Math. 1999;99(1):13–8. https://doi.org/10.1111/j.1949-8594.1999.tb17441.x.

Smith MU, Snyder SW, Devereaux RS. The GAENE—generalized acceptance of evolution evaluation: development of a new measure of evolution acceptance. J Res Sci Teach. 2016. https://doi.org/10.1002/tea.21328.

Tekkaya C, Akyol G, Sungur S. Relationships among teachers' knowledge and beliefs regarding the teaching of evolution: a case for Turkey. Evo Edu Outreach. 2012;5(3):477–93. https://doi.org/10.1007/s12052-012-0433-y.

Wagler A, Wagler R. Addressing the lack of measurement invariance for the measure of acceptance of the theory of evolution. Int J Sci Educ. 2013;35:2278–98. https://doi.org/10.1080/09500693.2013.808779.

Publisher's Note