The project is supported by

Evolution Education Questionnaire on Acceptance and Knowledge (**EEQ**)

Standardised and ready-to-use protocols to measure acceptance of evolution and knowledge about evolution in an international context

This method publication is based upon work from *COST Action (CA17127) Building on scientific literacy in evolution towards scientifically responsible Europeans,* supported by COST (European Cooperation in Science and Technology). www.euroscitizen.eu

This publication refers to Deliverable 2 (standardised and ready-to-use protocols [e.g. survey design, questionnaire(s), procedures, data preparation and statistics] in Action members' European languages) of Working Group 1 of EuroScitizen COST Action.

Authors:

Beniermann, A., Kuschmierz, P., Pinxten, A., Aivelo, T., Bohlin, G., Brennecke, J. S., Cebesoy, U. B., Cvetković, D., Đorđević, M., Dvořáková, R. M., Futo, M., Geamana, N., Korfiatis, K., Lendvai, A., Mogias, A., Paolucci, S., Petersson, M., Pietrzak, B, Porozovs, J., Realdon, G., Savković, U., Sofonea, M. T., Šorgo, A., Stermin, A. N., Torkar, G., Uitto, A., Vázquez-Ben, L., & Graf, D.

Abstract

The lack of standardised assessment of evolutionary knowledge and acceptance of evolution across Europe makes comparisons between studies difficult. The Evolution Education Questionnaire on Acceptance and Knowledge (EEQ) was constructed to measure attitudes and understanding across Europe and beyond. We aimed to compile a brief instrument to allow for easy application in school and university. The target group of the EEQ was freshman university students who had just finished their secondary education. However, several components of the questionnaire were developed and validated for additional target groups. Therefore, this questionnaire may, in addition, be suitable for students in secondary school, in-service teachers as well as the general public.

This method report describes the contents and application of the EEQ and provides information on survey conduction, data preparation, analyses and interpretation of results to serve as a standardised and ready-to-use protocol to measure the acceptance of and knowledge about evolution in a local, national or international context. To allow for sampling in different European countries, we present the EEQ in 23 European languages.

Keywords: evolution, acceptance of evolution, knowledge about evolution, cross-country survey, assessment, measurement, scientific literacy, creationism, Europe

The project is supported by

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

EuroSciTizen Building on scientific literacy in Evolution

Didactics of Biology Giessen

JLU NEW PATHS. SINCE 1607

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Outline

The project is supported by
cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY
EuroSciTizen Building on scientific literacy in Evolution
Didactics of Biology Giessen
JLU NEW PATHS. SINCE 1607
JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

1. Aim of the Evolution Education Questionnaire on Acceptance and Knowledge (EEQ)

Across Europe, a lack of standardised assessment of evolutionary knowledge and acceptance - and, therefore, a lack of reasonably comparable data - prevails (Kuschmierz et al., 2020a). In their systematic review, Kuschmierz et al. (2020a) revealed that only approximately one-third of all studies on acceptance and/or knowledge about evolution provided evidence for local validity and reliability.

The Evolution Education Questionnaire on Acceptance and Knowledge (EEQ) was constructed to measure attitudes and understanding across Europe and beyond. We aimed to compile a brief instrument to allow for easy application in school and university. Furthermore, the data analysis can be conducted without the efforts of content analysis since we almost exclusively implemented rating scales (attitudes towards evolution, religious faith and dualistic thinking) and multiple-choice questions (knowledge about evolution).

The EEQ was translated into many European languages (Fig. 1): Bulgarian, Croatian, Czech, Dutch, English (British), Finnish, French, German, Greek, Hungarian, Italian, Latvian, Macedonian, Polish, Portuguese, Romanian, Serbian, Slovak, Slovenian, Spanish, Swedish, Turkish and Ukrainian. Versions of the EEQ in these languages are published with this report (see Appendices).

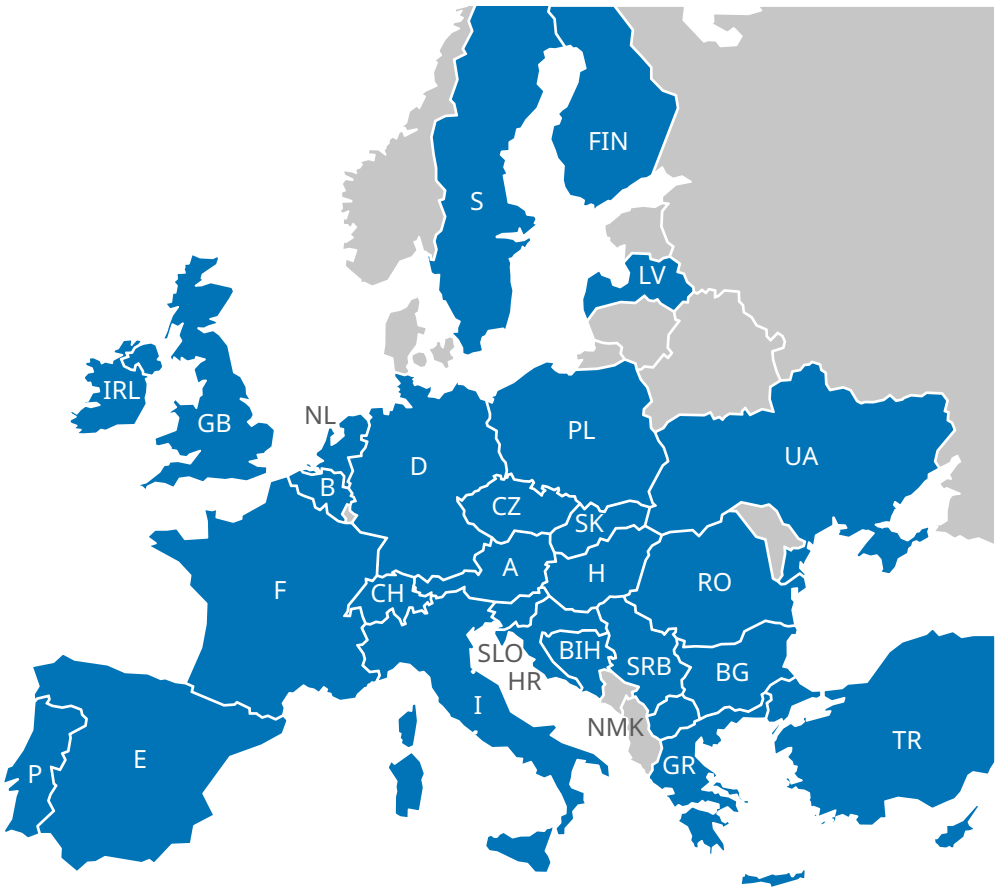

Figure 1: Countries for which translations of the EEQ are available.

