Malthus for kids: The impact of exploring Malthus' principle on elementary school students' understanding of evolution by natural selection.

Xana Sá-Pinto[1*], Alexandre Pinto[2], Joana Ribeiro[2], Inês Sarmento[2], Patrícia Pessoa[1,3], Leonor R. Rodrigues[4], Lucía Vázquez-Ben[5], Evangelia Mavrikaki[6] and J. Bernardino Lopes[3,1]

1- Research Centre Didactics and Technology in Education of Trainers; University of Aveiro, Campus Universitário de Santiago; 3810-193 Aveiro, Portugal. e-mail: xanasapinto@gmail.com

2- Polytechnic Institute of Porto School of Education, Rua Dr. Roberto Frias, 712, 4200-465 Porto, Portugal.

3- University of Trás-os-Montes e Alto Douro, Quinta de Prados, 5000-801 Vila Real, Portugal.

4- Centre for Ecology Evolution and Environmental Changes, Faculty of Sciences, University of Lisbon, Edifício C2, 3º piso, 1749-016 Lisboa, Portugal

5- Faculty of Educational Studies, University of Coruña, Department of Pedagogy and Didactics, Campus de Elviña s/n, 15071 A Coruña, Spain

6- Faculty of Primary Education, National & Kapodistrian University of Athens, Navarinou 13a, 10680 Athens, Greece.

* - Corresponding author

Abstract:

While several researchers have suggested that evolution should be explored from the initial years of schooling, little information is available on effective resources to enhance elementary school students' level of understanding of evolution by natural selection (LUENS). For the present study, we designed, implemented and evaluated an educational activity planned for fourth graders to explore concepts and conceptual fields that were historically important for the discovery of natural selection. Observation field notes and students' productions were used to analyse how the students explored the proposed activity. Additionally, an evaluation framework consisting of a test, the evaluation criteria and the scoring process was applied in two fourth-grade classes to estimate elementary school students' LUENS before and after engaging in the activity. Our results suggest that our activity allowed students to effectively link all of the key concepts in the classroom and produced a significant increase in their LUENS. These results indicate that our activity had a positive impact on students' understanding of natural selection. They also reveal that additional activities and minor fine-tuning of the present activity are required to further support students' learning about the concept of differential reproduction. We also observed a low level of teleological predictions for both pre- and post-tests.

Keywords: evolution education, teleological thinking, conceptual fields, history of science.

Introduction:

In 1798, Malthus published a famous essay in which he explored how species' reproductive potential to grow geometrically is affected by the constant (or in the case of humans, arithmetically growing) resources available for subsistence (Malthus, 1798). Inspired by this idea, Darwin and Wallace proposed that species evolved through natural selection 60 years later (Darwin & Wallace, 1858).

Given its great explanatory power, evolution soon became a central concept in biology since it provides a framework that allows us to make sense of and link facts and knowledge from distinct subdisciplines and make predictions about biological organisms and systems (Dobzhansky, 1973). Despite its fundamental importance to biology and many other research fields, several studies have shown that evolution is not understood—or even accepted—as a valid scientific theory by many, with frequent and persistent misconceptions being shared by people across several countries, ages, instructional levels and developmental stages (Bishop & Anderson, 1986; Rutledge & Warden, 2000; Miller et al., 2006; Nehm & Reilly, 2007; Prinou et al., 2011; Spiegel et al., 2012; Athanasiou & Mavrikaki, 2013; Yasri & Mancy, 2016; To et al., 2017).

This is particularly worrying since understanding evolution is fundamental to understanding the surrounding world, making informed choices and tackling personal and societal problems (National Research Council [NRC], 2012; Carrol et al., 2014). To overcome this problem, the NRC (2012) proposed evolution as one of the four core concepts in biology that should be explored since kindergarten and across students' entire educational routes with increasing complexity. Despite the NRC's (2012) recommendation, little information is available regarding what students in elementary schools can learn about evolution, their knowledge and misconceptions about this topic, or effective strategies to teach evolution at such young ages. Moreover, few studies have analysed elementary school students' understanding of evolution

(Nadelson, 2009; Campos and Sá-Pinto, 2013; Kelemen et al., 2014; Shtulman, Neal and Lindquist, 2016; Berti, Barbetta and Toneatti, 2017; Emmons, Lees and Kelemen, 2017; Sá-Pinto et al., 2017a, Brown et al., 2020, Frejd et al., 2020), while even fewer studies have explored their understanding of natural selection (Kelemen et al., 2014; Shtulman et al., 2016; Berti et al., 2017; Emmons et al., 2017; Sá-Pinto et al., 2017a; Brown et al., 2020; Frejd et al., 2020).

Notably, discordant results were obtained in these studies regarding elementary school students' ability to learn about natural selection after educational interventions. Campos and Sá-Pinto (2013), Kelemen et al. (2014), Emmons et al. (2017), Brown et al. (2020) and Frejd et al. (2020) reported that elementary school students were able to understand and apply the principle of natural selection to explain and predict biological evolution following pedagogic interventions.

However, in a study that tested a distinct pedagogical sequence, Berti et al. (2017) reported that only a minority of children were able to learn about natural selection. These results highlight the need for further studies analysing elementary school students' ability to learn about evolution by natural selection. Together, these studies will facilitate the design of effective strategies to promote such learning.

Research in evolution education shows us that unlike experts, novices tend to be sensitive to the superficial features of a situation/problem (Nehm & Ridgway, 2011). For conceptually equivalent problems, students may provide different sets of normative and non-normative ideas about evolution if these have distinct surface features (e.g., if the same problem is presented with animals evolving distinct traits or a plant is used instead) since these features activate distinct mental representations that will subsequently activate distinct concepts and problem-solving schemas (reviewed in Nehm, 2018).

An approach to overcome this problem involves allowing students to explore evolution in various biological scenarios, species and contexts by emphasising the common features of these situations while simultaneously highlighting their differences in terms of surface features (Nehm, 2018). Aligned with this view, Vergnaud (2009) argued that learning requires the development of conceptual fields, which he understands as a set of situations—that may be explored in different educational activities—and a set of linked concepts. Concepts and situations are tightly linked: a given situation can only be fully understood by applying and linking certain concepts, while the meaning of a concept can only be learnt by exploring a variety of distinct situations that highlight the set of a concept's invariants (i.e., objects, properties and relationships) that allow students to apply it to make sense of new situations and solve new problems (Vergnaud, 2009).

This also emphasises the need to have a set of good examples and educational activities that expose students to distinct situations involving evolution by natural selection that allow them to identify the concepts' invariants from surface features and promote evolution understanding. This need contrasts with the scarcity of educational activities described to promote evolution understanding in elementary school students.

A comparison between the historical development of scientific ideas and students' conceptual progression revealed striking similarities between them, including in evolution (see Ha and Nehm, 2014 and references therein). This led several authors to design educational activities inspired by important events in the history of science, which were shown to be effective in fostering students' learning (see Dedes & Ravanis, 2009 and references therein). Although students' learning processes and the development of ideas among 19th-century scientists are not exactly comparable (Ha and Nehm, 2014), the history of science may be used to identify the processes, contexts, concepts and conceptual fields that were important for scientific discoveries and inspire the design of educational activities.

In his autobiography, Darwin described how facing distinct situations during the Beagle's voyage and after returning to England allowed him to develop his conceptual field related to evolution (Barlow, 1958). After returning to England, Darwin collected data and information from diverse sources about variation in wild and domestic animals and plants (Barlow, 1958). However, according to Darwin, the discovery of the process of natural selection only took place in October 1838, when he "*happened to read for amusement 'Malthus on Population', and being well prepared to appreciate the struggle for existence (...) it at once struck me that under these circumstances, favourable variations would tend to be preserved, and unfavourable ones to be destroyed. The result of this would be the formation of new species*" (Barlow, 1958, p.120). This sentence reveals the importance of contrasting the potential for the geometrical growth of natural populations with the constant or arithmetical growth of subsistence (i.e., the basis of Malthus' (1798) essay on population) for Darwin to devise and operationalise the concept of natural selection. In support of this hypothesis, Wallace (Darwin & Wallace, 1858) used species' potential for geometrical growth to depict the "*struggle for existence*" and to describe the evolutionary process that Darwin called natural selection. Both of these observations suggest that understanding the concept of natural selection may be facilitated by exploring Malthus' principle.

While the specific situations Darwin and Wallace faced during their lives allowed them to discover evolution by natural selection, these are largely impossible to replicate in the classroom. Instead, we can design educational activities that would require students to explore situations that address Malthus' principle and to put in action concepts and conceptual fields that were important to the scientific discovery of natural selection. Therefore, our research question is: Will educational activities that require students to explore situations addressing Malthus' principle and put in action historically important concepts and conceptual fields effectively promote students' learning on evolution? To answer this question, we aimed to: *i)* design an educational activity that uses a situation developed for elementary school students to explore Malthus' principle and put in action key concepts and conceptual fields similar to those that were historically important for the scientific discovery of natural selection; *ii)* evaluate the impacts of the designed activity on students' evolution understanding.

Materials and methods:

To achieve our goals, we opted to use design research. This methodological approach (or series of approaches, as proposed by Barab and Squire, 2004) stems from the need for educational research to return to educational practice in order to inform education activity and policy (Van den Akker et al., 2006). According to Kelly (2013), this approach is particularly appropriate for studying wicked and open problems, where little is known about how to teach the content or there is a lack of effective instructional strategies and/or materials, for example.

Therefore, design research consists of designing and implementing (new) interventions aimed at solving a complex educational problem to either gain knowledge about the process of intervention design and development itself and/or validate (in an exploratory sense) new theories (Plomp, 2013).

Accordingly, two types of design research can be considered (i.e., development *vs* validation studies) and two principal outcomes can be obtained (i.e., design principles/local theory and empirically underpinned innovative interventions). However, both orientations could be combined since they share the interventionist, utility, and process-oriented nature of design research while also building on prior research, being cyclical and holistic, and involving practitioners (Plomp, 2013; Van den Akker et al., 2006).

