RESEARCH ARTICLE

Open Access

Effects of situated learning and clarification of misconceptions on contextual reasoning about natural selection

Helena Aptyka[1] , Daniela Fiedler[2] and Jörg Großschedl[1*]

Abstract

Background: Natural selection is a core principle of evolution. Understanding natural selection enables students to think about the evolution and the variability of life. Despite its great importance, understanding natural selection is challenging for students. This is evident in the phenomenon of contextual reasoning, showing that students can often explain natural selection in one context (e.g., trait gain) but not in another (e.g., trait loss). The study pursues the following aims: First, to examine the link between contextual reasoning and situated learning. Second, to explore whether different instructional strategies differ in their associated cognitive load. Third, to investigate whether clarifying common misconceptions about natural selection (no vs. yes) is an effective strategy to regular instructions when aiming to increase key concepts and reduce misconceptions. Fourth, to exploratively examine the effectiveness of different instructional strategies.

Method: In a 2 × 2 factorial intervention study with a total of $N = 373$ secondary school students, we varied the instructional material of a 90-min intervention in terms of the evolutionary context (trait gain vs. trait loss) and the availability of additional support in the form of a clarification of misconceptions (no vs. yes). We measured students' cognitive load immediately after instruction and assessed their ability to reason about natural selection (i.e., use of key concepts and misconceptions) later.

Results: We documented low knowledge about evolution in the pre-test and persisting misconceptions in the post-test. The results showed that the intervention context of trait loss elicited a higher intrinsic and extraneous cognitive load than trait gain. Moreover, when the clarification of misconceptions is analyzed in connection to the intervention context, it reveals a potential for reducing misconceptions in some contexts. Students who have learned in trait gain contexts with a clarification used significantly fewer misconceptions in later reasoning than students who learned in trait gain contexts without a clarification of misconceptions.

Conclusion: Our study creates new insights into learning about natural selection by outlining the complex interplay between situated learning, cognitive load, clarification of misconceptions, and contextual reasoning. Additionally, it advises researchers and educators on potential instructional strategies.

Keywords: Evolution education, Evolutionary knowledge, Concept, Secondary school students, Situated learning, Cognitive load, Contextual reasoning, Clarification of misconceptions

---

*Correspondence: j.grossschedl@uni-koeln.de
[1] Present Address: Institute for Biology Education, Faculty of Mathematics and Natural Sciences, University of Cologne, Herbert-Lewin-Straße 10, 50931 Cologne, Germany
Full list of author information is available at the end of the article

Introduction

Natural selection is a core principle of evolution and constitutes a pivotal role in evolution education (National Research Council [NRC] 2012; Secretariat of the

Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the federal republic of Germany [KMK] 2020). Understanding natural selection enables students to think and reason about the emergence and existence of biological variability and the evolution of species on earth. Concurrently, the principles of evolutionary biology can offer helpful approaches and strategies for facing global challenges (Carroll et al. 2014; Smith 2010). Despite its general importance, understanding and reasoning about evolutionary processes pose challenges that manifest themselves in unscientific or naïve ideas (in the following, referred to as misconceptions). These misconceptions can be traced back to early childhood and are also observed in secondary school students (e.g., Bishop and Anderson 1990; Opfer et al. 2012; Beggrow and Sbeglia 2019; Ha and Nehm 2014; Nehm and Ha 2011; Evans 2001; Kampourakis and Zogza 2009; Beniermann 2019; Kuschmierz et al. 2020a). Research with German secondary school students yielded that they have low (Beniermann 2019; Kuschmierz et al. 2020a, b) to moderate (Kuschmierz et al. 2020a; Fenner 2013; Lammert 2012) knowledge about evolution and often rely on teleological and 'Lamarckian' conceptions to explain natural selection (Beniermann 2019; Kuschmierz et al. 2020b; Fenner 2013; Lammert 2012). Additionally, other studies indicated that students use key concepts and misconceptions in a context-related manner when reasoning about natural selection (also termed contextual reasoning; Nehm and Ha 2011; Nehm et al. 2012). Context herein refers to the underlying content or topic. For example, when the context differs concerning the polarity of trait change, students use the key concept heritability of variation ('heritability') and the misconception use or disuse of particular body parts ('use/disuse') more often in trait loss than in trait gain scenarios (see also Opfer et al. 2012; Ha and Nehm 2014; Nehm et al. 2012). Although normative reasoning about natural selection would require using the same key concepts in all contexts, context-related reasoning occurs (Nehm and Ha 2011; Nehm et al. 2012; Federer et al. 2015). To promote normative reasoning in all contexts, there is an urgent need to ascertain factors linked to the phenomenon of contextual reasoning and identify effective instructional strategies that increase the use of key concepts and reduce prevailing misconceptions.

Current research suggests that the phenomenon of contextual reasoning could be related to situated learning (Opfer et al. 2012; Beggrow and Sbeglia 2019; Nehm and Ha 2011; Kirsh 2009). The situated learning approach (closely associated with situated cognition) posits that learning and knowledge are inextricably linked to a specific situation (Kirsh 2009; Brown et al. 1989; Reder et al. 1994; Sutton 2008). This approach can also be applied to

instructional situations. Again, situational entities determine learning and knowledge acquisition. For example, if students are learning with tasks on a certain context, this underlying context is decisive. As a result, acquired knowledge may only be accessible to a limited extent in new contexts. However, tasks in new contexts can be solved more effectively when students are familiar with similar ones (Kirsh 2009; Reder et al. 1994). If there is a lack of attention to situated learning, context-dependent lessons can occur, which means that natural selection may only be taught in one context but not in another. Context-dependent instructions do not appear to be uncommon, as Nehm and Ha (2011) ascertained, for example, that scenarios involving trait gain are covered more frequently in the curriculum than scenarios involving trait loss. Hence, students have more opportunities to develop key concepts of natural selection in trait gain contexts, while they lack practice in contexts of trait loss. Consequently, a mix of key concepts and misconceptions or relatively intuitive ideas (e.g., from childhood) may remain or are reinforced in trait loss contexts (Opfer et al. 2012; Evans 2001; Kirsh 2009; Reder et al. 1994). Notwithstanding, there is a lack of empirical evidence on whether contextual reasoning about natural selection is demonstrably related to situated learning.

Nehm and Ha (2011) extend the view on situated learning and contextual reasoning by pointing to a possible link between persistent misconceptions in less familiar contexts and a high load on working memory capacity. The emerging patterns in contextual reasoning may indicate that some contexts impose a greater cognitive load and are more difficult for students than others (Nehm and Ha 2011; Federer et al. 2015; Nehm 2018). For instance, it is more challenging for students to reason about trait loss or evolution in the plant kingdom than trait gain or evolution in the animal kingdom since individuals use fewer key concepts and more misconceptions (Nehm and Ha 2011; Federer et al. 2015; Nehm 2018; Ha et al. 2006; Großschedl et al. 2018). Similar patterns in concept use emerged throughout the study of the history of biology (Ha and Nehm 2014). Accordingly, difficulties in reasoning about trait loss scenarios are not necessarily due to familiarity. Some contexts may have always been more difficult than others (Ha and Nehm 2014). The link between different evolutionary contexts and their associated cognitive load has been a neglected aspect in previous studies, which makes this a worthwhile research direction of biology education (Nehm and Ha 2011; Klepsch and Seufert 2020).

Along with ascertaining factors associated with the phenomenon of contextual reasoning, it is important to consider instructional strategies to promote normative reasoning about natural selection. Andrews et al. (2011)

Aptyka *et al. Evolution: Education and Outreach*   (2022) 15:5

Page 3 of 21

highlighted that instructions could be more effective when key concepts of natural selection are taught and educators provide their students with an additional clarification of misconceptions. Clarifying misconceptions can enable students to recognize their misconceptions and compare them to scientific key concepts (Gartmeier et al. 2008; Tulis et al. 2016). Recognizing discrepancies when comparing concepts can play a vital role in restructuring inherent conceptualizations and developing scientific knowledge about natural selection (Kampourakis and Zogza 2009; Gartmeier et al. 2008; Tulis et al. 2016; Limón 2001; Oser et al. 2012; Rea-Ramírez and Clement 1998; Nelson 2008). Hence, it remains to be ascertained whether the clarification of misconceptions will lead to an increase in key concepts while reducing misconceptions.

In addition, a new perspective emerged from the literature suggesting that the aforementioned situated learning approach should be considered in combination with the instructional strategy of clarifying misconceptions. Studies on situated learning argue that concepts are generally developed in a context-related matter, including key concepts and misconceptions (Kirsh 2009; Goel et al. 2010; Barsalou 2016; Bechtel et al. 2009; Sadler 2009). Consequently, further developing situated concepts could be meaningful in contexts similar to those in which they originated (Barsalou 2005, 2016; Barsalou et al. 2009). Thus, it appears to be of importance to investigate situated learning in different contexts in combination with a clarification of misconceptions.