The target population of the EEQ are freshman university students who had just finished their secondary education. However, several components of the questionnaire were developed and validated for additional target groups (see Section 2). Therefore, this questionnaire may also be suitable for secondary school students in the 10th grade and above, undergraduates (biology majors and non-majors), in-service teachers and the general public.

By using the EEQ,

- teachers and educators can assess their students' knowledge about various evolutionary aspects of microevolution and macroevolution while identifying underlying misconceptions to make their teaching more efficient and expedient.
- instructors can utilise a single assessment to provide a broad overview of students' knowledge regarding various aspects of evolution.
- religious and non-religious persons can be surveyed.
- instructors can distinguish between attitudes towards evolution in general, as well as attitudes towards the evolution of humans, including personality traits.
- comparative studies (within or between country samples) can be achieved.

2. Background and content of the questionnaire

To investigate the acceptance of evolution, knowledge about evolution and the relationship between these factors, as well as other factors (e.g. religious faith), across studies or countries, the instruments utilised should measure equivalent constructs (i.e. content validity). Furthermore, comparative analyses require appropriate evidence for validation in the local context of the single surveys or survey populations (AERA et al., 2014). In recent years, the debate concerning measurement issues in evolution education has accelerated, and researchers have continuously addressed it (e.g. Anderson et al., 2010; Barnes et al., 2019; Beniermann, 2019; Kuschmierz et al., 2020a; McCain & Kampourakis, 2018; Mead et al., 2019; Nehm & Mead, 2019; Nehm & Schonfeld, 2008; Novick & Catley, 2012; Smith et al., 2016) These efforts are valuable since multiple challenges arise regarding measurement in evolution education research:

a) Inadequate definitions of key constructs

The key constructs in evolution education research, such as attitudes, acceptance, knowledge and understanding, are often not defined or inconsistently defined and used in publications, which may lead to different operationalisations (Ha et al., 2012; Konnemann et al., 2012; McCain & Kampourakis, 2018). In accordance with Smith and Siegel (2016), we use the term *knowledge* when referring to measurement instruments that focus on content knowledge: *'a student gains knowledge (via instruction, self-study, etc.), upon which she can build understanding'* (Smith & Siegel, 2016, p. 486). We use the term *acceptance* to describe a positive attitude towards evolution, while a negative attitude is called *rejection* (Beniermann, 2019).

b) Application of diverse measurement instruments

Barnes et al. (2019) demonstrated that different evolution acceptance instruments may produce different results, even when applied to the same population. This reveals a potential bias when comparing research results derived from studies that employed different instruments to assess acceptance of evolution. In particular, acceptance is generally higher for 1) microevolution than for human evolution (Barnes et al., 2019), as well as for 2) evolution in general than for evolution of the human mind (Beniermann, 2019). Hence, these differences must be considered when measuring evolution acceptance (Kampourakis & Strasser, 2015; Rughiniş, 2011).

c) The multidimensionality of knowledge about evolution

Most instruments that aim to assess knowledge or understanding of evolution focus on single evolutionary constructs, mostly natural selection and related concepts (e.g. Concept Inventory of Natural Selection [CINS], Anderson et al., 2002). However, the authors of the current report see knowledge about evolution as a multidimensional construct; thus, results may be dependent on the evolutionary concept that is assessed (Kuschmierz et al., 2020b). In addition, Nehm and Ha (2011) illustrated that specific contexts evoke different types and magnitudes of key concepts of natural selection and alternative conceptions.

d) Measurement standards

The application of measurement standards (AERA et al., 2014) is crucial for instruments measuring evolutionary knowledge and acceptance (Mead et al., 2019). Mead et al. (2019) revealed validity and reliability issues for some often-used instruments. Additionally, most instruments are validated for only one specific population (Mead et al., 2019), and many studies only utilise parts or modified versions of published instruments (Kuschmierz et al., 2020a). These issues may affect an instrument's ability to measure the intended construct (Mead et al., 2019).

The challenges above indicate the difficulty of comparing results gathered through the use of different instruments. Group comparisons - between students of different grades, people from different countries or even participants receiving different survey instructions - are only sound if comparable data is available for all compared groups. Therefore, it is essential to employ instruments for which there is supporting evidence to measure the same construct - or ideally even the same instrument - and to target similar sample groups to gain comparable data.

To enable sound comparative surveys, we compiled the Evolution Education Questionnaire on Acceptance and Knowledge (EEQ). The EEQ consists of several published scales for which evidence regarding local validity and reliability was provided. The scales utilised in the EEQ cover the topics of 'knowledge about evolution' (KAEVO 2.0; Kuschmierz et al., 2020b), 'attitudes towards evolution' (ATEVO; Beniermann, 2019), 'religious faith' (PERF; Beniermann, 2019) and 'dualistic thinking' (SD; Beniermann, 2019), as well as several socio-demographic questions. In the following sections, we describe the theoretical background behind these topics and the corresponding scales.

2.1 The Knowledge About EVOlution (KAEVO 2.0) instrument

Research has consistently identified many aspects of evolution as being difficult to understand, e.g. the tremendous timespan of Earth's history and the fact that evolution has no purpose or direction (Catley & Novick, 2009; Gregory, 2009). Several studies have also shown that students substitute non-scientific explanations for the mechanisms of evolution (e.g. Fiedler et al., 2017; Nehm et al., 2012). These misconceptions have a crucial impact on students' learning, and their handling is the foundation for evolution education. Conceptions regarding aspects of evolution often do not differ arbitrarily between individuals (Gregory, 2009), and it is possible to identify and describe typical and recurring patterns and types of conceptions.