Therefore, we present a study aimed at developing a research-based solution to improve natural selection understanding among elementary school students while also validating the domain-specific instruction theories underlying such learning processes. Consequently, two products result from our research: a transdisciplinary problem-based learning (PBL) activity and new insights into elementary students' understanding of natural selection and their learning processes in light of Malthus' principle. In this process, we joined the efforts of primary school teachers (JR, PP), researchers in science (XSP, AP, JBL, LVB, EM) and mathematical education (IS), and an evolutionary biologist (LR).

Sandoval (2014) recommended being clear and specific about how and why each design research study has been designed to overcome common criticism, such as a lack of accumulative grammar or the inability to both evaluate and test theory simultaneously. He suggested using a "conjecture map", where the embodiments (i.e., tools, materials, discursive practices, etc.) employed in the intervention, its expected outcomes and the mediating processes that are meant to occur in between are shown. Therefore, a conjecture map can contribute to testing the initially formulated hypothesis (or high-level conjecture). Figure 1 introduces our conjecture map, which presents a summary of our design and how its various elements relate to each other. The results presented in this paper only correspond to the first cycle of the design and application of this intervention.

FIGURE 1

Design and implementation of the educational intervention

Basic principles guiding our design

The ability to use natural selection to explain or predict biological situations requires students to understand, articulate and put in action several key concepts. Notably, many researchers in evolution education have listed some of the distinct key concepts involved (Anderson et al., 2002; Nehm & Ridgway, 2011; Tibell & Harms, 2017). We will follow the list of key and threshold concepts recently proposed by Tibell and Harms (2017), who considered published lists of key concepts and then summarised and organised them into main principles. Furthermore, they proposed key concepts that are more generalisable and less sensitive to the surface features of a situation/problem. One such example is the key concept of "selective pressure", which replaces other less generable key concepts such as "competition" and "limited resources", which were presented in Anderson et al. (2002) and Nehm and Ridgway (2011) and merely represent some of the many selective pressures that can cause evolution by natural selection. Finally, unlike Anderson et al. (2002) and Nehm and Ridgway (2011), Tibell and Harms (2017) included differential reproduction as one of the key principles of evolution by natural selection. This is particularly important since differences in fitness among individuals are determined by the differences in their contributions to the next generations' gene pool (Orr, 2009).

To identify which key concepts from the list of concepts by Tibell and Harms (2017) were acknowledged by Darwin as crucial in his development of the theory of natural selection, we searched for evidence in both Darwin's biography (Barlow, 1958) and his initial descriptions of evolution by natural selection (Darwin, 1857; Darwin and Wallace, 1858). This comparison is presented in Table 1.

Notably, we found evidence supporting the notion that Darwin articulated and put in action, most of the key concepts (KCs) proposed by Tibell and Harms (2017), with the exception of two key concepts: speciation and the (genetic) origin of variation (Table 1). Regarding the genetic origin of variation, although Darwin mentions that "*during millions of generations, individuals of a species will be occasionally born with some slight variation*" (Darwin & Wallace, 1858 p.52), he was unaware of the genetic basis and mechanisms behind these variations. Accordingly, the origin of variation was not addressed during the planned educational activity. Although we present a species with variable traits in our activity, the genetic basis of these traits was not discussed further than the traits being heritable. Despite Darwin mentioning speciation in his initial 1857 letter to Asa Gray (Darwin & Wallace, 1858) in his autobiography, he identifies this discovery as occurring later than the discovery of the process of natural selection (Barlow, 1958). Accordingly, we do not address this concept in the educational activity.

TABLE 1

Malthus' principle is based on mathematical models that describe population growth as a function of resource availability. Therefore, we aimed to design a transdisciplinary activity that would require mathematical and biology skills as well as knowledge to be solved. By designing an interdisciplinary activity that simultaneously explores natural selection and mathematical learning goals, we aimed to: *i)* link biology and maths disciplines and allow students' engagement in mathematical thinking and the development and use of models, which are two scientific practices that students are expected to learn (NRC, 2012); *ii)* allow elementary school teachers to include evolution in teaching—even if this topic is not explicitly in the learning goals of their national curriculum—to increase the likelihood of this concept being explored in these school grades. To further align our didactical proposal with the learning goals typically explored in elementary school classes, we aimed to design activities that further engage students in scientific practices included in Portuguese science standards (Portuguese Government/Ministry of Education, 2018a, 2018b, 2018c, 2018d, 2018e, 2018f) and those of other nations (NRC, 2012, 2013; Greek Government Gazette 303B/13-03-2003).

To engage the students in the activity, we aimed to have at least one practical activity that would promote contact with animals since exploring real animals (either through contact with living animals or through films) was shown to increase students' interest and competence

(Hummel & Randler, 2012). To achieve this, we had to choose an animal species that: *i*) could be easily brought to the classroom for the students to observe and "manipulate"; *ii)* would have at least two varieties with distinct features that result in fitness differences under selective pressures that we could easily manipulate; *iii)* would have a relatively short generation time so that students could easily work with and understand the time scale of the activity because "deep time" has been proven to be a difficult concept for many groups (Catley & Novick, 2009; Cotner, Brooks and Moore, 2010); *iv*) would have a high reproductive output and could easily grow to population sizes larger than the environmental carrying capacity.

The proposed didactic sequence

Our PBL activity consisted of three sessions of 150 minutes each. With the support of the students' teacher in each class, the three sessions took place within one week and were led by research team members experienced in teaching these grades. The aims of each session are detailed in Table 2.

TABLE 2

We used the two-spotted spider mite (*Tetranychus urticae*)—an agricultural pest—as our model organism. This species has a short life cycle (generation time of approximately 13 days), which allowed us to follow evolution over short time scales. Furthermore, individuals of this species are highly fecund, with females laying up to 10 eggs per day over a period of 20–30 days (Wrensch & Young, 1975). Consequently, populations experience exponential growth and quickly deplete their resources, making them ideal for exploring Malthus' principle. Moreover, *T. urticae* displays intraspecific variability, with different populations being adapted to different host plant species (Migeon et al., 2011). The targeted concepts and sessions in which these were explored are described in Table 1.

- Session 1

In the first session, we introduced the model species and students were asked to solve mathematical problems related to size measurements and scales. This allowed students to explore spatial scales, a threshold concept important for evolution understanding according to Tibell and Harms (2017) (see Table 2). Students were asked to individually draw and share what they thought a mite looked like with the class, which uncovered previous conceptions about this species. Students were then invited to observe spider mites using various instruments, without being informed of the magnifications used, so they could collaboratively propose strategies to estimate the size of the spider mites using mathematical thinking. After solving this mathematical problem, students' initial conceptions were compared to their observations.

- Session 2

During the second session, students were introduced to the research group MITE2: Multidisciplinary Investigation Targeting Ecology and Evolution from the Centre for Ecology, Evolution and Environmental Changes based at the University of Lisbon (https://ce3c.ciencias.ulisboa.pt/sub-team/mite2) through a short movie. This research team provided the spider mites used in the activities and the movie guided the students through their laboratories and introduced some of their research projects using this organism. Besides introducing students to an example of how researchers work, the video allowed us to provide a real context for the problem posed to students.

Students were informed that two individuals of one spider mite population that feeds on citrus tree leaves (henceforth referred to as the lemon specialist) and six individuals of another spider mite population that feeds on bean plant leaves (henceforth referred to as the bean specialist) would be sent by the MITE2 research group on that day to be presented to other classes for observation and to perform more experiments. The teacher of the class divided students into smaller groups (between 4 and 6 members) and these groups were asked to work collaboratively to propose strategies to mathematically model the growth of the population and to estimate and graphically represent the number of lemon specialists that were expected to exist in 45 days. The entire class discussed what information regarding species' biology would be needed. After reaching a consensus on the information needed to solve the mathematical problem—and to simplify the mathematical modelling—students were told to consider a sex ratio of 1:1, a generation time equal to a life expectancy of 15 days, and that each female lays approximately 100 eggs, from which 100 individuals are born. Using the aforementioned parameters, students discussed the best strategy to solve this problem in smaller groups and applied it to estimate the solution. Each group then presented the strategy they used and the results they obtained to the class, and all students ultimately discussed and decided on the best strategy to be applied. Each group was asked to estimate, using this method, the number of bean specialists within 45 days and to graphically represent the number of mites of each plant specialist for each 15-day period. While solving this problem—applicable to both plant specialists—students were expected to explore the mites' reproduction (KC5, Table 1) by estimating and graphically representing the geometric population growth expected under an unlimited resource scenario (no selective pressure present). They explored this pattern for two mite populations (KC2, Table 1) that differ in their heritable ability to feed on distinct food sources (KC4, Table 1).

- Session 3

During the third session, students were asked to do the same exercise as in the second session while considering the selective pressure (KC6, Table 1) imposed by resource availability: we could only provide 100 lemon tree leaves and 10 bean leaves per week to all the mites, which would be kept in a single large box. The number of leaves was chosen so the bean specialist, initially most frequent, would have fewer resources available to feed on, thereby changing its representation in the population (K9, Table 1). Once again, students decided what information they required regarding species biology and how could they use it to answer this question via an initial class discussion. Students were told that each leaf (regardless of the plant type) could feed a maximum of 100 mites in a week.

Again, the class teacher divided the students into small groups of 4 to 5 students. Students were then asked to propose a strategy to estimate the number of bean and lemon specialists on this limited resource scenario 45 days later, building from the procedure developed in the previous session. At the end of the activity, students were asked to observe the results of their mathematical model, discuss it in their groups and explain why the least frequent plant specialist had become the most frequent one (KC9, Table 1). Furthermore, they were asked to compare the results obtained in this scenario with those of the unlimited resource scenario and discuss the reasons for the observed differences. Solving the proposed tasks required that students understand that, in the proposed situation, resource availability (KC6, Table 1) resulted in distinct fitness (KC3, Table 1) between the two mite populations due to their differential survival (KC7, Table 1) and reproduction (KC8, Table 1).