This study combines fundamental and application-orientated research to address the previously outlined research gaps. We investigate factors related to contextual reasoning and explore instructional strategies for strengthening key concepts and reducing misconceptions about natural selection. Our main aims are to clarify (a) whether situated learning is linked to differences in contextual reasoning and (b) whether instructions and underlying contexts differ in the associated cognitive load. Additionally, (c) it is essential to investigate if an explicit clarification of misconceptions increases the probability that students use fewer misconceptions and simultaneously more key concepts. Finally, (d) there is a need to explore whether considering learning as a holistic, situated process could shed light on seminal instructional strategies in terms of different contexts (trait gain vs. trait loss) and the clarification of misconceptions (no vs. yes). By engaging with these aims, our study considers the conglomerate of interdependencies. It can shed light on the hitherto unexplained manifestations of contextual reasoning and allows for a more profound exploration of effective instructional strategies.

Background

The classical situated learning approach emphasizes that learning and knowledge acquisition is tied to the situation in which it occurs. Thus knowledge acquisition relates uniquely to situation-specific entities such as the underlying activity, setting, and culture (Reder et al. 1994; Goel et al. 2010; Barsalou 2016). Moreover, situated learning determines how individuals integrate new into existing knowledge. In particular, the context is a central element for the instructional situations, as it is associated with different amounts or types of concepts. Since the use of concepts is context-related, the development or transformation of concepts should occur in a context similar to the one of interest (Reder et al. 1994; Goel et al. 2010; Barsalou 2016; Hendricks 2001; Schaffernicht 2006). For example, if the goal is to develop concepts in the context of trait gain, it may be helpful to tailor the instructional material to trait gain scenarios, especially for novices. The learning situation is also decisive for the extent to which students can retrieve knowledge in new situations (Goel et al. 2010; Barsalou 2016; Johnson-Laird 1980; Paas and Ayres 2014). Much of what is acquired is often not directly accessible and transferable from one situation to another (e.g., Brown et al. 1989; Lave and Wenger 1991). Anderson et al. (1996) propounded that students who have learned in a specific situation ('source') can recall the acquired knowledge more easily in similar target situations ('target'). For instructional situations, this means that the likelihood of transferability depends on how educational situations are framed (e.g., in a meaningful and authentic context) and how familiar the situations are. Therefore, the transfer is also more effortless for novices in similar contexts. The more experienced students are with knowledge transfer, the higher the probability that the transfer will be effective, even between somewhat dissimilar contexts (Reder et al. 1994; Anderson et al. 1996). An empirical study in educational sciences showed that situated learning is associated with an increase in the immediate learning effect on conceptual knowledge, for example, on causal reasoning (Hendricks 2001). Further multi-faceted empirical findings from cognitive sciences on situated learning and conceptualization are presented in Barsalou (2005). Empirical research on these theoretical approaches to situated learning in evolution education is scarce. Nevertheless, Nehm and colleagues alluded that the situatedness of prior learning can be related to different patterns in the reasoning about natural selection (e.g., Opfer et al. 2012; Nehm and Ha 2011). Thus, it is worth examining whether students who learn in one evolutionary context transfer their knowledge equally well to familiar and unfamiliar contexts.

In addition to the relationship between situated learning and contextual reasoning, previous studies revealed

that cognitive load is associated with instructions (e.g., Klepsch and Seufert 2020; Klepsch et al. 2017; Sweller and Chandler 1994; Klahr and Robinson 1981; Cooper 1998). Different instructional situations can bind different amounts of cognitive resources. Depending on the nature of the learning situation, the type and extent of the load can vary. To be more specific, the cognitive load is composed of three interrelated types: intrinsic, germane, and extraneous cognitive load (Chandler and Sweller 1991). Intrinsic load increases as tasks become more complex (e.g., large numbers of interacting elements; underlying context). Relating these findings to insights from evolution education, the difficulties associated with trait loss (compared to trait gain) may also translate into an increased intrinsic cognitive load (Nehm and Ha 2011; Federer et al. 2015). Germane load represents the capacity used for cognitive processes during the acquisition of knowledge or concept development. It is the only load that promotes learning. Germane load increases when students learn with effective instructions (e.g., Klepsch and Seufert 2020; Mayer 2002; Moreno and Park 2010; Sweller 2011). An example of an effective instructional strategy may be to add a clarification of misconceptions to regular teaching (Andrews et al. 2011; Choi and Hannafin 1995). Extraneous load depends on design variation and represents the artificially induced cognitive load. Taking the three loads together, the general rule for cognitive load during learning is that instructions should neither be over- nor under-challenging because this can lead to ineffective learning, erroneous reasoning, and misconceptions (Klepsch and Seufert 2020; Goel et al. 2010; Johnson-Laird 1980; Taber 2017; Schneeweiß and Gropengießer 2019). Overall, assessing cognitive load can improve our understanding of how individual learning outcomes emerge. It can aid in clarifying which modifications in instructional materials (e.g., context, instructional strategy) elicit which type of cognitive load. Furthermore, a cognitive load assessment can indicate which instructional materials must be modified to obviate detrimental cognitive load and prevent students from developing misconceptions. Modifying instructions could free up working memory capacity for students to acquire key concepts effectively (Goel et al. 2010; Barsalou 2016; Johnson-Laird 1980).

Although adjusting instructional materials in terms of cognitive load can facilitate learning about key concepts, acquiring key concepts is complex. It is not merely a matter of replacing one concept with another (Limón 2001; Sinatra et al. 2014; Posner et al. 1982). Students may hold key concepts and misconceptions in mixed models (e.g., Opfer et al. 2012; Evans 2001). As they are not mutually exclusive, misconceptions can persist if they are not recognized. Recognizing misconceptions is challenging for students because they are shaped by the individual conceptual ecology (KMK 2020; Nelson 2008; Sinatra et al. 2014). The conceptual ecology provides a frame with variables that affect individual thinking and reasoning about natural selection, such as knowledge about evolution, dualistic thinking, attitudes towards evolution, and personal religious faith (Park 2007; Deniz et al. 2008; Großschedl et al. 2014). For example, knowledge about evolution can have a facilitating or inhibiting effect on learning gains, depending on its nature. If a student's conceptual ecology exhibits a low level of knowledge about evolution and many misconceptions, these misconceptions can inhibit the ability to develop scientific knowledge about natural selection. As these misconceptions are deeply rooted, educators can incorporate instructional strategies such as an explicit clarification of misconceptions into common teaching. Clarifying misconceptions explicitly simplifies the recognition of misconceptions and requires less effort on the part of students (Gartmeier et al. 2008; Oser et al. 2012). Knowledge about misconceptions can cause dissatisfaction with one's concepts (Nelson 2008) and enables students to distinguish between key concepts and misconceptions. Furthermore, students most likely revise and reconstruct their misconceptions, translating into an increase in germane cognitive load (Klepsch and Seufert 2020; Mayer 2002). It is important to note that when students understand that certain concepts are misconceptions and they recognize them in a given context, this does not necessarily mean that students will draw the connection to similar contexts. However, if they recognize misconceptions in one context and notice similarities to another context, this can prevent them from repeatedly using misconceptions and making similar mistakes in other contexts (Oser et al. 2012; Otto and Mandorli 2018).

Empirical studies on the clarification of misconceptions found heterogeneous results by showing no (e.g., Aptyka and Großschedl 2019), positive (e.g., Kampourakis and Zogza 2009; Nehm and Reilly 2007; Colton et al. 2018), and mixed-effects (e.g., Heemsoth and Heinze 2016) on conceptual knowledge. After one semester of teaching about natural selection and actively addressing misconceptions, Nehm and Reilly (2007) found small effects as the diversity of key concepts significantly increased, and misconceptions decreased. Nevertheless, students continued to use context-related misconceptions, which accounted for a significant part of their conceptual knowledge. Similarly, Colton et al. (2018) reported small to large pre–post-test effects after one semester of instructions on natural selection and clarifications of misconceptions. The results showed that students used more key concepts and fewer misconceptions. The strength of the effects differed depending on the

Aptyka *et al. Evolution: Education and Outreach* (2022) 15:5

Page 5 of 21

underlying context of the pre- and post-test tasks. Heemsoth and Heinze (2016) obtained mixed results in their study on conceptual knowledge in mathematics. They identified prior knowledge as a determinant of learning outcomes. In their research, clarifying what is wrong promoted conceptual knowledge when students had high prior knowledge about the underlying topic. Conversely, students with low prior knowledge showed higher learning success after only focusing on key concepts (Heemsoth and Heinze 2016). These empirical results indicate that clarifying misconceptions can benefit students' conceptual knowledge. Moreover, the findings allude that students' prior knowledge and the underlying context could be decisive for students' learning outcomes. What remains empirically unexplained is whether the context is already important when clarifying misconceptions is used as an instructional strategy.