The existing literature includes several multiple-choice instruments to assess knowledge about evolution (e.g. Evolution Concept Test [ECT], Bishop & Anderson, 1990; Concept Inventory of Natural Selection [CINS], Anderson et al., 2002; Concept Assessment of Natural Selection [CANS], Kalinowski et al., 2016; Measure of Understanding of Macroevolution [MUM], Nadelson & Southerland, 2009; Genetic Drift Inventory [GeDI], Price et al., 2014; Evo-Devo Concept Inventory [EvoDevoCI], Perez et al., 2013). These instruments differ concerning the examined evolutionary concepts and the target groups.

The Knowledge About Evolution (KAEVO) 2.0 instrument (Kuschmierz et al., 2020b) contains various aspects of evolution that address microevolution and macroevolution. Building upon KAEVO 1.0 (Beniermann, 2019), the development of KAEVO 2.0 extended and modified the instrument based on aspects of validity (AERA et al., 2014).

KAEVO 2.0 contains aspects of evolution that high school students should know. The development of the questionnaire was based on a curricula and textbook analysis to address content validity, and European experts in the fields of biology education and evolutionary biology reviewed the instrument (Kuschmierz et al., 2020b).

KAEVO 2.0 was then pre-tested in an assessment of high school and university undergraduate students' evolution knowledge. The instrument consists of three parts with different answer formats: A) multiple-choice questions, B) wrong/false questions and C) time scale questions in an open response format. The instrument covers the concepts of biological fitness, deep time, heredity, human evolution, mutations, natural selection and phylogenetics.

The KAEVO 2.0 is intended to be an 'allrounder' among instruments measuring knowledge about evolution, as it covers different concepts of evolution. Furthermore, evidence was provided that KAEVO 2.0 is suitable for various target groups (high school and university undergraduate students in biology-related and non-biology-related fields of study; Kuschmierz et al., 2020b). All data and analyses are available in Kuschmierz et al. (2020b).

2.2 The Attitudes Towards EVOlution (ATEVO) scale

The acceptance of evolution is mostly assessed using the MATE (Measure of Acceptance of the Theory of Evolution) scale (Rutledge & Warden, 1999). However, some indications suggest that the MATE scale is not suitable for all ages and educational levels (Wagler & Wagler, 2013). Furthermore, distinguishing between knowledge about evolution and attitudes towards evolution is difficult using the MATE scale (Konnemann et al., 2012). Moreover, the MATE scale includes questions referring to religious faith and/or creationism, which may be a problematic compound, at least in more secular populations (Beniermann, 2019).

The Attitudes Towards EVOlution (ATEVO) scale (Beniermann, 2019) consists of eight rating scale items organised into two subscales with four items each. The subscales reflect attitudes towards evolution in general (ATEVO-EG) and attitudes towards evolution of the human mind (e.g. consciousness, personality traits; ATEVO-EM). This focus on acceptance of the evolution of cognitive mechanisms in humans refers to the philosophical position of *evolutionary epistemology* (Popper, 1985; Wuketits, 1990). The assumption that cognitive abilities, culture and moral concepts have evolved is widely shared within the scientific community (Henrich & McElreath, 2003) and can be deduced from knowledge about the evolutionary history of humans: *"From the fact of (human) evolution it inevitably follows that (human) mental capacities, our cognitive and knowledge capacities, result from evolutionary processes"* (Wuketits, 1990, p. 33). In evolutionary epistemology it is considered that the organic evolution provides the preconditions for any cultural evolution (Popper, 1985). This position, which results from evolutionary theory may be hard to accept even for non-religious people (Voland, 2010), which was empirically shown in the USA (Paz-y-Miño-C & Espinosa, 2012) and Germany (Beniermann, 2019).

The development of the ATEVO scale commenced with an item pool of existing instruments (e.g. Astley & Francis, 2010; Ingram & Nelson, 2006; Johnson & Peeples, 1987; Lombrozo et al., 2008; Rutledge & Warden, 1999; Taber et al., 2011), as well as theoretical considerations concerning items that aim to measure attitudes towards evolution of the human mind (Beniermann, 2019). To further ensure content validity, experts from the fields of biology, biology education and biophilosophy, as well as people holding divergent views on evolution (creationists, naturalists and theologists), were asked to review and evaluate the items and offer feedback regarding the comprehensibility and operationalisation of the construct. Thus, the collection of items was subsequently selected, supplemented and refined in consultation with these experts.

The clause 'Personally, I think that…' introduces each item to ensure that participants know that this part of the questionnaire is about their personal opinions on evolution (rather than their knowledge about evolution or the scientific consensus' perspective on evolution). Another criterion applied to the operationalisation was the linguistic and content-related adequacy of instructions and items, especially for younger learners. For this purpose, pre-tests with high school students (N = 142; 13 to 20 years) were conducted to ensure validity of answer processes (AERA et al., 2014).

Evidence for local validity and reliability for the ATEVO scale was shown based on four studies (N_{total} = 9,311; Beniermann, 2019). Survey populations differed across the four studies, which further ensures that the ATEVO scale is a suitable instrument to measure attitudes towards evolution for groups of different ages and education, for the general public and for explicitly non-religious people. The populations for the four studies (Beniermann, 2019) were as follows: an unsystematic sample of

the general public (EGl study), four subgroups with diverse levels of biology education (high school seventh grade, high school 10-11th grade, university students and pre-service teachers; EWi study), a representative German sample (RED study) and participants with no religious affiliation (EKI study). Reliability for ATEVO, as well as for the two subscales, was good in all studies. Principal component analyses (PCA) revealed that an expectedly two-dimensional structure of the ATEVO scale was suitable in most of the samples. For the EGl study, however, the PCA revealed a one-dimensional structure, reflecting the observation that the size of the difference between acceptance of evolution and acceptance of evolutionary epistemology varies across survey populations (Beniermann, 2019).