Sampled classes

Two classes of fourth-grade students (ages 9–10 years old) from two distinct schools engaged in the previously described didactic activity. The two schools were from the northern region of Portugal. This involved convenience sampling since schools were not chosen randomly. Instead, they were chosen among those with which the research team had worked before in other classes and topics and that had at least two fourth grade classes. SA was a private school located in the centre of a big city in the northern region of Portugal, while SB was a public school located in a more rural area 20 km away from this city. According to publicly available information, most parents with children at SB only completed the 6th grade or below (58, 21 and 15% of parents had completed the 6th grade or below, 9th grade and 12th grade, respectively, while only 6% had instruction beyond the 12th grade) and 82% of the students were included in the 1º and 2º class of family support for social security due to their low family income. No information on parents' academic or income levels was available for SA.

The class in which the activities were implemented and evaluated in SA had 19 students (henceforth referred to as the SAT class), while 25 students attended the SB class (henceforth referred to as the SBT class). No personal information about students was collected since their answers were identified by a code made from their student number, class, and school. Informed consent was obtained from the students' parents, the school boards and teachers before the implementation of the activity and test. The procedures followed were approved by the school boards and are in accordance with the ethical standards of the Ethics and Deontology Council of the University of Aveiro and with the Helsinki Declaration of 1975, as revised in 2008.

Design and application of the evaluation framework

To evaluate students understanding of evolution by natural selection, we adapted and applied an evaluation framework. In the following sections, we describe: *i)* the evaluation instruments upon which we designed our framework; *ii*) the features of our evaluation instrument; *iii*) how the test was applied in the classrooms; *iv)* the procedure used to evaluate and score students' answers.

Evaluation instruments upon which we designed our framework

When we started this project, two evaluation frameworks were available to evaluate elementary school students' understanding of evolution by natural selection: the interview script used by Kelemen et al. (2014) and Emmons, Lees and Kelemen (2017) as well as the test proposed by Sá-Pinto et al. (2017a). Although we could not find information on the preferences of elementary school teachers for performing student evaluations, our lengthy experience and contact with this school grade suggests that these mostly use written tests. To elaborate on an instrument that could also be useful and applied by teachers, we followed Sá-Pinto et al. (2017a) and designed a written test. We also retained some features of this framework that distinguishes it from the one used by Kelemen et al. (2014) and Emmons et al. (2017), namely: *i*) the final outcome of the biological scenario was not provided to the students, which would allow students to reveal fixist ideas; *ii*) unlike Kelemen et al. (2014) and Emmons et al. (2017), Sá-Pinto et al. (2017a) did not ask students any isolated fact questions regarding the trait inheritance, trait constancy, survival or reproduction ability of each phenotype before or after asking them to predict the outcome of the biological scenario to avoid influencing students' predictions and justifications; *iv)* like in Sá-Pinto et al. (2017a), students were informed that the two phenotypes in the test were heritable—without this information, it would be impossible to evaluate how much the phenotypic differences could result from environmentally driven morphological plasticity.

The test and its implementation with students

The test used in this evaluation presented students with a biological scenario similar to the one explored in the educational activity (Table 1): *i*) an isolated population of butterflies (mites in the activity); *ii*) with a variable and heritable trait with two distinct phenotypes that influenced their ability to feed on two distinct food resources (i.e. butterflies with long or short proboscises feeding on flowers with long and short calyxes; bean and lemon specialists eating bean or lemon leaves in the activity); *iii*) the most frequent phenotype would have fewer resources available to feed on (in the activity the bean specialists in a box with more lemon leaves than bean leaves). The test is presented in detail in Figures A1 and A2 of the Appendix.

Students were asked to think forward in time and predict the outcome of this scenario, and then describe how the butterfly population would look in 100 years. The test was read aloud to the class (to overcome reading and interpretation difficulties that some students might have) and students were asked to write a justified prediction and draw it. After finishing these tasks, each student was individually asked to verbally explain her/his predictions and justifications to the researcher and, when the student provided more information at this stage, she/he was asked to complete her/his written answer in the test form. No corrective feedback or additional information was provided by the researcher during this phase. For students with writing difficulties, the answers were provided verbally and registered by the researcher using the students' exact words. This procedure was followed independently of the type of predictions and justifications put forward by the students. In total, between 20 and 30 minutes were required to obtain all of the students' answers for each class. This evaluation procedure was applied immediately before (pre-test) and approximately 20 days after the activity was performed (post-test).

Procedure to evaluate students' answers and score the evaluation criteria

To evaluate students' answers, we used criteria developed by other authors (Kelemen et al., 2014; Sá-Pinto et al., 2017a) in the context of the aforementioned framework. These were complemented with the inclusion of another criterion that targets whether students' predictions to integrate information about selective pressure: resource availability. These criteria formed the items of our rubric. The complete definitions of each rubric item are provided in Table 4. These rubric items allowed us to classify answers according to the student's type of prediction (i.e., fixist, fittest or equilibrium) and the justification provided (i.e., developmental, teleological, resource availability, differential survival or differential reproduction).

- *Fixist* answers predicted that the initially most common (and less fit, if no other biological meaningful justification was provided) phenotype would remain the most common in 100 years;
- *Fittest* answers predicted that the fittest haplotype would become the most frequent in 100 years (predicting a strong frequency change KC9, Table 1);
- *Equilibrium* predictions stated that both phenotypes would become equally frequent in 100 years (predicting a moderate frequency change KC9, Table 1).

The level of understanding of evolution by natural selection (LUENS) revealed by each answer was determined by the sum of the scores attributed for each rubric item identified in that answer, regarding both predictions and corresponding justifications. Population evolution heavily depends on *resource availability* (selective pressure KC6, Table 1) and *differential survival* (KC7, Table 1). Accordingly, we attributed a score of 1 to each of these two rubric items. However, the most important parameter determining evolutionary outcomes is *differential reproduction* (selective pressure KC8, Table 1) since this better correlates with individuals' contributions to the gene pool of the next generation (i.e., individuals' fitness); thus, we attributed a score of 2 to this rubric item. To determine the score of each type of prediction, we estimated the Spearman's correlation coefficient (and its corresponding statistical significance) between them and the rubric items related to evolution (namely *resource availability, differential survival* and *differential reproduction*). These results, depicted in Table 3, mostly confirm those obtained in previous studies (Sá-Pinto et al., 2017a), showing positive and significant correlations between *fittest* predictions and justifications mentioning *resource availability, differential survival* and *differential reproduction* and negative and significant correlations between these three rubric items and *fixist* predictions. While the results of previous studies (Sá-Pinto et al., 2017a) showed that *equilibrium* predictions were negatively and significantly correlated with justifications mentioning *resource availability, differential survival* and *differential reproduction*, no significant correlation was found in the present study. This suggests that students providing *equilibrium* predictions are not relating the frequency changes with biological important parameters, nor thinking evolutionarily. Based on these results we attributed a score of 1 to *fittest* predictions and a score of 0 to *fixist* and *equilibrium* predictions. All other rubric items received a score of 0. Given this score rating, LUENS can range between 0 (for answers with no evidence of evolution understanding) and 5 (for answers with evolutionary predictions justified by all components of the key concepts important to understanding natural selection). The present framework evaluates whether students can apply all KCs related to the principle of selection (Tibell & Harms, 2017; Table 1)—except for speciation since this KC was not addressed in this activity for the aforementioned reasons.

For a detailed explanation of how students' answers were coded, see examples in Figure 2 and Table 4.

FIGURE 2

TABLE 3

TABLE 4

Ensuring the validity of the evaluation instrument

To ensure that the chosen evaluation instrument was valid, we *i)* designed our instrument by adapting a previously validated instrument (Sá-Pinto et al., 2017a), *ii*) ensured that all key concepts required for evolution understanding (Tibell & Harms, 2017) that were explored in our activity were present in our evaluation instrument (Table 1), and *iii*) studied the correlation between the students' predictions and justifications to decide on the scoring procedure. Furthermore, we applied the same test procedure in two control classes, which were classes from the same schools in which we did not apply the aforementioned activity or any evolution-related activity in their educational programme (henceforth referred to as SAC [N=21] and SBC [N=19]) on the same days that the target classes were tested. Target and control classes were chosen by the school director and teachers based on their availability according to the school schedule. We used control classes to check for the impact of the double exposure of students to our test and to evaluate the internal validity of the process (Lahm, 2004). The pre- and post-tests of the two control classes (SAC and SBC) did not significantly differ (Z_{SAC}= - 0.447, p = 0.655 and Z_{SBC}= -1.604, p = 0.109), thus confirming the internal validity of the process (Lahm, 2004). Finally, two independent researchers—one evolutionary biologist with a background in science education (XSP) and one elementary school teacher (PP)—evaluated all of the students' answers. Interrater reliability was estimated as the percentage of the initial agreement between raters (McHugh, 2012). Answers not equally rated by the two researchers were discussed and, if a consensus could not be reached, these were removed from the analysis. Since interrater reliability was >89% for all analysed items, the reliability of this procedure was considered acceptable (Stemler, 2004, p.2).

Data analysis

McNemar and Wilcoxon tests were used to estimate the statistical significance of, respectively, changes in the frequency of each rubric item and students' LUENS between pre- and post-tests. All statistical analyses were performed using SPSS v23. The database housing the results of the students' answers analysis is deposited in the Dryad repository: (to be included after acceptance).

To complement the data collected from students' test answers and characterise the learning processes that occurred in the target classes, we collected field notes during participant observation in the sessions, took photos of students' productions and recorded their discussions. These documents were used to: *i)* analyse how students approached the mathematical modelling process by focusing on the biological parameters they described as important as well as the strategies proposed and implemented to solve the proposed problems and estimate the values; *ii)* look for evidence of the target key concepts (Table 1) being explored during the sessions through content analysis using Tibell and Harms' (2017) key concepts (Table 1) as the categories of analysis.