To summarise, analyzing learning from a diverse vantage point illuminates that different situated learning contexts involve different concepts and levels of cognitive load. A high cognitive load increases the difficulty for students to recognise misconceptions so that misconceptions may persist. When misconceptions are clarified (in targeted contexts), it can stimulate the germane cognitive load and accelerate the recognition and reconstruction of misconceptions. Eventually, this may lead to students having fewer misconceptions and developing normative reasoning.

Current study

Based on the theoretical background, we aim to fulfill four aims: First, to investigate the effect of situated learning in different evolutionary contexts on later contextual reasoning (cf. H1). Second, to examine whether instructional materials and underlying contexts vary significantly in associated cognitive load (cf. H2). Third, to explore the effectiveness of a clarification of common misconceptions for later reasoning about natural selection (cf. H3). Fourth, to analyze which instructional strategy effectively supports learning about natural selection and shows increased use of key concepts and fewer misconceptions in trait gain and trait loss scenarios in later reasoning (cf. exploratory analysis). Thus, the following hypotheses and exploratory analysis guided our research:

H1 Students learning in a specific evolutionary context use more key concepts and fewer misconceptions in later reasoning about the same context than students who learned in a different one before. Hence, we hypothesize that students learning in the instructional context of trait gain perform better in subsequent trait gain scenarios than in trait loss scenarios. Students learning in the instructional context of trait loss perform

better in subsequent trait loss scenarios than in trait gain scenarios.

H2 Given that trait gain contexts appear easier than trait loss contexts, we suppose students perceive lower intrinsic cognitive load when provided with the instructional context of trait gain than trait loss. Furthermore, supplying a clarification of misconceptions should ease conceptual development. Thus, students who receive no clarification of misconceptions should perceive a lower germane cognitive load than students who receive a clarification. Generally, we do not expect the extraneous load to differ between instructional materials. However, it should be noted that the context of instruction and the clarification of misconceptions might interact and affect differences in the three types of loads. Working memory is limited, and the three loads are mutually dependent.

H3 Since clarifying misconceptions should promote reconstructing conceptions, we expect students who received instructions without a clarification of misconceptions to use fewer key concepts and more misconceptions in later reasoning about natural selection than students who received a clarification.

Exploratory analysis We analyse which instructional strategy (differing in the context of instructions [trait gain vs. trait loss] and the clarification of misconceptions [no vs. yes]) effectively supports learning about natural selection and shows increased use of key concepts and fewer misconceptions (both in trait gain and trait loss scenarios in later reasoning).

An overview of the corresponding research design, including the variables of the individual hypotheses, can be found in Fig. 1.

Method
Sample
Overall, we recruited $N = 373$ upper secondary school students, of which nine students had to be excluded from the data analyses as preliminary tests showed strong outliers for these students. The students were $M = 16.7$ years old ($SD = 1.0$ years, 61% female), enrolled in 16 upper secondary schools in Germany (gymnasium, comprehensive school, and vocational training) and attended the grades 10–13 ($M = 10.9$, $SD = 0.7$).

Upper secondary education is equivalent to the third level of the International Standard Classification of Education (ISCED 3; Eurydice 2021). It covers natural selection to a comparable extent for the school types concerned (Eurydice 2021; Ministry for School and Further Education of the State of North Rhine-Westphalia [MSW

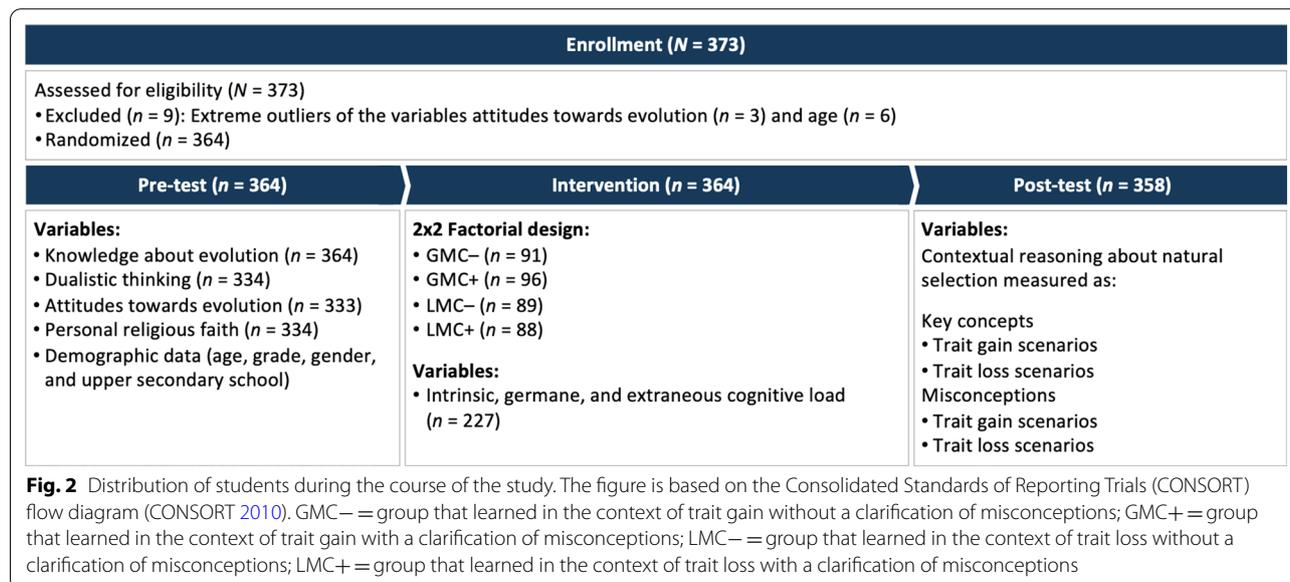**Fig. 1** Research design. *DV* dependent variable, *IV* independent variable
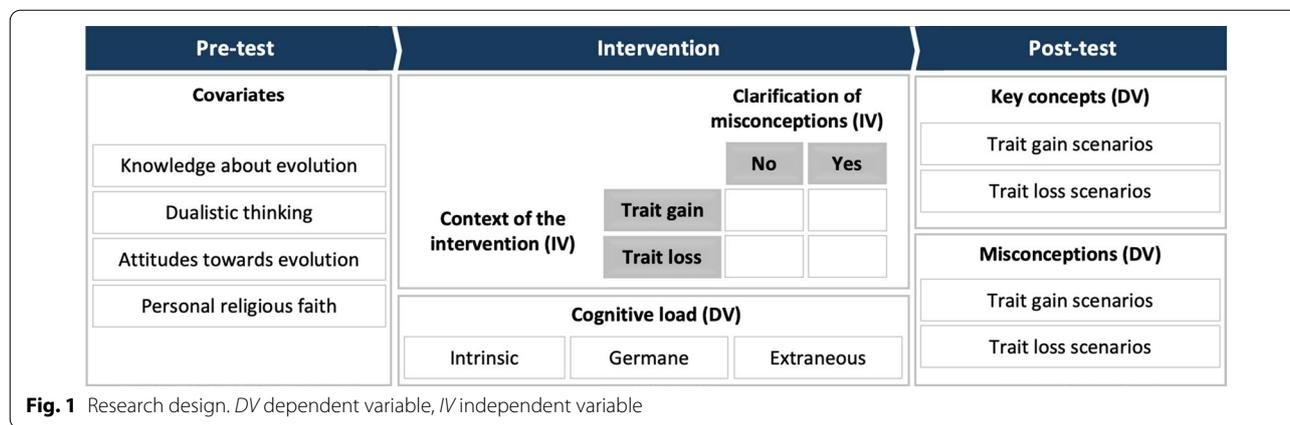
NRW] 2013a, 2015; KMK 2016). Informed by previous research on evolutionary preconceptions in young children (Evans 2008; Kampourakis et al. 2012) as well as secondary students (Beniermann 2019; Kuschmierz et al. 2020a, b; Fenner 2013; Lammert 2012) and by the core curricula for lower secondary schools (ISCED 2; KMK 2020; MSW NRW 2011a, b, 2013b; Ministry for School and Education of the State of North Rhine-Westphalia [MSB NRW] 2019), we cannot exclude that students hold at least fragmented knowledge about natural selection.

Research design and procedures
We conducted a 2×2 factorial intervention study with an experimental design. The study consisted of three phases: a pre-test, an intervention, and a post-test (cf. Fig. 2). It was conducted in students' regular school environments to ensure ecological validity. To warrant comparability

across schools, we revised the core curricula of the lower secondary school (previous education) and upper secondary schools (current education) regarding the evolutionary contexts (Eurydice 2021; KMK 2016; MSB NRW 2019; MSW NRW 2011a, b, 2013a, b, 2015). We instructed the implementers of the study on standardized procedures. The design and instructions (e.g., paper-based material, amount of time to work on tasks, content, attention to key concepts, group-specific attention to misconceptions, learning form) were similar between the schools and classes.