Thus, evidence demonstrates that the ATEVO scale is suitable for different target groups (high school and university undergraduate students, in-service teachers and the general public; Beniermann, 2019). All data and analyses are available in German in Beniermann (2019).

2.3 The PErsonal Religious Faith (PERF) and Short Dualism (SD) scales

In addition to the KAEVO 2.0 and ATEVO scales, the EEQ incorporates two more scales to assess respondents' religious faith (Personal Religious Faith [PERF]; Beniermann, 2019) and dualistic thinking (Short Dualism [SD]; Beniermann, 2019). Their development was based on the same measurement standards and procedure described above for the ATEVO scale.

The 10 items of the PERF scale focus on respondents' personal religious beliefs, addressing different aspects of monotheistic faith and behaviour. Meanwhile, the SD scale measures the degree to which people favour dualistic theories of mind. This scale is a shortened (five items) and modified version of Stanovich's Dualism scale (Stanovich, 1989). All items utilise a five-point rating scale.

Evidence reveals that the PERF and SD scales are suitable for different target groups (high school and university undergraduate students, in-service teachers and the general public; Beniermann, 2019). All data and analyses on validity and reliability are available in German in Beniermann (2019).

3. Survey conduction

To facilitate the aim of the initial survey (Kuschmierz et al., 2021) - to survey freshmen university students directly within university courses and thereby maintain greater control over the surveyed group and assessment situation - the EEQ was constructed for application as a pen-and-paper questionnaire.

a) Instructors who wish to produce data that is comparable to Kuschmierz et al. (2021) should utilise the following guidelines:

- Sample freshman university students. It is important to survey the students, if possible, shortly after they have finished their secondary education.
- Ensure that participants are neither currently attending nor have already attended an evolutionary biology university course.

- Focus on students who are enrolled in a biology-related university programme, or divide the population between students from a biology-related university programme and those from a non-biology-related university programme. We defined the university programmes in Infobox 1 as biology-related university programmes. However, this list is not exhaustive.

• Agricultural sciences	• General biology
• Anatomical sciences	• Genetics
• Biochemistry	• Immunology
• Bioinformatics	• Life science
• Biology education	• Medicine
• Biomedical sciences	• Microbiology
• Biophysics	• Molecular biology
• Biotechnology	• Natural sciences
• Cellular biology	• Nursing college
• Ecology	• Nutrition science
• Environmental management	• Pharmacology and toxicology
• Environmental protection	• Physiology and related sciences
• Environmental sciences	• Veterinary medicine
• Evolutionary biology	

Infobox 1: Biology-related university programmes.

Instructors can, of course, aim for other target groups or survey settings but should remain aware that data will not be directly comparable to that of the initial study that employed the EEQ (Kuschmierz et al., in 2021).

b) Instructors who wish to apply the EEQ in different contexts should utilise the following guidelines:

- The assessment of socio-demographic data can be adjusted.
- The single scales and instruments of the EEQ (KAEVO 2.0, ATEVO, PERF and SD) can be used separately and selectively (e.g. only KAEVO 2.0 and ATEVO) based on the instructor's needs.
- The three parts of the KAEVO 2.0 (A, B and C) can be used separately based on the instructor's needs.
- The EEQ can be transferred to an online format. However, the items in KAEVO 2.0, Part C are not directly transferrable. Therefore, the best option is to use a slide bar.
- Depending on the country or region where the survey is conducted, instructors might aim to survey multiple areas to account for variation in the geographical distribution of denominations and other variables.

To ensure comparable survey situations, survey instructions are provided for instructors who conduct the EEQ surveys in classrooms or university courses (see Infobox 2).

Instructions for survey conductors:

The aim of our study is to measure attitudes towards and understanding of evolution across Europe. We survey freshmen students to gather insight regarding their upper secondary education learning outcomes. The results of this survey will provide a more detailed understanding of the reasons for acceptance or rejection of evolution and relations between various factors.

It is important that the students follow your instructions and work alone, so please be very clear and strict. Additionally, please refrain from answering student questions regarding the questionnaire. If the students do not understand a question, they should write their problem beside the question in the questionnaire.

If you notice any conspicuous behaviour or special occurrences, please take note and inform the principal investigator about them.

Completing the questionnaire should take approximately 30 minutes.

Please provide students with the following instructions:

1. This questionnaire has been developed in the context of European research on evolution. Of course, the survey will be anonymous, and we will treat the data with care.
2. The questionnaire is printed double-sided.
3. Unless otherwise stated in the question, check only one answer per question.
4. Please work alone.
5. Please read all instructions carefully before answering the questions.
6. In Sections A, B and C, which are about knowledge, please mark what you think is the scientifically accurate point of view.
7. If you do not know the correct answer, please mark 'I don't know'. Do not guess an answer or copy it from a neighbour.
8. Thank you in advance for your cooperation.

Infobox 2: Instructions for survey conductors.

4. Data preparation and application of statistical analyses

4.1 Data preparation

If applied as a pen-and-paper survey, the results of the EEQ must be digitised. For this purpose, we constructed a spreadsheet for data entry (see Appendix 8.25). Within the spreadsheet, a legend explains the method for transferring all questionnaire answers to the spreadsheet (e.g. which numbers represent which denominations). This legend should also be used when converting the EEQ to a digital survey for online administration.

The items of the KAEVO 2.0 instrument can be coded in three ways, depending on the objective of the analysis:

a) Focus on all items to create a knowledge score:

Items are coded based on the scientific accuracy of the chosen answer, with participants who stated that they did not know the answer and those who offered incorrect answers combined:

- Correct answer: 1
- All incorrect answers and 'don't know' responses: 0

This method may be employed to display the results of the entire instrument as a KAEVO 2.0 score and for the KAEVO 2.0 subscales (see calculation instructions), respectively, and to calculate correlations with other variables.

b) Focus on single items regarding scientific accuracy:

Items are coded based on the scientific accuracy of the chosen answer while distinguishing between participants who stated that they did not know the answer and those who offered incorrect answers:

- Correct answer: 1
- All incorrect answers: 2
- 'Don't know' responses: 3

This method may be employed to display the results of single items.

c) Focus on single items regarding misconceptions:

Items are coded based on the different misconceptions represented in the chosen answers; participants who stated that they did not know the answer are combined in an additional category:

- Assign one number to each misconception and code answers based on their inclusion of any misconception(s).
- Assign an additional category for 'Do not know' responses.