Results:

Evidence of the mediating processes during the educational activity

Students' engagement

During the three sessions, students were actively engaged in the proposed tasks (see examples of students' engagement in the tasks in Figure 3). They used the materials provided to them and collaboratively (in both small and large groups) proposed, discussed, implemented and revised solutions for the problems and identified the parameters important for population growth, mathematical modelling and calculation strategies to estimate population sizes. In both large and small groups, they also graphically depicted the results. Moreover, they further discussed these results in the large group.

Mathematical modelling and linkage to historically important key concepts

During the first session, students from both classes observed mites using several magnifying devices (e.g., a magnifying glass, tripod magnifier, digital stereo microscope, and magnifying camera; see Figure 3b,c,d). From the magnified image projected on the board, students successfully estimated the size of mites by applying two solutions to the same problem. In both cases, they used an object that was measurable in both the magnified and non-magnified images as a size scale. In SBT, they used a line drawn in the Petri dish by the researchers before the activity for this purpose. Although this line was drawn in all Petri dishes used in both classes, the students in SAT first proposed and used a piece of vegetable material in the petri dish. After finding a solution to determine the spatial scale of magnification, students mathematically estimated the size of the mite (Figure 3e). This allowed students to explore spatial scales and compare their previous ideas (Figure 3a) with real-life observations and their mathematical estimations.

FIGURE 3

In session 2, students estimated the number of mites of each variety that would be available for other students to observe in 45 days. When asked to consider the biological information they would need to make these estimations, students mentioned variables such as the number of progeny a mite could have, their life expectancy, the number of males and females in each population, the time it takes for the eggs to hatch, the age at which individuals start to reproduce (i.e., generation time) and the size of individuals. During the discussion, students decided that individuals' size and hatching time were not necessary to solve the problem. Different groups of students presented various strategies to solve the problem and organise the data to their class, including strategies solely based only on mathematical language and strategies combining this with tables (Figure 4a and 3b, respectively). After identifying and discussing the correct approaches in both classes, the combination of mathematical language with tables was chosen by the students to perform the estimations for the second mite population (KC2, Table 1), reproduction (KC5, assuming the trait heritability KC4, Table 1) and population growth (Figure 4c).

FIGURE 4

In the third session, when asked about the type of biological information they would require to estimate the number of mites of each variety by the end of 45 days, students mentioned two additional parameters: the number of mites that could be accommodated on one leaf and the number of mites that one leaf could feed. In one of the classes (SAT), students initially estimated the space occupied by the mites estimated in the previous session (Figure 5a). However, after estimating the number of mites that could fit on one leaf, students noticed that leaf size would change since the mites feed on leaves. Accordingly, the students decided to instead use the number of mites that could be fed by one leaf to estimate the maximum number of each plant specialist mite that could survive based on the number of leaves supplied (Figure 5b). In the second class (SBT), the second approach was chosen by the students from the start. After estimating this parameter, students estimated the number of each mite variety after 45 days by applying a strategy similar to the one used in session 2 (again applying KC2, KC5 and KC4) while also accounting for resource availability (KC6, Table 1). When asked to explain why the least frequent variety of mites become the most frequent after this estimation (Figure 3c; KC9, Table 1), students introduced and discussed the concepts of differential survival (KC7, Table 1) and differential reproduction (KC8, Table 1) as well as their impacts on the number of offspring remaining over generations (KC3, Table 1). During this discussion, students in both classes orally described the process of natural selection applied to this biological scenario. These results suggest that during the three sessions, the students explored all of the key concepts that were planned (Table 1).

FIGURE 5

Evaluation of the impacts of the activity in students' LUENS

The impact of our proposed activity was examined in the two target groups that we applied the activity with (SAT and SBT). Significant differences in LUENS (Z_{SAT}= -2.961, p = 0.003 and Z_{SBT}= -2.591, p = 0.010) were recorded between the pre- and post-tests in the two target classes, with post-tests revealing a better understanding of evolution (Figure 6).

The percentage of students' answers falling under the category of each rubric item is presented in Table 5. Differences between pre- and post-tests were observed in *i*) the type of prediction made by the students and *ii*) the justification of this prediction. A significant increase in *fittest* predictions and a significant decrease in *fixist* predictions were observed between pre- and post-tests in both target classes ($p<0.05$) (Figure 7).

TABLE 5

At the pre-test, more than half of the students in the SAT class provided *fixist* predictions (see Table 5), with *fittest* predictions being the second most frequent. However, in the SBT target class, students mostly provided *fittest* predictions, with *fixist* predictions being the second most frequent. *Equilibrium* predictions were the least frequent in all classes. In post-tests, the *fittest* predictions increased in both classes and become the most frequent in both classes. Notably, many of the *fixist* predictions were justified with a mathematical model for population growth that only accounts for the number of offspring an individual can have (see, for example, Figure 2a). The changes observed to the *fittest* predictions involved students introducing additional biological parameters to this model, namely resource availability and the consequent differential survival and reproduction of individuals in the diverse population. In agreement with this, when comparing post-tests across classes, there was a stronger increase in the frequency of students justifying their predictions with *resource availability*, *differential survival* or *differential reproduction* of the phenotypes in target than in control classes (see Figure 2 and Table 4 for examples, Table 5 for frequencies and Figure 7 for a graphical representation). However, these differences were only significant in the target SA school, and only for the items *resource availability* and *differential survival* ($p = 0.002$ and $p = 0.031$ respectively; see Table 5).

FIGURE 6

FIGURE 7

Although most predictions were related to the proboscides size in butterfly populations, in post-tests, one student predicted the evolution of calyx tube size in the plant population since the plants were pollinated by the butterflies (see Student C, SBT post-test in Table 4). *Teleological* and *developmental* justifications were rare in most classes for both pre- and post-tests, and no significant differences between pre- and post-tests were observed for these two types of justifications in any of the classes (see frequencies in Table 5 and examples in Table 4).

Discussion:

The results of the present study indicate that our approach allowed elementary school students to explore and link all of the historically important key concepts. Notably, this approach was able to promote elementary school students' understanding of evolution by natural selection.

During session 2, students applied three of the eight historically important key concepts. In session 3, all eight of these concepts were applied to solve and discuss the results of the proposed problem. Moreover, using this approach led to a high and significant increase in students' LUENS (average increase of LUENS of 1.51 on a scale from 0 to 5), which was mostly due to: *i*) the significant increase of *fittest* predictions and the significant decrease of *fixist* predictions; *ii*) the strong (and statistically significant, in the case of the SAT class) increase in justifications mentioning the *resource availability*, *differential survival* and *differential reproduction* of the phenotypes (see Figure 7).

Although additional studies are required, our results support the hypothesis that PBL activities designed to explore concepts and conceptual fields that were important during the historical process of scientific discoveries may foster science understanding in students. The history of science has been widely used to design activities that allow students to learn about the nature of science and develop important scientific and critical thinking skills (Clough, 2010; Gooday et al., 2008; Mavrikaki & Kapsala, 2014). Regarding evolution, many textbooks mention the important contribution of Malthus' principle for developing the concept of natural selection (see, for example, Silva et al., 2004; Mader, 2009; Levine & Miller, 1994). However, to the best of our knowledge, no educational activities have been designed for students to link the concepts underlying this principle with those of intraspecific variability and habitat diversity through active learning. Our results are promising and highlight the potential of applying educational activities designed to promote historically important conceptual fields about evolution.

It is interesting to note that in pre-tests, many students that provided fixist explanations based these predictions on simple mathematical models that only consider a few parameters (namely the initial proportions of the varieties (KC2), trait heritability (KC4) and, in some cases, the potential reproductive output of the species (KC5); see Table 4 and Figure 2 for examples). In fact, the observed improvement in LUENS was achieved because students accounted for other biologically meaningful parameters (and evolution key concepts) in their answers, especially the selective pressure imposed by the available resources (KC6, Table 1) and the resulting differential survival (KC7, Table 1) and reproduction of the distinct populations (KC8; Table 1), which allowed them to predict the frequency change (KC9; Table 1). During the activity, these concepts were linked through increasingly complex mathematical models that incorporated several meaningful biological parameters and were collaboratively built by the students to solve the real-life problem posed to them. This further supports the potential of educational transdisciplinary activities that use mathematical modelling to promote and support science learning (see review in NRC, 2007).

Other features of our activity also likely contribute to its success, namely: *i)* the engagement of students with real organisms that they have observed and measured (Broder et al., 2018); *ii)* the context of the activity was a real-life problem (i.e., the need to grow mites in order to repeat the activity in other schools); *iii)* the cooperative PBL approach followed, with repeated cycles of learning and knowledge application; *iv*) the short life cycle of the mites, which would allow evolution to be observable in a very short period of time. We acknowledge that the model organism we used and the contact with the research team may not be easy to replicate in some schools. This could be a limitation for teachers that wish to apply this activity in their schools. However, this limitation might be easily overcome by using other organisms that have already been explored in schools. For instance, despite its longer life cycle (one year), the silk moth (*Bombyx mori*) has great reproductive potential and is heavily dependent on a specific type of food, which rapidly becomes a limiting resource. In this scenario, students can be asked what would happen if one individual is born with a heritable difference in its ability to eat other types of food.

Although other activities have been reported to explore natural selection with elementary school students (see, among others, Kelemen et al., 2014; Shtulman et al., 2016; Berti et al., 2017; Sá-Pinto et al., 2017a; Frejd et al., 2020), to the best of our knowledge, no other activity has engaged students in mathematical modelling to achieve this type of goal. However, mathematical thinking and the ability to develop and use models have been recognised as important scientific practices that students should learn since their initial years of schooling (NRC, 2012). When considering evolution, the ability to think mathematically while using and extending Malthus' mathematical model on population growth by including other biological parameters was fundamental for Darwin and Wallace to reason about natural selection and, according to our results, may also influence students' learning about this evolutionary process.