In the pre-test (45 min), we collected variables considered essential for analyzing learning, including knowledge about evolution, dualistic thinking, attitudes towards evolution, personal religious faith, and demographic data (Park 2007; Deniz et al. 2008; Großschedl et al. 2014). In the subsequent intervention (90 min),

Fig. 2 Distribution of students during the course of the study. The figure is based on the Consolidated Standards of Reporting Trials (CONSORT) flow diagram (CONSORT 2010). GMC−=group that learned in the context of trait gain without a clarification of misconceptions; GMC+=group that learned in the context of trait gain with a clarification of misconceptions; LMC−=group that learned in the context of trait loss without a clarification of misconceptions; LMC+=group that learned in the context of trait loss with a clarification of misconceptions

Aptyka *et al. Evolution: Education and Outreach*   (2022) 15:5

Page 7 of 21

students of each class were randomly assigned to four intervention groups (i.e., random assignment was at the student level). Each group received instructional material covering natural selection, which was manipulated in terms of the evolutionary context (trait gain [G] vs. trait loss [L]) and the clarification of misconceptions (no clarification of misconceptions [MC−] vs. clarification of misconceptions [MC+]). Thus, there were a total of four groups of which one each learned in contexts of trait gain without a clarification of misconceptions (GMC−), trait gain with a clarification of misconceptions (GMC+), trait loss without a clarification of misconceptions (LMC−), and trait loss with a clarification of misconceptions (LMC+). Directly after learning with the instructional materials, students reported their cognitive load using the cognitive load questionnaire that we provided to them on the last page of the instructional materials. A few days after the intervention (on average $M = 1.4$), the post-test (45 min) was administered to gain insights into students' contextual reasoning about natural selection. We used three test versions comprising the same four tasks but in different orders to reduce possible influences of task sequencing effects (Federer et al. 2015). Each student received all four tasks.

Instructional materials for the intervention

The instructional materials covering natural selection (see Additional file 1) commenced with definitions of basic terms, such as 'population', 'individual', and 'trait', to prevent comprehension problems. For presenting natural selection in a didactically valuable way, we developed the tasks concerning research on situated learning, knowledge building, and cognitive load theory. We chose text-based worked examples in the instructional materials to facilitate profound, interconnected, coherent knowledge and increase the learning-relevant germane load while reducing the extraneous load (Barsalou 2016; Leppink and Heuvel 2015; Paas et al. 2003; Renkl 2005; Richey and Nokes-Malach 2015). The worked examples were composed of writing tasks with best-practice solutions. They addressed four trait change scenarios based on natural selection, two animal scenarios (i.e., locusts [trait: dichlorodiphenyltrichloroethane (DDT) resistance], flies [trait: flight ability]) and two plant scenarios (i.e., grapes [trait: tendrils], cacti [trait: thorns]). In the writing tasks, students had to explain processes of natural selection based on the four mentioned scenarios. To ensure the validity of the conclusions drawn from the intervention and the post-test and imply familiarity with the response format of the post-test, the intervention materials already contained tasks inspired by the Assessment of Contextual Reasoning about Natural Selection (ACORNS) instrument (Nehm et al. 2012). For example, one task

was defined as follows: '*Explain how a locust population without DDT resistance could develop into a locust population resistant to DDT. First, try to answer this question for yourself*'. The best-practice solutions introduced an explanation of the underlying scenarios based on the following key concepts: Presence and cause of variation ('variation'), the heritability of variation ('heritability'), differential survival of individuals ('individual fitness'), and limited resources ('resource limitation'; for short definitions of the concepts, see Additional file 2).

The manipulation of the instructional materials in terms of the evolutionary context (i.e., trait gain vs. trait loss) and the clarification of misconceptions (i.e., no vs. yes) was implemented as follows: The groups learning in trait gain or trait loss scenarios only differed in the fact that the first-mentioned group received scenarios of *trait gain* and the second-mentioned group received scenarios of *trait loss* (e.g., '*Random changes (mutations)[…] may have led to some locust individuals **to gain/to lose** resistance to DDT*'). Besides, the differences in clarification of misconceptions were realized through four additional reasoning tasks embedded in the existing four trait change scenarios. These tasks consisted of three steps: First, the material presented an example of a misconception about one of the four existing scenarios (e.g., locust). Second, the tasks encouraged the students to explain why the statement was technically incorrect. Third, the misconception in the example was explained by an informational text. Accordingly, half of the materials contained global informational text on the following misconceptions: Need as a driving force for evolution ('need'), the use or disuse of particular body parts ('use/disuse'), the intentionality to change ('intentionality'), and the active adaptation to environmental conditions ('adapt'; e.g., *The 'adaptation' to environmental conditions has no effect on the hereditary characteristics of individuals in a population. It, therefore, plays no role in the transmission of traits to the next generation*; for short definitions of the concepts, see Additional file 2). Aside from the presented manipulations, other parts of the instructional materials were isomorphic.

Measures

The following outlined measures are appropriate for the examined target sample (Beniermann 2019; Kuschmierz et al. 2020b; Nehm et al. 2012; Klepsch et al. 2017). For the measures that students rated on a 5-point or 7-point Likert scale, a high score represents an increased expression of the characteristic in question. Table 1 presents reliability scores and descriptive statistics of all measures.

Pre-test variables

Knowledge about evolution We employed the Knowledge About Evolution 2.0 (KAEVO 2.0; Kuschmierz et al. 2020b) to assess students' knowledge about evolution, especially about different aspects of micro- and macro-evolution. We used the KAEVO 2.0 as this measure was validated using multiple evidence for validity (e.g., content validity, internal structure, and reliability), developed for German high school students, and it was thematically orientated towards biology curricula and textbooks contents (Beniermann 2019; Kuschmierz et al. 2020b; Beniermann et al. 2021). Moreover, its underlying context is similar to the post-test assessment (see the following section 'Contextual reasoning about natural selection'). Both allow the evaluation of knowledge about concepts of natural selection (e.g., variation) in animal and plant scenarios concerned with the gain or loss of traits (Nehm et al. 2012). The KAEVO 2.0 entails multiple-choice questions, true/false statements, and distractors based on existing misconceptions about evolution. In total, students could score up to 24 points by answering all tasks correctly (Kuschmierz et al. 2020b). Assessing knowledge about evolution as a covariate is crucial for the following analyses, as previous studies showed a positive effect of knowledge on learning. Depending on its nature, knowledge can act as a filter or reinforcer in learning (Barsalou 2016; Deniz et al. 2008).

Dualistic thinking We assessed dualistic thinking with the Short Dualistic scale (SD-scale; Beniermann 2019). This scale represents an abridged version of Stanovich's

Dualism scale (Stanovich 1989). The substantive, content, and internal validity were ensured using expert interviews, a pre-test for testing comprehensibility as well as reliability, and factor analyses (Beniermann 2019). The scale comprised five items on a 5-point Likert scale and has been applied in previous studies to examine the extent to which students hold dualistic theories about the brain and the mind and reject materialistic accounts (Beniermann 2019; Beniermann et al. 2021; Stanovich 1989). This variable is an important covariate because it is part of the belief system of individuals and thus of the conceptual ecology (Deniz et al. 2008). Furthermore, current research emphasises the need to investigate dualistic thinking because it negatively relates to knowledge about evolution (Beniermann 2019; Beniermann et al. 2021).

Attitudes towards evolution We captured the attitudes towards evolution by employing the Attitudes Towards Evolution 2.0 (ATEVO 2.0; Beniermann 2019). Internal validity for this version was determined using statistical analyses such as principal component analysis (Beniermann 2019). Overall, this measure contains eight items on a 5-point Likert scale that address attitudes towards the philosophical position of evolutionary epistemology. Four items focus on attitudes towards the human spirit and four on attitudes towards evolution in general (Beniermann 2019; Beniermann et al. 2021). Prior studies showed that assessing the attitudes towards evolution as a covariate is crucial for an informed interpretation and analysis of the student's learning process. They are positively related to knowledge about evolution and learning

Table 1 Reliability scores and descriptive statistics of the measures

Variable	Number of items	*M*	*SD*	*α*	*ω**	*κ**
Knowledge about evolution	24	12.39	4.77	.82	.83	N/A
Dualistic thinking	5	2.96	0.97	.75	.75	N/A
Attitudes towards evolution	8	4.00	0.54	.65	.68	N/A
Personal religious faith	10	2.60	1.26	.96	.96	N/A
Cognitive load	7	3.64	1.16	.80	.81	N/A
Intrinsic cognitive load	2	4.00	1.53	.72	.72	N/A
Germane cognitive load	2	3.98	1.41	.76	.76	N/A
Extraneous cognitive load	3	2.94	1.34	.71	.72	N/A
Key concepts	4	6.87	4.13	N/A	N/A	.61
Trait gain scenarios	2	3.57	2.26	N/A	N/A	.58
Trait loss scenarios	2	3.29	2.28	N/A	N/A	.64
Misconceptions	4	2.26	2.69	N/A	N/A	.64
Trait gain scenarios	2	0.90	1.47	N/A	N/A	.62
Trait loss scenarios	2	1.37	1.63	N/A	N/A	.66

[*] We calculated scores with the RStudio packages 'psych' (ω) and 'irr' (κ). We included the reliability indices for *α* and *ω* in this table to indicate the scales' internal consistency and ensure comparability with previous studies such as Klepsch and Seufert (2020), who only used *ω* for reporting the reliability. *α*=Cronbach's alpha, *κ*=Cohen's Kappa (Inter-Rater-Reliability), *M*=mean score, *SD*=standard deviation, *ω*=McDonald's omega

(Beniermann 2019; Kuschmierz et al. 2020; Fenner 2013; Lammert 2012; Deniz et al. 2008; Beniermann et al. 2021).