This method may be used to display the results of single items while emphasising the share of participants holding specific misconceptions.

See Appendix 8.24 for a questionnaire version with solutions for all KAEVO 2.0 items.

The EEQ contains several items that are reverse-worded (e.g. expressing rejection of evolution, while most items of the scale express acceptance of evolution). These negatively worded items must be recoded after data entry. In general, all rating scales (Parts D-F) are coded as 1-5 (agree-disagree).

High scores indicate a strong manifestation of the respective trait (D: high dualistic thinking; E: high acceptance of evolution; F: high personal religious faith).

Therefore, the following items require recoding:

- SD scale: D1, D3
- ATEVO scale: E1, E2, E3, E5, E6, E7
- PERF scale: all items (F1-F10)

The same is true for two rating scale items that assess socio-demographic data and are coded as 1-5 (Item 5) or 1-7 (Item 8). Therefore, the following items require recoding:

- Socio-demographic data: Item 5
- Socio-demographic data: Item 8

Exclusion criteria:

Since we aimed to measure students' upper secondary education learning outcomes in our initial pan-European comparative study (Kuschmierz et al., 2021), we applied strict data exclusion criteria to maintain focus on freshmen university students. Including students who were too old or whose graduation was long ago could have impacted the results. For researchers who aim to collect data comparable to our survey data, we recommend the same exclusion criteria:

- Exclude cases with an age of > 25 years
- Exclude cases with graduation from upper secondary school > 2 years before completing the questionnaire
- Exclude countries with $N < 150$

4.2 Calculation instructions

To interpret the results and provide a foundation for further analyses, we built individual scores for every measured parameter. To build a score for the KAEVO 2.0 instrument, apply the data preparation described in 4.1a.

For ATEVO, SD and PERF scores, we summed the individual scores of all items (1–5) on the respective scale. ATEVO scores range from 8–40 (while scores of ATEVO subscales range from 4–20), PERF scores range from 10–50 and SD scores range from 5–25.

For ATEVO, SD and PERF scale reliability analyses, Cronbach's α should be applied to ensure local instrument reliability. For KAEVO 2.0, reliability analyses cannot be applied because the KAEVO 2.0 instrument represents several evolutionary biology concepts (see Kuschmierz et al., 2020b). We recommend conducting a reliability analysis via Cronbach's α for the four items regarding evolutionary adaptation and natural selection (Items A1, A3, A5 and A6).

We also recommend testing the dimensionality of the single constructs within the EEQ. The evolutionary concepts within KAEVO 2.0 represent separate factors of the construct 'knowledge about evolution'. To test for local validity of these dimensions, confirmatory factor analyses may be conducted according to Kuschmierz et al. (2020b).

Depending on the sampling group, the ATEVO scale can either reveal a one- or two-dimensional structure (Beniermann, 2019). Therefore, we recommend conducting principal component analyses (PCA) to test for dimensionality before conducting further statistical analyses. If two dimensions result, these dimensions should represent the two ATEVO subscales, which reflect distinct constructs. However, this division into two distinct constructs is only expected when a majority of the survey population differs in their views on evolution in general compared to their views on evolutionary epistemology (Beniermann, 2019). When performing correlational and multivariate analyses, PCA results can inform decisions to combine item values into one score for the whole scale (PCA one-dimensional) or to maintain two separate scores for both subscales (PCA two-dimensional).

The PCA of the SD and PERF scales should, in both cases, reveal a one-dimensional structure.

To test relations between the constructs within the EEQ, correlation or regression analyses of the scale scores may be applied.

To illustrate the results of the whole sample stemming from KAEVO-C 2.0 (deep time tasks 1 and 2), we suggest employing heat maps (see Fig. 2). To prepare the heat map, divide the timeline into 29 sections, with a length of 9 mm each. While the entire timeline represents the age of the Earth (about 4.54 billion years), one section reflects about 158.6 million years. The frequencies of the data for each millimetre of the time axis are reflected by the colours in the heat map. Thus, the heat maps display participants' perceptions about the existence of dinosaurs and humans in a colour gradient.

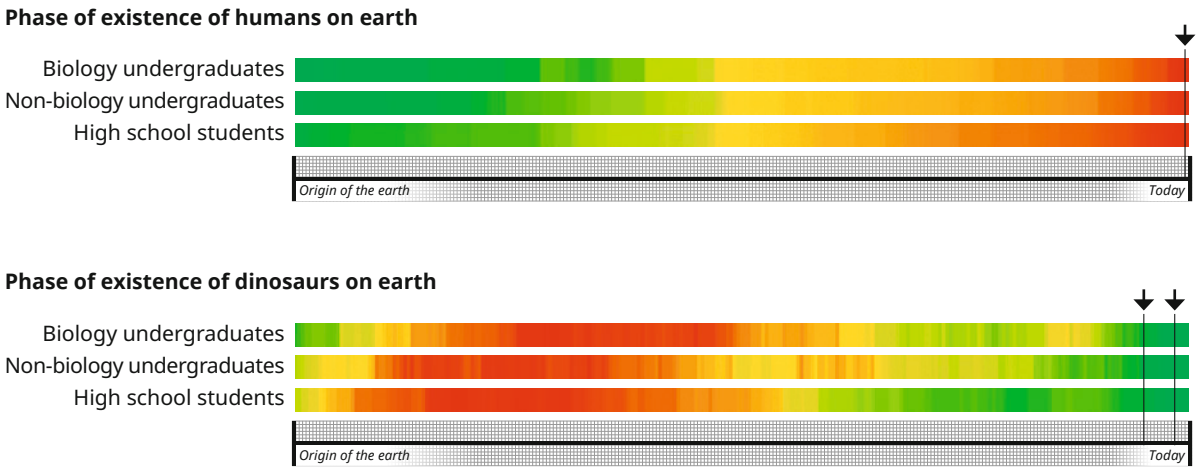

Figure 2: Heat map example for deep time tasks 1 and 2 (Kuschmierz et al., 2020b). A heat map (total length: 26 cm) ranging from green (no answers in this section) to red (greatest number of answers in this section). The scientifically accurate classification is indicated by the black arrows (the existence of humans: 25.9 cm [200,000 years ago] – today; the existence of dinosaurs: 24.7 cm [235 m years ago] – 25.6 cm [65 m years ago]).