Among all of the important evolution-related KCs required to understand the principles of selection (Tibell & Harms, 2017; see Table 1), *differential reproduction* (and consequently *fitness* from the principle variation) was least commonly applied by students to justify their predictions in both tests. Moreover, although there was an increase in the frequency of its use from pre- to post-test, this difference was not statistically significant. Our results are in line with those obtained by Brown et al. (2020), who reported that 32% of students used this key concept after a storytelling intervention. These results suggest that additional effort should be made to increase students' understanding of and ability to apply this key concept. To achieve this goal, we propose extending session 3 by asking students to estimate (and graphically represent) the number of viable offspring per individual that were able to survive and reproduce for each generation. Additionally, an activity that explicitly asks students to link the different key concepts (e.g., a conceptual map) could contribute to scaffolding their conceptual field of evolution by natural selection. This exercise is expected to improve students' perceptions of these two key concepts. Additional possibilities that allow students to explore the importance of *differential reproduction* to drive frequency change involve the use of activities that directly explore sexual selection as the process driving reproductive success (see Sá-Pinto et al., 2017b for a review on the importance of sexual selection for evolution and evolution understanding as well as activities that aim to explore this process).

An interesting result from the present work is the low level of teleological justifications identified (< 2% of the total number of answers). These results strongly contrast with those of previous studies with older students, which suggests that teleological thinking is one of the main difficulties precluding evolution understanding (see review in Galli & Meinardi, 2011). Many studies report a high frequency of misconceptions related to teleological thinking in older students, which are persistent and difficult to change—even through educational programmes specifically designed to address them (Bishop & Anderson, 1986; Nehm & Reilly, 2007). Younger students were also shown to provide teleological explanations for biological scenarios involving natural selection before instruction (Brown et al., 2020).

Some researchers have proposed that teleological thinking is innate or developed in early childhood (reviewed in Kelemen, 1999a). While young children apply teleological thinking promiscuously to both biological and non-biological agents, adults seem to selectively apply it to human-made artefacts and biological agents (Kelemen, 1999b). The study by Kelemen (1999b) also revealed that a very large proportion of adults still provide teleological explanations to biological parts of living organisms and even to biological entities (e.g., babies or plants). Therefore, the low frequency of teleological answers reported here and in other studies that analysed elementary students' understanding of evolution by natural selection using this and similar frameworks (Sá-Pinto et al., 2017a) is surprising. This contrasts with recently published results showing that at pre-tests, 28% of elementary school students provided teleological explanations for biological scenarios involving natural selection (Brown et al., 2020).

Some work suggests that adults provide more teleological explanations when their causal knowledge is eroded or in high processing demands situations (see Kelemen & Rosset, 2009 and references therein). This suggests that both factual knowledge and time to process information are essential for people to inhibit teleological explanations. According to this view, the low frequency of teleological predictions reported here could be partially explained by the features of the evaluation framework used in this study. While people are usually asked to explain an extant scenario or entity (e.g., Kelemen, 1999a, 199b), we asked students to predict the evolutionary outcomes of a given biological scenario and told them that the target trait was heritable. This provided the students with factual information about the departure scenario (namely the existence of intraspecific variability and trait heritability), which may facilitate the process of suppressing teleological explanations. However, factual information on the initial population composition was also provided by Brown et al. (2020), suggesting that this factor may not be the only explanation for the reduced teleological explanations reported here. It is possible that asking students to predict and explain these predictions result in ways of thinking that differ from those required to explain observations. To further test this hypothesis, studies comparing elementary students' performance with distinct evaluation frameworks would be necessary.

Another possible explanation for the low level of teleological explanations found in this and previous studies on evolutionary thinking (Sá-Pinto et al., 2017a; Emmons et al., 2017), when compared to those found in adults and older students (Bishop & Anderson, 1986; Rutledge & Warden, 2000; Miller et al., 2006; Nehm & Reilly, 2007; Prinou et al., 2011; Spiegel et al., 2012), could be the reinforcement of this misconception during people's lives. Several studies have suggested that teleological thinking in evolution can be reinforced by teachers, books, the media and even by the way evolutionary biologists speak about evolution (Nehm et al., 2010, Prinou et al., 2011). This would support the importance of an early introduction of students to evolutionary processes, which has been advocated by several authors (e.g., Nadelson, 2009; Wagler, 2010, 2012; Campos & Sá-Pinto, 2013; Kelemen et al., 2014; Berti et al., 2017; Pires et al., 2016; Emmons et al., 2017; Sá-Pinto et al., 2017a, 2017b; Brown et al., 2020; Frejd et al., 2020). As suggested by Emmons et al. (2017), early instruction on evolution may preclude the development and strengthening of misconceptions on the topic, thereby providing children with scientifically accurate explanations to compete with inaccurate ideas in multiple learning and reasoning contexts. To further support this idea, the work of Brown et al. (2020) suggested that teleological reasoning in elementary school students may be easy to overcome with instruction, a pattern that contrasts with what has been reported for older learners and adults (Bishop & Anderson, 1986; Nehm & Reilly, 2007).

In Portuguese official curricula, evolution by natural selection is not present as a learning goal until the 11th grade. Therefore, it is highly improbable that the students which engaged in our activity had previously explored this process in school. Both the present work and the work previously published on these grades (Kelemen et al., 2014; Shtulman et al., 2016; Berti et al., 2017; Emmons et al., 2017; Sá-Pinto et al., 2017a; Brown et al., 2020; Frejd et al., 2020) only evaluated the impact of students' engagement in one activity exploring natural selection. However, as suggested by both Nehm (2018) and Vergnaud (2009), a clear understanding of natural selection and its key concepts (or 'invariables'; Vergnaud, 2009) can only be achieved through the exploration of this process in distinct situations. Therefore, future studies should attempt to understand how addressing natural selection under distinct situations contributes to elementary school students' understanding of evolution by natural selection.

Conclusions

In the present work, we present an innovative and effective approach to explore natural selection and promote evolution understanding in elementary school students. To foster learning about evolutionary processes, we designed a transdisciplinary activity that uses real-world problems to engage students in mathematical modelling that links concepts that were historically important to Darwin discovering the process of natural selection. Our activity allowed students to put in action all the historically important key concepts and resulted in a significant increase in their understanding of evolution by natural selection. Despite this, the activity did not significantly increase students' ability to use the key concept of differential reproduction, which suggests that this is a proximal development zone that additional activities could improve. The in-depth study of the activity implementation revealed that some fine-tuning of the activity may further enhance learning about this key concept. In contrast to what has been reported for older students and adults, we observed an unexpectedly low level of teleological answers from elementary school students. Together, these results contradict the general assumption that young children are unable to learn evolution by natural selection and mostly apply teleological thinking to biological processes. This result highlights the importance of early learning about evolution and raises new research questions related to the development and use of teleological explanations during a person's life.

References:

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.

Athanasiou, K. & Mavrikaki, E. (2013). Conceptual Inventory of Natural Selection as a tool for measuring Greek university students' evolution knowledge: Differences between novice and advanced students. *International Journal of Science Education,* 36:8, 1262-1285https://doi.org/10.1080/09500693.2013.856529

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences*, *13*(1), 1–14.

Barlow, N. ed. (1958). The autobiography of Charles Darwin 1809–1882. With the original omissions restored. Collins.

Berti, A. E., Barbetta, V., & Toneatti, L. (2017). Third-graders' conceptions about the origin of species before and after instruction: an exploratory study. *International Journal of Science and Mathematics Education*, *15*, 215–232.

Bishop, B. A., & Anderson, C. W. (1986). Student conceptions of natural selection and its roles in evolution. Research Series N 165. East Lansing, MI: The Institute for Research on Teaching.

Bransford, J., Brown, A. L., Cocking, R. R., & National Research Council (U.S.). (2000). How people learn: Brain, mind, experience, and school. National Academy Press.

Broder, E. D., Angeloni, L. M., Simmons, S., Warren, S., Kaitlin D., Knudson, K.D., & Ghalambor, C. K. (2018). Authentic science with live organisms can improve evolution education. *The American Biology Teacher*, *80*(2), 116–123. https://doi.org/10.1525/abt.2018.80.2.116

Brown, S. A., Ronfard, S., & Kelemen, D. (2020). Teaching natural selection in early elementary classrooms: Can a storybook intervention reduce teleological misunderstandings? *Evolution: Education and Outreach*, *13*, 12. https://doi.org/10.1186/s12052-020-00127-7

Campos, R., & Sá-Pinto, X. (2013). Early evolution of evolutionary thinking: Teaching biological evolution in elementary schools. *Evolution: Education and Outreach*, *6*(25), 1–13 https://doi.org/10.1186/1936-6434-6-25

Carroll, S. P., Jørgensen, P. S., Kinnison, M. T., Bergstrom, C. T., Denison, R. F., Gluckman, P., Smith, T. B., Strauss, S. Y., & Tabashnik, B. E. (2014). Applying evolutionary biology to address global challenges. *Science*, *346*(6207):,1245993.

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.

Cotner, S., Brooks, D. C., & Moore, R. (2010). Is the age of the earth one of our "sorest troubles"? Students' perceptions about deep time affect their acceptance of evolutionary theory. *Evolution, 64*(3), 858–864. https://dx.doi.org/10.1111/j.1558-5646.2009.00911.x

Clough, M. P. (2010). The story behind the science: bringing science and scientists to life in post-secondary science education. *Science and Education* 20, 701–717 (2011). https://doi.org/10.1007/s11191-010-9310-7

Darwin, C., & Wallace, A. R. (1858). On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Journal of the Proceedings of the Linnean Society, Zoology*, *20*, 45–62.