Personal religious faith   We used the Personal Religious Faith 2.0 (PERF 2.0; Beniermann 2019) to assess monotheistic faith and religious behaviors (Beniermann et al. 2021). This measure comprises ten items on a 5-point Likert scale. Experts from philosophy, theology, religion, psychology, and sociology evaluated and modified the instrument to guarantee content validity (Beniermann 2019). It is essential to determine personal religious faith and use it as a covariate in analysing learning, as this variable is part of the conceptual ecology. The examination of this variable is also essential as current studies on the correlation between personal religious faith and knowledge about evolution are inconsistent. They mostly show no or a negative correlation between the two variables (Beniermann 2019; Kuschmierz et al. 2020; Deniz et al. 2008).

Intervention variable

Cognitive load   We assessed students' cognitive load or mental effort by using the cognitive load questionnaire of Klepsch et al. (2017). The substantive and content validity of this questionnaire was demonstrated by deriving its items from literature, comparing it to former instruments, and contrasting informed as well as naïve ratings. Also, Klepsch et al. (2017) tested the internal structure of this questionnaire through statistical analyses such as expert rating agreements and confirmatory factor analysis (Klepsch and Seufert 2020; Klepsch et al. 2017). The questionnaire consists of seven items on a 7-point Likert scale. Since the cognitive load encompasses three types, three scales are assessed. Two items of the questionnaire measure the intrinsic, two measure the germane, and three measure the extraneous cognitive load. The questionnaire has proven valuable in identifying which parts of tasks are cognitively challenging for students (Klepsch and Seufert 2020; Klepsch et al. 2017). Similarly, we used this measure to gain insight into the perceived difficulties of the interventions' instructional materials.

Post-test variable

Contextual reasoning about natural selection   We applied the ACORNS (Nehm et al. 2012) in the post-test. The ACORNS was originally validated by providing convergent validity. This instrument is used regularly to gain insights into students' contextual reasoning about natural selection. It differs from the previously presented KAEVO 2.0 in, among other things, the response process (Opfer et al. 2012; Nehm and Ha 2011; Federer et al. 2015). We applied four open-ended tasks and promoted students to reason about natural selection in written form. The four tasks had an isomorphic structure but differed in the evolutionary contexts (trait gain in a species of snails [toxicity] or elms [winged seeds]; trait loss in a species of penguins [ability to fly] or roses [spines]). Trained independent human raters manually coded each item according to Nehm et al. (2010). We used the manual to identify and quantify the number of key concepts and misconceptions used in trait gain and trait loss scenarios. For this analysis, we focused on the four key concepts which were used in the instructional material, namely the 'variation', 'heritability', 'individual fitness', and 'resource limitation', as well as the four misconceptions, 'need', 'use/disuse', the 'intentionality', and 'adapt' (Bishop and Anderson 1990; Nehm and Ha 2011; Großschedl et al. 2018; Nehm et al. 2010; Rachmatullah et al. 2018; Rector et al. 2013). The presence of a concept was tallied and dichotomously coded for each task (e.g., concept absent = 0, concept present = 1). We calculated a sum score for the number of used key concepts for each of the four post-test tasks and assigned them to the two different scenarios, trait gain and trait loss. Correspondingly, we obtained scales for the number of key concepts used in trait gain scenarios and trait loss scenarios, where the maximum score for each scale was eight (e.g., '8 concepts used'). We calculated the number of misconceptions used in trait gain and trait loss scenarios in the same way. To ensure the reliability of the open response rating, we evaluated the data material of the two raters and received a substantial overall score (cf. Table 1; Landis and Koch 1977).

Data analysis

We analysed the data using SPSS IBM Statistics (version 27.0) and resorted to RStudio for calculating reliability measures ω and κ. Initially, we screened the data and its distribution. Missing values were partly missing at random (MAR), so we excluded the respective cases from individual data analyses. The data also showed extreme statistical outliers for the attitudes towards evolution ($n = 3$) and age ($n = 6$), which we eliminated from all analyses (cf. Fig. 2). We used descriptive statistics for the overall sample specification. We explicitly examined knowledge about evolution in the pre-test and the type of used key concepts and misconceptions in the post-test. We set up a formal model description as a guideline for our following analyses (cf. Additional file 3). For null hypothesis significance testing, we used inferential statistics and set the significance level for the statistical analyses of our research hypotheses to 5%.

Previous studies have shown that pre-test knowledge about evolution, dualistic thinking, attitudes towards evolution, and personal religious faith are related to thinking and reasoning about natural selection (Park

2007; Deniz et al. 2008; Großschedl et al. 2014). We applied these pre-test variables as covariates for all following analyses related to reasoning about natural selection to improve our power and reduce unexplained variability between groups (Maxwell et al. 2017). Since the study design allowed examining the first hypothesis, third hypothesis, and the exploratory analysis in a single step, we performed only one analysis to answer these. Although the analysis is carried out jointly, we presented the results in chronological order of the previously stated hypotheses (see the section 'Current study'). We conducted a two-way multivariate analysis of covariance (MANCOVA). Explicitly, we investigated the effect of the evolutionary context of the intervention (trait gain vs. trait loss; H1), the impact of the clarification of misconceptions (no vs. yes; H3), and the interaction effect of both (exploratory analysis) on the later use of key concepts and misconceptions, both in trait gain and trait loss scenarios of the post-test. The two-way MANCOVA was used, as it is essential for identifying the most effective instructional materials. Even if the two main effects of this analysis are not significant, the interaction term can still be significant. From a statistical perspective, the significant interaction means that the two factors should not be interpreted globally but in combination to obtain a holistic analysis (Maxwell et al. 2017).

Afterward, we performed a two-way multivariate analysis of variance (MANOVA) to address our second hypothesis (H2). We examined the effects of the intervention material, explicitly the evolutionary context of the intervention (trait gain vs. trait loss), the clarification of misconceptions (no vs. yes), and the interaction effect of both on students' perceived intrinsic, germane, and extraneous cognitive load.

As a continuation of the exploratory analysis, we also aimed to provide a more general statement about the effects of the four experimental conditions of the intervention (differing in the evolutionary context [trait gain vs. trait loss] and the clarification of misconceptions [no vs. yes]) on the use of key concepts and misconceptions in the post-test per se. Thus, we calculated a total score for the used key concepts (key concepts in trait gain plus trait loss scenarios) and misconceptions (misconceptions in trait gain plus trait loss scenarios). Next, we applied a two-way MANCOVA. In this analysis, we were primarily interested in the interaction effect of the evolutionary context of the intervention (trait gain vs. trait loss) and the clarification of misconceptions (no vs. yes) for the use of key concepts and misconceptions in the post-test.

Results

Baseline description

We used descriptive statistics to visualize the comparability of the pre-test variables and demographic data among the four intervention groups (cf. Table 2).

About the pre-test variables, we explicitly examined knowledge about evolution in more detail. Our sample achieved $M = 12.39$ points in the pre-test, which is classified as low knowledge according to the score categories of Kuschmierz et al. (2020b). Moreover, we examined four items of the KAEVO 2.0 that are similar to the ACORNS in that the items cover evolutionary adaptation and natural selection. In the tasks, students used more key concepts when trait gain (37.2%) compared to trait loss (28.8%) was addressed. Most of the misconceptions used in all items are based on teleological ideas, especially concerning the organism (22.4%) itself.

Subsequently, we analyzed the mean frequency of the used key concepts and misconceptions when reasoning about evolutionary scenarios within the four ACORNS items in the post-test (cf. Fig. 3). The figure shows that students used about twice as many keys concepts as misconceptions when reasoning natural selection. They used the key concept 'resource limitation' most frequently and 'heritability' least frequently. In addition, the misconceptions 'adapt', 'need', and 'use/disuse' occurred nearly equally often, with 'intentionality' being used least frequently.