4.3 Interpretation of results

Previous studies have published suggestions for interpreting the results for all constructs within the EEQ.

Table 1 displays suggestions for interpreting scores from the KAEVO 2.0 instrument and its subscale scores based on score categories. Score categories facilitate comparisons between different data sets. Variations in the interpretations of the results for the KAEVO 2.0 subscales stem from the fact that the true/false items of KAEVO-B 2.0 have a high guessing rate.

Table 1: Score categories for KAEVO 2.0, KAEVO-A 2.0 and KAEVO-B 2.0 (based on Kuschmierz et al., 2020b).

KAEVO 2.0 score (whole scale)	KAEVO-A subscale score	KAEVO-B subscale score	Interpretation
23–24	12	12	High knowledge
19–22	10–11	11	Rather high knowledge
15–18	8–9	9–10	Moderate knowledge
11–14	6–7	7–8	Low knowledge
0–10	0–5	0–6	Very low knowledge

KAEVO-C 2.0 provides a more qualitative measure for participants' perceptions of deep time. In previous studies, most students situated all three events (the phase of the existence of humans on Earth, the phase of the existence of dinosaurs on Earth and the origin of life on Earth) far too early in the history of Earth. Displaying the results from these tasks within heat maps makes it possible to evaluate whether participants place these events in the correct relative order and how they estimate the scale of time between them. In contrast to multiple-choice scales, this timeline without absolute dates allows participants to express conceptions of deep time events.

Score categories for the interpretation of results for the ATEVO, PERF and SD scales are displayed in Tables 2–4.

Table 2: Score categories for ATEVO and subscales ATEVO-EM and -EG (based on Beniermann, 2019).

ATEVO score (whole scale)	ATEVO score (subscales)	Interpretation
35–40	18–20	Acceptance
29–34	15–17	Rather acceptance
20–28	10–14	Indifferent position
14–19	7–9	Rather rejection
8–13	4–6	Rejection

Table 3: Score categories for SD scale (based on Beniermann, 2019).

SD score	Interpretation
22–25	Strong dualistic thinking
18–21	Dualistic thinking
13–17	Indifferent position
9–12	Monistic thinking
5–8	Strong monistic thinking

Table 4: Score categories for PERF scale (based on Beniermann, 2019).

PERF score	Interpretation
43–50	High religious faith
35–42	Rather high religious faith
26–34	Indifferent position
18–25	Low religious faith
10–17	Very low religious faith

5. Further information on limitations and possible future adjustments

To increase the consistency and clarity of the questionnaire and its application, we collected suggestions regarding limitations and possible future adjustments of the questionnaire and its administration from all authors. In the following sections, we present information regarding issues of content, translation and administration.

5.1 Content and translations of the EEQ

Since the content of the KAEVO 2.0 is based on the German biology curriculum (see Kuschmierz et al., 2020b), the instrument may not address some concepts of evolutionary biology that are of interest for educators (e.g. genetic drift, special focus on human evolution). Future versions of the KAEVO could include additional concepts of evolutionary biology while maintaining the initial structure and content of the instrument.

For additional translations, it is helpful to compare existing translations to identify cases in which the translators speak several languages. In most cases, for the presented translations, we employed the method of translation/back-translation and subsequent discussion between translators. For some languages, the Translation–Review–Adjudication–Pre-test–Documentation (TRAPD) method (European Social Survey, 2014) was applied. For future translations, we recommend using the TRAPD method since it is generally more suitable to identify problematic translations (Harkness, 2003).

Questions regarding socio-demographic data vary between countries due to country-specific differences (e.g. adapting to official terms for religious denominations). Depending on the country, survey group and aim of the study, the socio-demographic data may be adapted further.

Other specific translation issues include the following:

- The Dutch translation of 'mind' (i.e., 'geest') in the SD scale has a connotation of something supernatural. There is, however, no more suitable alternative in Dutch.
- The word 'ape' in Item B3 is not directly translatable to Greek. The Greek translation interprets the term 'ape' as 'monkey' (i.e., 'πίθηκος'), but the language lacks a more appropriate term.
- The Spanish name for banded snails is lengthy and can be misleading (i.e., 'caracoles'); therefore, we used the translation for 'snails' instead of 'banded snails'.

5.2 Administration of the EEQ

Depending on the setting, online administration of the questionnaire may be more suitable. The advantages of online administration include greater sustainability and lower costs. Online administration may also facilitate the involvement of other educators or even countries. On the other hand, it may be more difficult for instructors to motivate students and ensure that they concentrate when they complete the survey at home. In addition, it is more difficult to ensure that only the target group completes the questionnaire. Most of these problems could be solved via online questionnaire administration in a controlled setting (in school or university during teaching time with someone supervising), where possible.

For researchers who aim for group comparisons, we emphasise including a balance of biology-related and non-biology-related students in the sample population. For a comparative survey between countries, the ratio of the biology-related and non-biology-related students should be more or less equal across all countries in the sample. An alternative approach could focus on only one special interest group (e.g. only biology or biology-related students). In any case, it is advisable to specify the group of 'non-biology university students' since differences may exist between and among non-biology students from scientific fields of study (e.g. chemistry, geology) and non-biology students from non-scientific fields of study (e.g. law, theology).