Darwin, F. ed. (1909). The foundations of the origin of species, a sketch written in 1842 by Charles Darwin. Cambridge University Press.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher*, *35*, 125–129

Emmons, N., Lees, K., & Kelemen, D. (2017). Young children's near and far transfer of the basic theory of natural selection: An analogical storybook intervention. *Journal of Research in Science Teaching*, 55(3), 321-347. https://doi.org/10.1002/tea.21421

Frejd, J., Stolpe, K., Hultén, M. & Schönborn K.J. (2020): Making a fictitious animal: 6-7 year-old Swedish children's meaning making about evolution during a modelling task. *Journal of Biological Education* https://doi.org/10.1080/00219266.2020.1799843

Galli, L. M. G., & Meinardi, E. N. (2011). The role of teleological thinking in learning the Darwinian model of evolution. *Evolution Education and Outreach*, *4*, 145–152.

Gooday, G., Lynch, J. M., Wilson, K. G., & Barsky, C. K. (2008). Does science education need the history of science? *Isis*, *99*, 322–330.

Ha, M., & Nehm, R. H. (2014). Darwin's difficulties and student's struggles with trait loss: Cognitive-historical parallelisms in evolutionary explanation. *Science Education*, *23*, 1051–1074.

Helle, W., & Sabelis, M.W. (1985) Spider Mites: Their Biology, Natural Enemies and Control, Elsevier Science Publishing Company.

Hmelo-Silver, C.E. (2004). Problem-based learning: What and how do students learn. *Educational Psychology Review*, *16*(3), 253–266.

Hummel, E., & Randler, C. (2012). Living animals in the classroom: A meta-analysis on learning outcome and a treatment–control study focusing on knowledge and motivation. *Journal of Science Education and Technology*, *21*, 95–105. https://doi.org/10.1007/s10956-011-9285-4

Kapsala N., & Mavrikaki E. (2020). Storytelling as a pedagogical tool in nature of science instruction. In W. McComas (Ed.), *Nature of science in science instruction*. *Science: philosophy, history and education*. Springer. https://doi.org/10.1007/978-3-030-57239-6_27

Kelemen, D. (1999a). Function, goals and intention: Children's teleological reasoning about objects. *Trends in Cognitive Science*, *3*(12), 461–468.

Kelemen, D. (1999b). The scope of teleological thinking in preschool children. *Cognition*, *70*, 241–272.

Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition*, *111*(1), 138–143.

Kelemen, D., Emmons, N. A., Schillaci, R. S., & Ganea, P. A. (2014). Young children can be taught basic natural selection using a picture storybook intervention. *Psychological Science,* 25(4), 893-902. https://doi.org/10.1177/0956797613516009

Kelly, A. E. (2013). When is design research appropriate? In T. Plomp & N. Nieveen (Eds.), *Educational design research. Part-A: An introduction* (pp. 135–151). Netherlands Institute for Curriculum Development (SLO).

Lian, J., & He, F. (2013). Improved performance of students instructed in a hybrid PBL format. *Biochemistry and Molecular Biology Education*, *41*(1), 5–10.

Malthus, T. (1798). An essay on the principle of population, as it affects the future improvement of society with remarks on the speculations of Mr. Godwin, M. Condorcet and other writers. Electronic Scholarly Publishing Project.

Mavrikaki, E., & Kapsala, N. (2014). Teaching biology by storytelling. In G. Katsiampoura (Ed.), *Scientific cosmopolitanism and local cultures: Religions, ideologies, societies* (pp. 612–617), Proceedings of the 5th International Conference of the European Society for the History of Science – 5ESHS. ISBN 978-960-98199-3-0

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

Migeon, A., Nouguier, E., & Dorkeld, F. (2011). Spider Mites Web: A comprehensive database for the Tetranychidae. In: Sabelis M., Bruin J. (eds) *Trends in Acarology*. Springer, Dordrecht. https://doi.org/10.1007/978-90-481-9837-5_96

Miller, J. D., Scott, E. C., & Okamoto, S. (2006). Public acceptance of evolution. *Science*, *313*, 765–766.

Nadelson, L., Culp, R., Bunn, S., Burkhart, R., Shetlar, R., Nixon, K., & Waldron, J. (2009). Teaching evolution concepts to early elementary school students. *Evolution: Education and Outreach*, *4*, 267–274.

National Research Council (2007). Taking science to school: Learning and teaching science in grades K-8. The National Academies Press.

National Research Council (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. The National Academies Press.

Nehm, R. (2018) Evolution. In K. Kampourakis & M.J. Reiss (Eds.). *Teaching biology in schools: Global research, issues, and trends.* Routledge Taylor and Francis Group.

Nehm, R. H., & Reilly, L. (2007). Biology major's knowledge and misconceptions of natural selection. *BioScience*, *57*, 263–272.

Nehm, R. H., Megan, A. R., & Ha, M. (2010). “Force-talk” in evolutionary explanation: Metaphors and misconceptions. *Evolution Education and Outreach*, *3*, 605–613.

Nehm, R. H., & Ridgway, J. (2011). What do experts and novices “see” in evolutionary problems. *Evolution Education and Outreach*, *4*, 666–679.

Orr, H. A. (2009). Fitness and its role in evolutionary genetics. *Nature Reviews Genetics*; 10(8), 531–539. doi:10.1038/nrg2603.

Pires, Y., Sá-Pinto, X., Martins, A. P., Pinto, A. (2016) Bacalhau, sardinhas, cotas de pesca e evolução. Sensos-e 3 (2) http://sensos-e.ese.ipp.pt/?p=12615

Plomp, T. (2013). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *Educational design research. Part-A: An introduction* (pp. 10–51). Netherlands Institute for Curriculum Development (SLO).

Portuguese Government / Ministry of Education. Essential learning goals. 1st grade: Basic Education. 1st Cycle. Study of the Environment. 2018a https://www.dge.mec.pt/sites/default/files/Curriculo/Aprendizagens_Essenciais/1_ciclo/1_estudo_do_meio.pdf

Portuguese Government / Ministry of Education. Essential learning goals. Basic Education. 2nd grade:1st Cycle. Study of the Environment. 2018b https://www.dge.mec.pt/sites/default/files/Curriculo/Aprendizagens_Essenciais/1_ciclo/2_estudo_do_meio.pdf

Portuguese Government / Ministry of Education. Essential learning goals. 3rd grade: Basic Education. 1st Cycle. Study of the Environment.2018c https://www.dge.mec.pt/sites/default/files/Curriculo/Aprendizagens_Essenciais/1_ciclo/3_estudo_do_meio.pdf

Portuguese Government / Ministry of Education. Essential learning goals. 4th grade: Basic Education. 1st Cycle. Study of the Environment. 2018d https://www.dge.mec.pt/sites/default/files/Curriculo/Aprendizagens_Essenciais/1_ciclo/4_estudo_do_meio.pdf

Portuguese Government / Ministry of Education. Essential learning goals. 5th grade: Basic Education. 2nd Cycle. Natural Sciences. 2018e https://www.dge.mec.pt/sites/default/files/Curriculo/Aprendizagens_Essenciais/2_ciclo/5_ciencias_naturais.pdf

Portuguese Government / Ministry of Education. Essential learning goals. 6th grade: Basic Education. 2nd Cycle. Natural Sciences. 2018f

Prinou, L., Halkia, L., & Skordoulis, C. (2011). The inability of primary school to introduce children to the theory of biological evolution. *Evolution: Education and Outreach*, *4*(2), 275–285.

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

Sá-Pinto, X., Pinto, A., Cardia, P., Fonseca, M.J, & Lopes, J.B. (2017a) Proposal for a framework to evaluate elementary school students' understanding of natural selection. Enseñanza de las Ciencias. extra number: 1083-1088.

Sá-Pinto, X., Cardia, P., Campos, R. (2017) Sexual selection: A short review on its causes and outcomes, and activities to teach evolution and the nature of science. *The American Biology Teacher*, *79*(2), 135–143. http://dx.doi.org/10.1525/abt.2017.79.2.135

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences,* *23*(1), 18–36 https://doi.org/10.1080/10508406.2013.778204.

Spiegel, A. N., Evans, E. M., Frazier, B., Hazel, A., Tare, M., Gram, W., & Diamond, J. (2012). Changing museums visitor's conceptions of evolution. *Evolution: Education and Outreach*, 5, 43–61.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation, 9*(4). https://doi.org/10.7275/96jp-xz07.

Shtulman, A., Neal, C., & Lindquist, G. (2016). Children's ability to learn evolutionary explanations for biological adaptation. *Early Education and Development*, *27*(8), 1222–1236.

Tibell, L. A. E., & Harms, U. (2017). Biological principles and threshold concepts for understanding natural selection: Implications for developing visualizations as a pedagogic tool. *Science and Education*, *26*, 953–973.

To, C., Tenebaum, H.R., & Hogh, H. (2017). Secondary school students' reasoning about evolution. *Journal of Research in Science Teaching*, *54*(2), 247–273.

van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (2006). Introducing educational research. In J. van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 3–7). Routledge.

Wagler, R. (2010). A missing link: K–4 biological evolution content standards. *Evolution: Education and Outreach*, *3*, 443–450.

Wagler, R. (2012). Assessing "the framework" for kindergarten through fifth-grade biological evolution. *Evolution: Education and Outreach*, *5*, 274–278.

Wrensch, D. L., & Young, S. S. Y. (1975). Effects of quality of resource and fertilization status on some fitness traits in the two-spotted spider mite*, Tetranychus urticae* Koch. *Oecologia*, *18*, 259–267.

Yasri, P., & Mancy, R. (2016). Student positions on the relationship between evolution and creation: What kinds of changes occur and for what reasons? *Journal of Research in Science Teaching*, *53*(3), 384–399.