Effects of situated learning on later reasoning

We hypothesized that students use more key concepts and fewer misconceptions in a context if they had previously learned in a similar one. We interpreted only the factor context of the intervention (trait gain vs. trait loss) of the two-way MANCOVA and used Wilks's statistics to test our first hypothesis. Results did not reveal a significant effect of the evolutionary context of the intervention (trait gain vs. trait loss) on the use of key concepts and misconceptions, both in trait gain and trait loss scenarios in the post-test, Wilk's $\Lambda = 0.99$, $F(4,301) = .69$, $p = .603$, $\eta_p^2 = .01$. The results did not support our first hypothesis since groups that learned in the context of trait gain (or trait loss) did not use significantly more key concepts nor fewer misconceptions in trait gain (or trait loss) tasks after instructions. Likewise, descriptive statistics only showed marginal differences. Both groups used more key concepts and fewer misconceptions in trait gain scenarios than in trait loss scenarios (cf. Table 3).

Cognitive load while learning

Regarding our second hypothesis, we expected that only the intrinsic load would be lower in trait gain than

Table 2 Descriptive statistics of the groups

Variable	GMC−			GMC+			LMC−			LMC+		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Knowledge about evolution	91	12.00	4.78	96	12.69	5.26	89	12.22	4.65	88	12.63	4.36
Dualistic thinking	82	3.04	0.92	87	2.87	1.06	82	2.88	0.96	83	3.04	0.95
Attitudes towards evolution	82	4.04	0.45	87	3.99	0.61	82	3.97	0.50	82	3.98	0.59
Personal religious faith	82	2.88	1.29	87	2.54	1.24	82	2.53	1.26	83	2.45	1.21
Age	83	16.68	0.95	87	16.54	0.91	82	16.65	1.08	85	16.77	1.01
Grade	91			96			89			88		
Gender	82			87			82			85		
Male	32			32			26			34		
Female	50			55			56			51		
Upper secondary school	91			96			89			88		
Gymnasium	62			69			61			63		
Comprehensive school	14			13			12			11		
Vocational training	15			14			16			14		

We measured knowledge about evolution on a scale ranging from 0 to 24. We assessed dualistic thinking, attitudes towards evolution, and personal religious faith on a 5-point Likert scale ranging from 1 = low to 5 = high. GMC− = group that learned in the context of trait gain without a clarification of misconceptions; GMC+ = group that learned in the context of trait gain with a clarification of misconceptions; LMC− = group that learned in the context of trait loss without a clarification of misconceptions; LMC+ = group that learned in the context of trait loss with a clarification of misconceptions
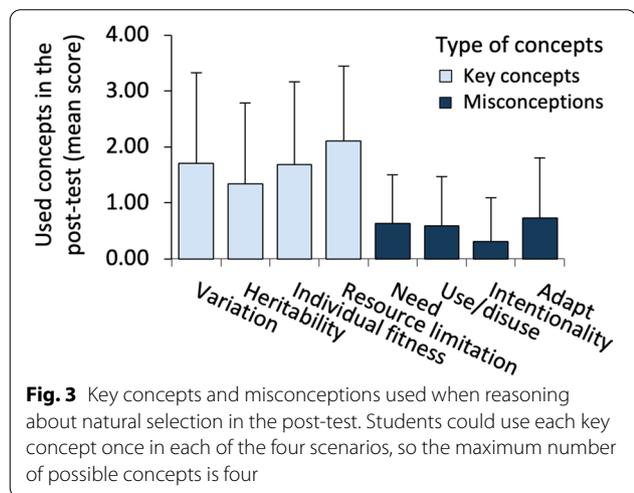

Fig. 3 Key concepts and misconceptions used when reasoning about natural selection in the post-test. Students could use each key concept once in each of the four scenarios, so the maximum number of possible concepts is four

Table 3 The used key concepts and misconceptions when reasoning about different evolutionary scenarios in the post-test

ACORNS items	G		L	
	M	*SD*	*M*	*SD*
Key concepts				
Trait gain scenarios	3.60	2.26	3.55	2.25
Trait loss scenarios	3.18	2.26	3.41	2.29
Misconceptions				
Trait gain scenarios	0.86	1.39	0.93	1.55
Trait loss scenarios	1.27	1.50	1.43	1.71

Students could use a maximum of eight concepts in each case. G = group that learned in the context of trait gain. L = group that learned in the context of trait loss in the intervention

in trait loss contexts of the intervention, that only the germane load would be lower when receiving no clarification of misconceptions than receiving a clarification, and that the extraneous would not differ between the intervention groups. We also tested whether the intervention context and clarification of misconceptions interact and translate into differences in the three types of loads. We conducted a two-way MANOVA using Wilk's statistics which revealed a significant difference between the experimental conditions of the evolutionary context of the intervention (trait gain vs. trait loss) for students' perceived cognitive load, Wilk's $\Lambda = .97$, $F(3,221) = 2.67$, $p = .049$, $\eta_p^2 = .04$, no significant

difference for the clarification of misconceptions (no vs. yes), Wilk's $\Lambda = .99$, $F(3,221) = 1.07$, $p = .361$, $\eta_p^2 = .01$, and a significant interaction effect, Wilk's $\Lambda = .96$, $F(3,221) = 3.08$, $p = .028$, $\eta_p^2 = .04$. Thus, the evolutionary context of the intervention (trait gain vs. trait loss) and the interaction effect relate to students' cognitive load differences. To identify the types of cognitive load which differed between the intervention groups, we conducted Bonferroni-corrected post-hoc ANOVAs. The analyses revealed a significant difference for the evolutionary context of the intervention (trait gain vs. trait loss) for the intrinsic load, $F(1,223) = 4.67$, $p = .032$, $\eta_p^2 = .02$, no significant effect for the germane load, $F(1,223) = 1.69$, $p = .195$, $\eta_p^2 = .01$, and significant effects for the extraneous load, $F(1,223) = 4.90$,

$p = .028$, $\eta_p^2 = .02$. The post-hoc analyses for the interaction effect did not show any significant differences for the intrinsic, germane, or extraneous cognitive load ($p > .05$). Groups that were provided with intervention materials on trait gain contexts (GMC− and GMC+) perceived a lower intrinsic load ($M_{Diff} = − .44$, 95%-CI [− .83, − .04]; $d_{Cohen} = .29$) and extraneous ($M_{Diff} = − .39$, 95%-CI [0.04, 0.74]; $d_{Cohen} = 0.29$) than groups that were provided with trait loss contexts. The results supported our hypothesis on intrinsic but not germane and extraneous load (cf. Fig. 4).

Effects of the clarification of misconceptions on later reasoning

The third hypothesis we posited was that students who received no clarification of misconceptions in the intervention would use fewer key concepts and more misconceptions in reasoning about natural selection in the post-test than students who received one. Contrary to our expectations, the results of the factor clarification of misconceptions (no vs. yes) in the two-way MANCOVA did not support our third hypothesis, as they did not show significant differences between groups in the use of key concepts and misconceptions, both in trait gain and trait loss scenarios, Wilk's $\Lambda = .97$, $F(4,299) = 2.01$, $p = .093$, $\eta_p^2 = .03$.

Identifying the most effective instructional strategy

To investigate whether it makes a difference in which context students received the clarification of misconceptions and identify the most effective instructional strategies of our 2×2 factorial design, we analysed the interaction term of the factors evolutionary context of the intervention (trait gain vs. trait loss) and clarification of misconceptions (no vs. yes) of the two-way

MANCOVA in respect to the use of key concepts and misconceptions, both in trait gain and trait loss scenarios in the post-test. We did not find any supporting evidence for an interaction effect, Wilk's $\Lambda = .98$, $F(4,299) = 1.74$, $p = .142$, $\eta_p^2 = .02$. From a purely descriptive perspective, the group LMC− used the most key concepts, and the group GMC+ used the fewest misconceptions, both in trait gain and trait loss scenarios (cf. Fig. 5).

Subsequently, we carried out another two-way MANCOVA to examine the interaction effect of two factors (evolutionary context of the intervention [trait gain vs. trait loss] and clarification of misconceptions [no vs. yes] on the general use of key concepts and misconceptions in the post-test. We found a statistically significant disordinal interaction effect, Wilk's $\Lambda = .98$, $F(2,301) = 3.06$, $p = .049$, $\eta_p^2 = .04$. After that, we conducted post-hoc ANCOVAs for both dependent variables. These analyses showed no statistically significant differences between the four groups regarding the use of key concepts, $F(1,302) = .78$, $p = .377$, $\eta_p^2 < .01$, but did regarding the use of misconceptions, $F(1,302) = 5.92$, $p = .016$, $\eta_p^2 = .02$ (cf. Fig. 6). Bonferroni-corrected post-hoc tests revealed that students who learned in trait gain contexts without a clarification of misconceptions (GMC−) used significantly more misconceptions than students who learned in trait gain contexts with the clarification of misconceptions (GMC+), $p = .004$ ($M_{Diff} = 1.22$, 95%-CI [0.40, 2.05]; $d_{Cohen} = .54$). The other groups did not differ significantly from each other.