6. References

AERA (American Educational Research Association), American Psychological Association, National Council on Measurement in Education (2014). *The standards for educational and psychological testing.* American Psychological Association.

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching, 39*(10), 952–978. https://doi.org/10.1002/tea.10053

Anderson, D., Fisher, K., & Smith, M. (2010). Support for the CINS as a diagnostic conceptual inventory: Response to Nehm and Schonfeld. *Journal of Research in Science Teaching, 47*(3), 354–357.

Astley, J., & Francis, L. J., (2010). Promoting positive attitudes towards science and religion among sixth-form pupils: Dealing with scientism and creationism. *British Journal of Religious Education, 32*(3), 189–200. https://doi.org/10.1080/01416200.2010.498604

Barnes, M. E., Dunlop, H. M., Holt, E. A., Zheng, Y., & Brownell, S. E. (2019). Different evolution acceptance instruments lead to different research findings. *Evolution: Education and Outreach, 12*(1), 4. https://doi.org/10.1186/s12052-019-0096-z

Beniermann, A. (2019). *Evolution—von Akzeptanz und Zweifeln: Empirische Studien über Einstellungen zu Evolution und Bewusstsein [Evolution—about acceptance and doubts: Empirical studies on attitudes towards evolution and consciousness].* Springer Spektrum. https://doi.org/10.1007/978-3-658-24105-6

Bishop, B. A., & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching, 27*(5), 415–427. https://doi.org/10.1002/tea.3660270503

Catley, K. M., & Novick, L. R. (2009). Digging deep: Exploring college students' knowledge of macroevolutionary time. *Journal of Research in Science Teaching, 46*(3), 311–332. https://doi.org/10.1002/tea.20273

European Social Survey (2014). *ESS Round 7 Translation Guidelines.* ESS ERIC Headquarters, Centre for Comparative Social Surveys, City University London.

Fiedler, D., Tröbst, S., & Harms, U. (2017). University students' Conceptual knowledge of randomness and probability in the contexts of evolution and mathematics. *CBE—Life Sciences Education, 16*(38), 1–16. https://doi.org/10.1187/cbe.16-07-0230

Gregory, T. R. (2009). Understanding natural selection: Essential concepts and common misconceptions. *Evolution: Education and Outreach, 2*(2), 156–175. https://doi.org/10.1007/s12052-009-0128-1

Ha, M., Haury, D. L. & Nehm, R. H. (2012). Feeling of certainty: Uncovering a missing link between knowledge and acceptance of evolution. *Journal of Research in Science Teaching, 49*(1), 95–121. https://doi.org/10.1002/tea.20449

Harkness, J. (2003). Questionnaire Translation. In J. Harkness, F. van de Vijver, & P. Mohler (Eds.), *Cross-Cultural Survey Methods* (pp. 35–56). Wiley.

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology: Issues, News, and Reviews, 12*(3), 123–135. https://doi.org/10.1002/evan.10110

Ingram, E. L., & Nelson, C. E. (2006). Relationship between achievement and students' acceptance of evolution or creation in an upper-level evolution course. *Journal of Research in Science Teaching, 43*(1), 7–24. https://doi.org/10.1002/tea.20093

Johnson, R. L., & Peeples, E. E. (1987). The role of scientific understanding in college: Student acceptance of evolution. *The American Biology Teacher, 49*(2), 93–98. https://doi.org/10.2307/4448445

Kalinowski, S. T., Leonard, M. J., & Taper, M. L. (2016). Development and validation of the conceptual assessment of natural selection (CANS). *CBE—Life Sciences Education, 15*(4), 1–11. https://doi.org/10.1187/cbe.15-06-0134

Kampourakis, K., & Strasser, B. J. (2015). The evolutionist, the creationist, and the 'unsure': picking-up the wrong fight? *International Journal of Science Education, Part B, 5*(3), 271–275. https://doi.org/10.1080/21548455.2014.931613

Konnemann, C., Asshoff, R. & Hammann, M. (2012). Einstellungen zur Evolutionstheorie: Theoretische und messtheoretische Klärungen [Attitudes towards evolutionary theory and Evolution and clarifications regarding measurement theory]. *Zeitschrift für Didaktik der Naturwissenschaften, 18,* 55–79.

Kuschmierz, P., Beniermann, A., Bergmann, A., Pinxten, R., Aivelo, T., Berniak-Woźny, J., Bohlin, G., Bugallo-Rodriguez, A., Cardia, P., Cavadas, B. F. B. P., Cebesoy, U. B., Cvetković, D. D., Demarsy, E., Đorđević, M. S., Drobniak, S. M., Dubchak, L., Dvořáková, R. M., Fančovičová, J., Fortin, C., … & Graf., D. (2021). *European first-year university students accept evolution but lack substantial knowledge about it: a standardized European cross-country assessment.* Manuscript submitted for publication.

Kuschmierz, P., Beniermann, A., & Graf, D. (2020b). Development and evaluation of the knowledge about evolution 2.0 instrument (KAEVO 2.0). *International Journal of Science Education, 42*(15), 1–29. https://doi.org/10.1080/09500693.2020.1822561

Kuschmierz, P., Meneganzin, A., Pinxten, R., Pievani, T., Cvetković, D., Mavrikaki, E., Graf, D., & Beniermann, A. (2020a). Towards common ground in measuring acceptance of evolution and knowledge about evolution across Europe: A systematic review of the state of research. *Evolution: Education and Outreach, 13*(1), 1–24. https://doi.org/10.1186/s12052-020-00132-w

Lombrozo, T., Thanukos, A., & Weisberg, M. (2008). The importance of understanding the nature of science for accepting evolution. *Evolution: Education and Outreach, 1*(3), 290–298. https://doi.org/10.1007/s12052-008-0061-8

McCain, K., & Kampourakis, K. (2018). Which question do polls about evolution and belief really ask, and why does it matter?. *Public Understanding of Science, 27*(1), 2–10. https://doi.org/10.1177%2F0963662516642726