TABLE 1 List of the key concepts of evolution by natural selection (from Tibell & Harms, 2017), evidence for these in Darwin's initial publication on the process of natural selection (Darwin and Wallace, 1858) as well as his diary (Barlow 1958), and how have these been addressed in the activities and biological scenario presented to students in the evaluation framework

Principles	Key concepts (KCs)	Evidence for this KC and how was it addressed by Darwin	How this KC is addressed in the activity	How this KC is addressed in the evaluation framework
Variation	**KC1:** Origin of variation (genetic changes)	*"In nature, we have some slight variation occasionally in all parts; and I think it can be shown that* ***changed conditions of existence is the main cause of the child not exactly resembling its parents****"* (Darwin and Wallace, 1858) *"(...)****during millions of generations individuals of a species will be occasionally born with some slight variation, profitable to some part of their economy"*** (Darwin and Wallace, 1858)	Although we present a species with variable traits, the genetic basis of these is not discussed further than the traits being heritable	Although we present a species with variable traits, the genetic basis of these is not discussed further than the traits being heritable
	KC2: Individual (phenotypic) variation	*"(...)* ***those individuals with the lightest forms, longest limbs, and best eyesight, let the difference be ever so small, would be slightly favoured****, and would tend to live longer, and to survive during that time of the year* ***when food was scarcest****; they would also rear more young, which would tend to inherit these slight peculiarities"* (Darwin and Wallace, 1858)	Sessions 2 and 3: Spider mites populations differ in their ability to feed on distinct food sources	Butterflies differ in their ability to feed on distinct food sources
	KC3: Differential	*"(...) those individuals with the lightest forms, longest limbs, and best eyesight, let the difference be*	Session 3: Individuals of the two populations of	Individuals of the two varieties of butterflies

	fitness (likelihood to survive and reproduce)	*ever so small, would be slightly favoured, and* ***would tend to live longer, and to survive during that time of the year when food was scarcest; they would also rear more young, which would tend to inherit these slight peculiarities"*** (Darwin and Wallace, 1858)	spider mites differ in their probability of surviving and reproducing in the described environment	differ in their probability of surviving and reproducing in the described environment
Reproduction	**KC4:** Heritable traits	*"(…) those individuals with the lightest forms, longest limbs, and best eyesight, let the difference be ever so small, (…) they would also rear more young,* ***which would tend to inherit these slight peculiarities"*** (Darwin and Wallace, 1858)	Sessions 2 and 3: The ability of mites to feed from distinct food sources is a variable trait that passes from parents to offspring	The ability of butterflies to feed from distinct food sources is a variable trait that passes from parents to offspring
	KC5: Reproduction	*"Suppose in a certain spot there are eight pairs of birds, and that only four pairs of them annually (including double hatches) rear only four young, and that these go on rearing their young at the same rate, then at the end of seven years (a short life, excluding violent deaths, for any bird) there will be 2048 birds, instead of the original 16"* (Darwin and Wallace, 1858)	Session 2: Each adult female lays 100 eggs and dies soon thereafter. From these eggs, 100 individuals are born (half males, half females)	Each adult butterfly lays four eggs and dies soon thereafter. From these eggs, four individuals are born
Selection	**KC6:** Selection pressure	*"But for animals without artificial means, the amount of food for each species must, on an average, be constant, whereas the increase of all organisms tends to be geometrical, and in a vast majority of cases at an enormous ratio"* (Darwin and Wallace, 1858)	Session 3: Resource availability imposes a selective pressure on the mite population, thereby limiting population growth. This selective pressure was distinct for the two distinct mite populations	Resource availability imposes a selective pressure on the butterfly population, thereby limiting population growth. This selective pressure was distinct for the two distinct butterfly varieties

	KC7: Differential survival	*"(…)during millions of generations individuals of a species will be occasionally born with some slight variation, profitable to some part of their economy.* ***Such individuals will have a better chance of surviving****, and of propagating their new and slightly different structure (…)"* (Darwin and Wallace, 1858)	Session 3: In the described environment the mites that can feed from lemon tree leaves had increased probability of survive, when compared to those that feed on bean leaves.	In the described environment, butterflies that can feed from flowers with a long calyx had an increased probability of survival when compared to those that feed from flowers with a short calyx
	KC8: Differential reproductio n	*"(…) those individuals with the lightest forms, longest limbs, and best eyesight, let the difference be ever so small, would be slightly favoured, and would tend to live longer, and to survive during that time of the year when food was scarcest;* ***they would also rear more young,*** *which would tend to inherit these slight peculiarities"* (Darwin and Wallace, 1858)	Session 3: In the described environment, the mites that can feed from lemon tree leaves had increased probability of reproduce, when compared to those that feed on bean leaves.	In the described environment, butterflies that can feed from flowers with a long calyx had an increased probability of reproducing when compared to those that feed from flowers with a short calyx
	KC9: Frequency change	*"(…) during millions of generations individuals of a species will be occasionally born with some slight variation, profitable to some part of their economy. Such individuals will have a better chance of surviving, and of propagating their new and slightly different structure, and* ***the modification may be slowly increased by the accumulative action of natural selection to any profitable extent. The variety thus formed will either coexist with or, more commonly, will exterminate its parent form"*** (Darwin and Wallace, 1858)	Session 3: In the context presented in session 3, the mites that can feed from lemon tree leaves survive more have a higher probability of survival and have more offspring than those that feed from bean leaves. Over generations, this results in a higher frequency of the lemon	In the environment presented in the figure, butterflies that can feed from flowers with a long calyx have a higher probability of survival and have more offspring than those that feed from flowers with a short calyx. Over generations, this results in a higher

			tree population	frequency of the variety with long proboscides
	KC10: Speciation	*"But at that time, I overlooked one problem of great importance (...). This problem is the tendency in organic beings descended from the same stock to diverge in character as they become modified. (...) The solution, I believe, is that the modified offspring of all dominant and increasing forms tend to become adapted to many and highly diversified places in the economy of nature"* (Barlow, 1958)	Not addressed	Not addressed

TABLE 2 Biology and mathematics learning goals for each of the three sessions

Session	Learning goals Biology	Mathematics
1	Scientific instrument manipulation skills Designing solutions for problems Exploring spatial scales	Length measurements Scales Mathematical problem-solving skills
2	Individual phenotypic variation (KC2) Heritable traits (KC4) Using mathematical thinking and modelling to estimate population growth due to reproduction (KC5)	Algebraic operations Geometric progressions Graphic representation of data Mathematical problem-solving skills
3	Resource availability (KC6) Differential survival (KC7) and reproduction (KC8) Frequency changes due to evolution by natural selection (KC9) Using mathematical thinking and modelling to estimate population growth considering individuals' fitness (KC5 and KC3)	Algebraic operations Geometric progressions Graphic representation of data Mathematical problem-solving skills

Abbreviation: KC, key concepts.

TABLE 3 Spearman's correlation coefficient and the statistical significance obtained between distinct types of predictions and rubric items related to evolution in students' justifications

Prediction type	Resource availability	Differential survival	Differential reproduction
Fixist	-0.672**	-0.398**	-0.329**
Equilibrium	0.010	0.082	-0.046
Fittest	0.873**	0.515**	0.465**

Abbreviation: **, statistically significant at p<0.01.

TABLE 4 Definition of each rubric item, its score contributing to the level of evolution understanding and examples based on students' answers

	Criteria	Definition	Score	Examples
Predictions*	*Fittest*	Student writes and/or draws that the fittest phenotype will become the most frequent	1	Student P, SA target, post-test (LUENS=3): ***I expect to find more butterflies with big noses*** *since there are more flowers with long calyxes and the others will die* Student A, SA control, pre-test (LUENS=2): *In 100 years,* ***I expect to find more butterflies with long proboscides*** *since there are more butterflies with short proboscides now, which will use more nectar from the flowers with short calyxes, leaving the butterflies with short proboscides without food, unlike those with long proboscides that will continue to feed themselves as before* Student F, SA target, post-test (LUENS=5): ***I think I will find more butterflies with long proboscides*** *since this butterfly has a lot of food and, for this reason, will have offspring faster and these offspring will still have the initial food. And one year after that, more flowers will grow and these butterflies can get food. The other butterflies will be left with no food and only some of them will be able to survive* Student C, SB target post-test (LUENS=4): *In 100 years, I expect to find butterflies with short proboscides since* ***there will be more flowers with short calyxes*** *and therefore these butterflies have food.* ***Since there are more butterflies with short proboscides, they help flowers with short calyxes to reproduce more***

	Equilibrium	Student writes and/or draws that the two phenotypes will become equally frequent	0	Student B, SA control, post-test (LUENS=0): ***I expect to find the same quantity of butterflies with long and short proboscides*** *since each butterfly can lay four eggs*
	Fixist	Student writes and/or draws that the initially most frequent haplotype will continue to be the most frequent one	0	Student L, SA control, post-test (LUENS=0)*: I expect to* ***find 16 butterflies with short proboscides and 4 with long proboscides since each butterfly lays 4 eggs*** Student X, SB target, pre-test (LUENS=0)*: In one hundred years, there will be butterflies since butterflies live one year and then die, and in one hundred years* ***there will be the same number as today but four times more. I will find 16 with short beaks and 4 with long***
Justifications	*Developmental*	Student states that the size of the proboscides depends on the individuals' developmental stage	0	Student K, SB control, post-test (LUENS=0)*: There will be more butterflies with short beaks.* ***Because these are younger.*** *Also,* ***a butterfly with a long beak is older***
	Teleological	Student justifies her/his prediction with a purpose, need or goal	0	Student Y, SB target, post-test (LUENS=0): *I expect to find butterflies since these are living beings and* ***they need to stay alive. And more with short proboscides*** Student M, SB target, post-test (LUENS=0): *I think I am going to find those with short and long proboscides* ***since, like this, there is lots of biodiversity***
	Resource availability	Prediction is justified by resource availability	1	Student P, SA target, post-test (LUENS=3): *I expect to find more butterflies with big noses* ***since there are more flowers with long calyxes*** *and the others will die*