Discussion

A large body of empirical research in evolution education has analysed key concepts and misconceptions in reasoning about natural selection (Opfer et al. 2012; Beggrow and Sbeglia 2019; Nehm and Ha 2011; Federer et al. 2015) and found that students' reasoning depends on the
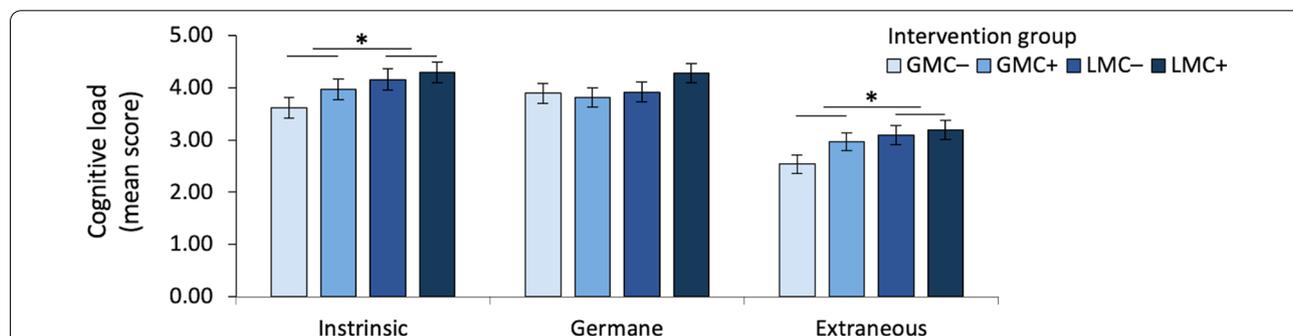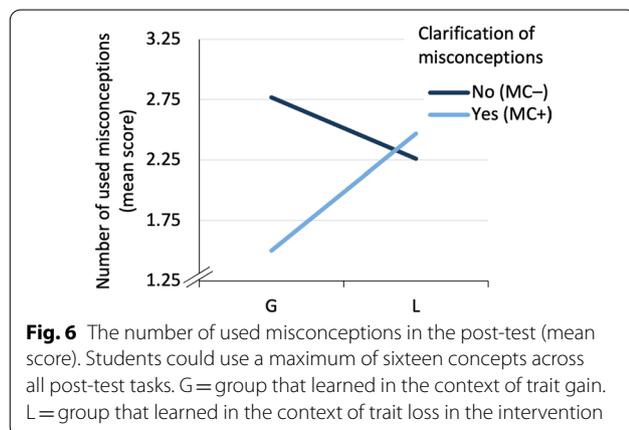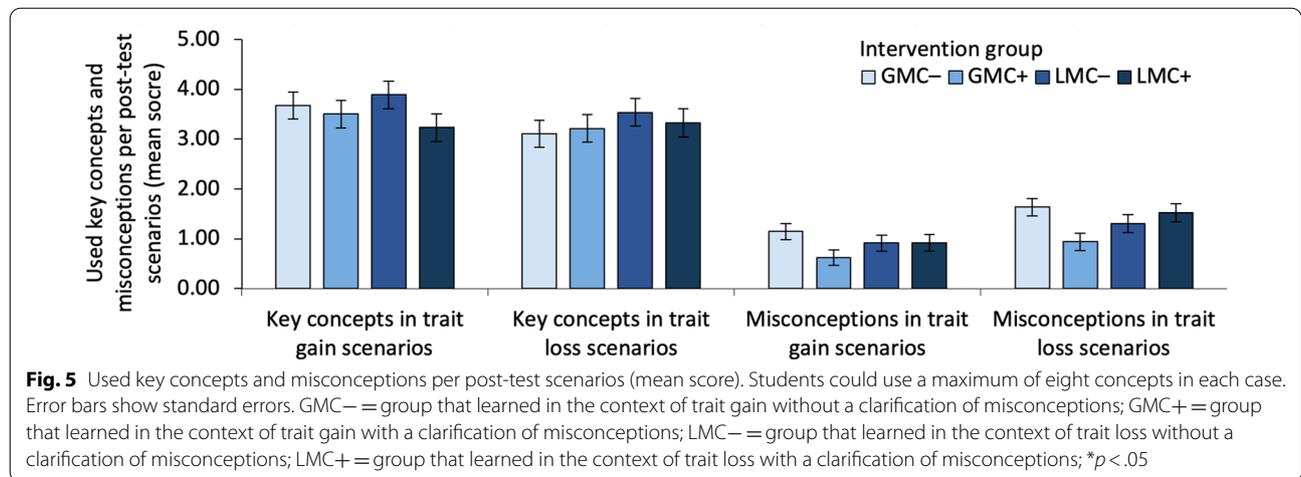

Fig. 4 Intrinsic, germane, and extraneous cognitive load in the intervention (mean score). Cognitive load was measured on a 7-point Likert scale ranging from 1 = low to 7 = high. Error bars show standard errors. GMC− = group that learned in the context of trait gain without a clarification of misconceptions; GMC+ = group that learned in the context of trait gain with a clarification of misconceptions; LMC− = group that learned in the context of trait loss without a clarification of misconceptions; LMC+ = group that learned in the context of trait loss with a clarification of misconceptions; *$p < .05$

Fig. 5 Used key concepts and misconceptions per post-test scenarios (mean score). Students could use a maximum of eight concepts in each case. Error bars show standard errors. GMC− = group that learned in the context of trait gain without a clarification of misconceptions; GMC+ = group that learned in the context of trait gain with a clarification of misconceptions; LMC− = group that learned in the context of trait loss without a clarification of misconceptions; LMC+ = group that learned in the context of trait loss with a clarification of misconceptions; *$p < .05$

Fig. 6 The number of used misconceptions in the post-test (mean score). Students could use a maximum of sixteen concepts across all post-test tasks. G = group that learned in the context of trait gain. L = group that learned in the context of trait loss in the intervention

underlying evolutionary context. Furthermore, previous studies stressed that it is easier for students to reason about some evolutionary contexts (e.g., natural selection covering animals; trait gain) than others (Nehm and Ha 2011; Federer et al. 2015; Großschedl et al. 2018). Moreover, previous studies repeatedly addressed the persistence of misconceptions (Ha and Nehm 2014; Nehm and Ha 2011; Nehm and Reilly 2007). However, it remained unclear which factors are linked to differences in contextual reasoning, whether students perceive different contexts to vary in difficulty, and what kind of instructional strategies promote learning about natural selection. We investigated prior situated learning to clarify which explanatory approach can explain contextual reasoning. In addition, we explored whether students perceive different contexts to vary in difficulty by monitoring their cognitive load while learning. Moreover, we explored learning instructions that differed in the underlying evolutionary context (trait gain vs. trait loss) and in the clarification of misconceptions (no vs. yes) to identify the most effective opportunities.

Evolutionary contexts and the situated learning approach

Regarding our research hypothesis on situated learning, the results indicated that the intervention groups learning with trait gain contexts did not use significantly more key concepts or fewer misconceptions than the groups learning with trait loss contexts when reasoning about natural selection in trait gain scenarios in the post-test. The same was true for the students who learned in trait loss contexts.

We suspect that prior knowledge, cognitive load, the design of the instructional materials, or effectiveness of knowledge transfer between tasks may explain the lack of effect as discussed below:

As our pre-test results showed, we cannot assume that each student resembled a *tabula rasa* when participating in the study. The students showed little knowledge about evolution and used a mix of key concepts and misconceptions to answer adaptation and natural selection items. Therefore, there exists the possibility that the students either did not have sufficient prior knowledge to draw on when working on the intervention materials or that the existing misconceptions posed a barrier to learning and overwhelmed the working memory when students endeavored to overcome them (e.g., Nehm and Ha 2011; Goel et al. 2010; Barsalou 2016).

It also seems possible that the cognitive load in the intervention was the decisive factor in why students' concept use did not significantly differ. For example, the intrinsic cognitive load differed significantly between the groups that learned in trait gain and trait loss contexts. Since the intrinsic cognitive load was already relatively high in the intervention group, which learned with trait loss contexts, it could have tied up a large proportion of the cognitive resources. As a result, students may have had fewer resources to acquire knowledge in the intervention and apply it to the scenarios in

Aptyka *et al. Evolution: Education and Outreach*   (2022) 15:5

Page 14 of 21

the post-test. Thus, the results of the second hypothesis could explain why students who learned in trait loss contexts in the intervention did not perform better in trait loss scenarios than in trait gain scenarios in the post-test.