Mead, L. S., Kohn, C., Warwick, A., & Schwartz, K. (2019). Applying measurement standards to evolution education assessment instruments. *Evolution: Education and Outreach, 12*(1), 1–14. https://doi.org/10.1186/s12052-019-0097-y

Nadelson, L. S., & Southerland, S. A. (2009). Development and preliminary evaluation of the measure of understanding of macroevolution: Introducing the MUM. *The Journal of Experimental Education, 78*(2), 151–190. https://doi.org/10.1080/00220970903292983

Nehm, R. H., Beggrow, E. P., Opfer, J. E., & Ha, M. (2012). Reasoning about natural selection. Diagnosing contextual competency using the ACORNS instrument. *The American Biology Teacher, 74*(2), 92–98. https://doi.org/10.1525/abt.2012.74.2.6

Nehm, R. H., & Ha, M. (2011). Item feature effects in evolution assessment. *Journal of Research in Science Teaching, 48*(3), 237–256. https://doi.org/10.1002/tea.20400

Nehm, R. H., & Mead, L. S. (2019). Evolution assessment: Introduction to the special issue. *Evolution: Education and Outreach, 12*(7). https://doi.org/10.1186/s12052-019-0098-x

Nehm, R. H., & Schonfeld, I. S. (2008). Measuring knowledge of natural selection: A comparison of the CINS, an open-response instrument, and an oral interview. *Journal of Research in Science Teaching, 45*(10), 1131–1160. https://doi.org/10.1002/tea.20251

Novick, L. R., & Catley, K. M. (2012). Assessing students' understanding of macroevolution: Concerns regarding the validity of the MUM. *International Journal of Science Education, 34*(17), 2679–2703. https://doi.org/10.1080/09500693.2012.727496

Paz-y-Miño-C, G., & Espinosa, A. (2012). Educators of Prospective Teachers Hesitate to Embrace Evolution Due to Deficient Understanding of Science/Evolution and High Religiosity. *Evolution: Education and Outreach, 5*(1), 139–162. https://doi.org/10.1007/s12052-011-0383-9

Perez, K. E., Hiatt, A., Davis, G. K., Trujillo, C., French, D. P., Terry, M., & Price, R. M. (2013). The EvoDevoCI: A Concept Inventory for Gauging students' understanding of evolutionary developmental biology. *CBE—Life Sciences Education, 12*(4), 665–675. https://doi.org/10.1187/cbe.13-04-0079

Popper, K. (1985). Evolutionary Epistemology. In G. Tarozzi & A. Merwe (Eds.), *A New International Book Series on The Fundamental Theories of Physics: Their Clarification, Development and Application: Vol. 10. Open Questions in Quantum Physics: Invited Papers on the Foundations of Microphysics* (pp. 395–413). Springer.

Price, R. M., Andrews, T. C., McElhinny, T. L., Mead, L. S., Abraham, J. K., Thanukos, A., & Perez, K. E. (2014). The genetic drift inventory:. A tool for measuring what advanced undergraduates have mastered about genetic drift. *CBE—Life Sciences Education, 13*(1), 65–75. https://doi.org/10.1187/cbe.13-08-0159

Rughiniş, C. (2011). A lucky answer to a fair question: Conceptual, methodological, and moral implications of including items on human evolution in scientific literacy surveys. *Science Communication, 33*(4), 501–532. https://doi.org/10.1177%2F1075547011408927

Rutledge, M.L., & Warden, M.A. (1999). The development and validation of the Measure of Acceptance of the Theory of Evolution instrument. *School Science and Mathematics, 99*(1), 13–18. https://doi.org/10.1111/j.1949-8594.1999.tb17441.x

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. https://doi.org/10.1007/s11191-016-9836-4

Smith, M. U., Snyder, S. W., & Devereaux, R. S. (2016). The GAENE—Generalized Acceptance of EvolutioN Evaluation: Development of a new measure of evolution acceptance. *Journal of Research in Science Teaching, 53*(9), 1289–1315. https://doi.org/10.1002/tea.21328

Stanovich, K. E. (1989). Implicit philosophies of mind: The dualism scale and its relation to religiosity and belief in extrasensory perception. *The Journal of Psychology, 123*(1), 5–23. https://doi.org/10.1080/00223980.1989.10542958

Taber, K. S., Billingsley, B., Riga, F., & Newdick, H. (2011). Secondary students' responses to perceptions of the relationship between science and religion: Stances identified from an interview study. *Science Education, 95*(6), 1000–1025. https://doi.org/10.1002/sce.20459

Voland, E. (2010). Die Evolution der Religiosität. Hat Gott Naturgeschichte? [The evolution of religiousness. Is there a natural history of God?] *Biologie in unserer Zeit, 40*(1), 29–35. https://doi.org/10.1002/biuz.201010411

Wagler, A., & Wagler, R. (2013). Addressing the lack of measurement in-variance for the Measure of Acceptance of the Theory of Evolution. *International Journal of Science Education, 35*(13), 2278–2298. https://doi.org/10.1080/09500693.2013.808779

Wuketits, F. M. (1990). *Evolutionary epistemology and its implications for humankind. SUNY series in philosophy and biology.* State University of New York.

7. Contributions

Initial discussions on questionnaire constructions: Anna Beniermann, Umran Betul Cebesoy, Momir Futo, Tamara Milosevic, Naïd Mubalegh, Rianne Pinxten, Marie Rahn, Barbora Trubenova and Jelle Zandveld (at Evoke Conference, Porto 2017)

Preliminary literature analysis: Anna Beniermann and Rianne Pinxten

Development and design of the questionnaire: Paul Kuschmierz, Anna Beniermann and Dittmar Graf, in addition to peer feedback from the entire WG1

Information on instructions and protocols for data preparation: Paul Kuschmierz and Anna Beniermann

Information on analysis and interpretation: Anna Beniermann and Paul Kuschmierz

The principal investigator on the initial study with the EEQ: Paul Kuschmierz

Writing of the current method report: Anna Beniermann

Translations and adjustment information: All authors

8. Appendices: Questionnaire in European languages & solutions