				Student A, SA control, pre-test (LUENS=2): *In one hundred years, I expect to find more butterflies with long proboscides* ***since there are more butterflies with short proboscides now, which will use more nectar from the flowers with short calyxes, leaving the butterflies with short proboscides without food, unlike those with long proboscides that will continue to feed themselves as before*** Student F, SA target, post-test (LUENS=5): *I expect to find more butterflies with long proboscides since* ***this butterfly has a lot of food*** *and, for this reason, will have offspring faster, and these offspring will still have the initial food. And one year after that, more* ***flowers will grow and these butterflies can get food. The other butterflies will be left with no food*** *and only some of them will be able to survive* Student C, SB target post-test (LUENS=4): *In one hundred years, I expect to find butterflies with short proboscides since* ***there will be more flowers with short calyxes*** *and therefore these butterflies have food.* ***Since there are more butterflies with short proboscides, they help flowers with short calyxes to reproduce more***
	Differential survival	Student mentions that individuals with the fittest phenotype will survive more, or those with the least fit phenotype will die more	1	Student P, SA target, post-test (LUENS=3): *I expect to find more butterflies with big noses since there are more flowers with long calyxes* ***and the others will die*** Student F, SA target, post-test (LUENS=5): *I expect to find more butterflies with long proboscides since this butterfly has a lot of food and, for this reason, will have offspring faster and their offspring will still have the initial food. And one year after that, more flowers will grow and these butterflies*

				can get food. The other butterflies will be left with no food and ***only some of them will be able to survive***
	Differential reproduction	Student mentions that individuals with the fittest phenotype reproduce more or have more offspring than those with the least fit phenotype	2	Student D, SB target, post-test (LUENS=4): *I expect to find those with longer proboscides in one hundred years since* ***these can produce more eggs.*** *As such, there will be more butterflies on the island. This is going to happen because those with the longest proboscides are more capable, I mean, they have more food.* Student F, SA target, post-test (LUENS=5): *I expect to find more butterflies with long proboscides since this butterfly has a lot of food and, for this reason,* ***will have offspring faster*** *and their offspring will still have the initial food. And one year after that, more flowers will grow and these butterflies can get food. The other butterflies will be left with no food and only some of them will be able to survive* Student C, SB target post-test (LUENS=4): *In one hundred years, I expect to find butterflies with short proboscides since* ***there will be more flowers with short calyxes*** *and therefore these butterflies have food.* ***Since there are more butterflies with short proboscides, they help flowers with short calyxes to reproduce more***

Note: The fittest haplotype was considered to be the one with the longest proboscides, except when other phenotypes were considered by the students and correctly justified with differential survival and/or reproduction. Boldface font indicates the sections from the students' answers assigned to the rubric item.

Abbreviations: *, whenever drawn and written predictions differed, we considered the latter prediction to be the valid one; LUENS, level of evolution understanding; SA, School A; SB, School B; Target, classes subjected to the proposed educational activity to promote students' evolution understanding; Control, classes not subjected to the proposed educational activity to promote students' evolution understanding.

TABLE 5 Results obtained for each class in pre- and post-tests for each rubric item

Class/ test		**Fittest**	**Equilibrium**	**Fixist**	**Developmental**	**Teleological**	**Resource availability**	**Differential survival**	**Differential reproduction**
SAT pre-test	N	19	19	19	19	19	19	19	19
	%	15.8	5.3	63.4	0	0	15.8	10.5	0
	IR	1	1	1	1	1	1	0.95	0.95
SAT post-test	N	19	19	19	19	19	18	19	19
	%	68.4**	0	21.1*	0	0	72.2**	42.1*	26.3
	IR	1	0.95	1	1	1	1	0.89	1
SBT pre-test	N	24	20	23	25	21	25	25	24
	%	41.7	0	34.8	0	0	36.0	8.0	8.3
	IR	1	1	1	1	1	1	0.96	1
SBT post-	N	23	21	23	24	21	23	23	22
	%	69.6*	0	13.0*	4.2	4.8	60.9	30.4	27.3

test	IR	0.96	1	1	1	1	0.96	0.96	0.95

Abbreviations: N, number of answers that could be classified for the rubric item; %, percentage of answers assigned to the rubric item; IR, interrater reliability; *, value significantly different from pre-test result with McNemar test p-values lower than 0.05; **, value significantly different from pre-test result with McNemar test p-values lower than 0.01.

Figure Legends

FIGURE 1 Conjecture map of our design research, adapted from Sandoval (2014). Based on prior research, we suggest that elementary school students' understanding of natural selection could be fostered through a transdisciplinary problem-based activity that includes exploring Malthus' principle and intraspecific diversity in heritable characters (high-level conjecture). Therefore, we designed a task consisting of a collaborative inquiry where students would explore population growth with and without selection pressure using mites as the model organism (embodiments). When engaging in this task, students design, implement and evaluate different mathematical models of population growth while observing/analysing the effects of the different factors involved and linking the historical key concepts (mediating processes). This should help them to better understand natural selection and allow them to produce natural selection-based explanations. Also, they would improve their math and science skills (expected outcomes)

FIGURE 2 Examples of answers given by students. (a) example of an answer with a *fixist* prediction (Student L, SA control, post-test, text translation: *I expect to find 16 butterflies with short proboscides and 4 with long proboscides since each butterfly lays 4 eggs*; LUENS= 0); (b) example of an answer with an *equilibrium* prediction and a *teleological* justification (Student M, SB target, post-test, text translation: *I think I am going to find those with short and long proboscides because, like this, there is lots of biodiversity*; LUENS= 0); (c) example of a *fittest* prediction justified by *resource availability*, *differential survival* and *differential reproduction* (Student F, SA target, post-test, text translation: *I expect to find more butterflies with long proboscides since this butterfly has a lot of food and, for this reason, will have offspring faster and their progeny will still have the initial food. And one* *year after that, more flowers will grow and these butterflies can get food. The other butterflies will be left with no food, and only some of them will be able to survive.* LUENS= 5); (d) example of a *fittest* prediction justified by *resource availability* and *differential survival* (Student P, SA target, post-test, text translation: *I expect to find more butterflies with big noses since there are more flowers with a long calyx and the others will die*; LUENS= 3). SA, School A; SB, School B; Target, classes that were subjected to the educational activity developed to promote students' evolution understanding; Control, classes not subjected to the educational activity developed to promote students' evolution understanding

FIGURE 3 Evidence collected from class observations and students' produced materials in the first session. In students' materials, English translations of what was written by students are provided in the printwritting (a) example of a students' pre-concept of a mite; (b) mite observed under a tripod magnifier; (c) students observing mites under a microscope; (d) preparation of mites to be observed with a magnifying camera; (e) example of how a group of students proposed to estimate mites size

FIGURE 4 Evidence collected from class observations and students' produced materials in the second session. In students' materials, English translations of what was written by students are provided in the printwritting (a) a group of students explaining how they estimated the mites' population size using only mathematical language to their peers; (b) another group of students explain the strategy they used to estimate population size, using a combination of tables and mathematical language; (c) example of a students' group work depicting the estimation and graphic representation of both mites population sizes

FIGURE 5 Evidence collected from class observations and students' produced materials in the third session. In students' materials, English translations of what was written by students are provided in the printwritting (a) students' estimated the space occupied by lemon and beans' mites; (b) example of the work of one group of students, estimating the number of lemon and beans' mites that could be supported by the available resources; (c) example of a students' group work depicting the estimation of the population sizes of the two mites under resources limitation.

FIGURE 6 Average level of understanding of evolution by natural selection (LUENS; maximum level 5, based on the evolution understanding evaluation framework) revealed by students' answers in pre- and post-tests in target classes. White and grey bars indicate pre-tests and post-tests, respectively. * indicates a value significantly different from the one obtained by the students in pre-tests according to Wilcoxon test results ($p < 0.05$). Vertical lines represent standard errors of the mean. SAT– School A target class; SBT – School B target class

FIGURE 7 Frequencies of students' answers assigned to each coding rubric item in pre- and post-tests. White bars indicate the SAT class pre-test. Black bars indicate the SAT class post-test. White dotted bars indicate the SBT class pre-test. Black dotted bars indicate the SBT class post-test. Asterisks (*) denote significant differences between pre- and post-tests according to McNemar test results ($p < 0.05$). Vertical lines represent the standard errors of the difference between two proportions

Data Accessibility Statement: The database housing the results of the students' answers analysis is deposited in the Dryad repository: (to be included after acceptance).

Competing Interests Statement: The authors declare no competing interests.

Author Contributions section

The conceptualization of this research was done by XSP, JR, AP, IS and JBL. The educational activity was designed by JR with the support of XSP, AP, IS and LR and implemented by JR and XSP. The research methodology was designed by XSP, JR, LVB, EM, JBL and implemented by XSP, JR, PP. Data curation and analysis was performed by JR, XSP, PP, LR and EM. The writing of the paper was led by XSP with the contribution of all the authors. All the authors reviewed and approved the final version of the manuscript. Funding acquisition and project administration was led by XSP and EM.

Acknowledgements:

We would like to thank the students and teachers that engaged in our project and thus made it possible. We would also like to thank the MITE2 team for kindly providing mites and producing the movie used in this activity. Finally, we would like to thank Pedro Cardia for his help in image production. Xana Sá-Pinto is funded by Portuguese national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., within the scope of the framework contract foreseen in numbers 4, 5 and 6 of article 23 of the Decree-Law 57/2016 of August 29, changed by Law 57/2017 of July 19 and under the project UID/CED/00194/2019. Leonor Rodrigues has an ERC-funded postdoctoral grant (COMPCON, GA 725419). Lucía Vázquez-Ben is funded by the Ministerio de Ciência, Innovación y Universidades (Spain) (Project PGC2018-096581-B-C22). Patrícia Pessoa is supported by Portuguese national funds through the PhD grant I.P.2020.05634.BD funded by FCT – Fundação para a Ciência e a Tecnologia. This paper is the result of collaborative work by researchers participating in the EuroScitizen COST Action (CA17127; Working Group 2).