Taking this idea further, another reason could be that students benefited from the instructional materials ('source'), but the learned information was not directly accessible in other situations ('target'). This means that although students have learned about certain contexts and concepts of natural selection in the intervention, their knowledge initially remained tied to the concrete instructional situation. According to Kirsh (2009), mental representations are always bound to a specific situation. For students to transfer knowledge to other situations, they must understand the deeper structure of the given problem and internalize its abstract representation. Consequently, understanding the post-test tasks ('target') and underlying abstract structures is essential for the transfer. If students did not fully understand the post-test tasks, they might not have understood more profound analogies between the source and the target, and the transfer was unlikely to occur. The transfer of knowledge between the source and target tasks could have been hampered if the students perceived the tasks as dissimilar (Reder et al. 1994; Anderson et al. 1996). Kirsh (2009) described a comparable case in which individuals could not transfer their problem-solving strategies from *tic-tac-toe* to the *game of fifteen*, even though both games were based on the same problem and differed only in surface features.

However, the opposite could also be true, meaning that students could transfer knowledge to trait gain and trait loss scenarios of the post-test. We cannot exclude that similar structures of the intervention materials and post-test tasks contributed to the students being able to transfer their acquired knowledge equally well to all post-test tasks. Therefore, students might have had the subjective perception that the post-test tasks matched in terms of required knowledge. For instance, students who learned in trait gain contexts could apply their knowledge equally well in post-test scenarios of trait gain and trait loss.

The complex interplay of intrinsic, germane, and extraneous cognitive load while learning

Regarding our research on cognitive load, the results showed that intrinsic cognitive load was significantly lower when students received interventional material on trait gain rather than trait loss contexts. This effect supports assumptions posed by the cognitive load theory concerning intrinsic load, as intrinsic load represents the perceived complexity placed on the students by context (Klepsch et al. 2017). Former results show that students use more key concepts (Federer et al. 2015) and fewer misconceptions (Nehm and Ha 2011) when explaining natural selection in trait gain scenarios. We added to the body of research by highlighting that students perceive a lower intrinsic load while working on trait gain compared to trait loss contexts.

Due to research suggesting that a clarification of misconceptions can promote conceptual transformation (Kampourakis and Zogza 2009; Andrews et al. 2011; Limón 2001; Nelson 2008; Colton et al. 2018), we expected the clarification of misconceptions to foster deeper learning and thus increase the germane cognitive load (Klepsch and Seufert 2020; Klepsch et al. 2017). Nevertheless, we did not find significant differences between the intervention groups. Klepsch and Seufert (2020) explicated that if high intrinsic and extraneous load levels are imposed on the working memory, only a little working memory capacity is available for germane load. Aligned with the researchers, we argue that all groups' intrinsic and extraneous load could have demanded substantial working memory. Compared to other studies, the assessed cognitive load was relatively high (Klepsch and Seufert 2020; Klepsch et al. 2017). Thus, possibly insufficient capacity was available to acquire new knowledge. Another reason might be the clarification of misconceptions which was not supportive for every student. Our students showed low levels of knowledge about evolution. The results of Heemsoth and Heinze (2016) supported this argument by indicating that a clarification of misconceptions can lead to cognitive overload and disadvantages for students with little prior knowledge but advantages for students with higher knowledge levels.

From a theoretical viewpoint, we also expected that the groups' extraneous cognitive load would not differ. From a practical point of view, the reality is less straightforward. Data revealed that the extraneous load was significantly lower when students learned in trait gain than trait loss contexts in the intervention. Given that the instructional design was the same in both contexts, differing only in the information that traits were *gained* or *lost*, other confounding variables must explain this difference in extraneous load. This result could be attributable to the fact that the students were generally more familiar with contexts of trait gain. The overall load in trait gain was lower than in trait loss contexts (Nehm and Ha 2011; Orru and Longo 2019). Comparably, Shen et al. (2020) found that working memory is not significantly strained when students work with familiar compared to unfamiliar components, as they investigated learning with familiar and unfamiliar icons.

In line with the findings of Klepsch and Seufert (2020), the natural complexity of the instructional materials could have prevented students from distinguishing

between the complexity of tasks (intrinsic load) and instructional design (extraneous load), leading to both being rated as high (Klepsch and Seufert 2020). Overall, the interplay of the three loads may have almost exhausted the maximum capacity of the working memory. The measured values were higher than in other studies, and working memory capacity is limited (Klepsch and Seufert 2020; Klepsch et al. 2017). As a result, students could have struggled with differentiating between the loads. It is also not yet evident what exact ratio of the three loads promotes the best possible learning success and what the critical level of cognitive overload is (Jong 2010).

The loads increased mostly uniformly from the groups GMC− to LMC+, whereas in other studies, the loads tended to change individually. Thus, more research is needed to determine whether the validity of this instrument is limited when students perceive a relatively high level of cognitive load. In addition, we would like to accentuate the practical significance of the fact that trait gain contexts are associated with a lower intrinsic and extraneous cognitive load than trait loss contexts. The effect sizes of the differences for intrinsic and extraneous load are small. The interpretation of extraneous load is limited and would require further research.

The clarification of misconceptions as instructional strategy

Contrary to our expectations, the sole inclusion of clarification of misconceptions did not significantly increase the germane load or reduce the use of misconceptions in trait gain or trait loss scenarios. A plausible reason is that students did not perceive the clarification as supportive. The intervention may have been insufficiently designed or implemented. Similar to the results of Heemsoth and Heinze (2016), a clarification of misconceptions could have been effective for students with higher prior knowledge but inhibiting for students with lower prior knowledge. Since the MANCOVA only provides a mean value for the groups, no differences are apparent. The average of the students did not benefit more from the clarification than the students who did not receive it.

Furthermore, the allotted interaction time may not have been sufficient to impact learning and subsequent outcomes in the post-test. Developing scientific knowledge of natural selection can take longer than one semester (Nehm and Reilly 2007). Misconceptions about evolution are cognitively, deeply rooted and resistant to teaching (Gregory 2009). Thus, students could benefit from more time to revise individual concepts and connect new with existing knowledge.

The combination of the clarification of misconceptions and different evolutionary contexts in learning instructions

Our exploratory analyses aimed to identify the instructional strategies that indicate the highest potential for learning about natural selection. Therefore, we investigated the interaction of the 2×2 intervention design and its effects on the use of key concepts and misconceptions, both in trait gain or trait loss scenarios. The results showed no evidence of an effect on the dependent variables. Nevertheless, from a general perspective of contextually detached use of key concepts and misconceptions in the post-test, the results revealed a small but noteworthy significant interaction effect. The group that learned in trait gain contexts without the clarification of misconceptions used significantly more misconceptions than students who learned in the same context and received a clarification of misconceptions. One explanation for why students only benefited from clarifying misconceptions in trait gain contexts suggests that the intrinsic and extraneous cognitive load was significantly lower in the groups GMC− and GMC+ than in the groups LMC− and LMC+. Accordingly, the cognitive load could have inhibited learning in the trait loss groups. An alternative explanation proposes that students were more familiar with the context of trait gain than with trait loss and were, therefore, better positioned to use the free working memory capacity to avoid misconceptions (Kirsh 2009; Reder et al. 1994; Barsalou 2016).

Although the students knew little about evolution during the pre-test, they could have been more familiar with contexts of trait gain. Being familiar with the underlying context could have provided a basis for transforming students' prior knowledge in the intervention because familiarity can facilitate memory formation (Poppenk et al. 2010). There may be parallels to Colton et al. (2018), who observed that learning gains between pre- and post-test scores were significantly higher in the trait gain than trait loss scenarios.

With respect to the covariates used in this model, we noticed that knowledge about evolution was of primary importance, as it was significant for all dependent variables. More prior knowledge is associated with higher numbers of used key concepts and lower numbers of used misconceptions in the post-test (Nehm et al. 2012; Deniz et al. 2008). The attitude towards evolution indicated that it is significant for the key concepts used in the trait gain scenarios of the post-test. These results are consistent with Kuschmierz et al. (2020), who summarised that the link between knowledge about evolution and attitude appears to be absent or weak, especially in primary and secondary school students. Dualistic thinking showed significant results for key concepts used in the gain scenarios and, similarly to attitudes towards

evolution, appeared to be of secondary importance. Personal religious faith represented no significant covariate. The results on religiosity suggested that religiosity is not directly associated with the learning outcomes of the sample in question and may be more closely related to other constructs of conceptual ecology, such as attitudes toward evolution (Beniermann 2019; Kuschmierz et al. 2020; Deniz et al. 2008). Overall, the pre-test knowledge about evolution was the most meaningful covariate in our analyses and should be included in future studies.

Implications for research and education
Our work presents statistical and practical significance and research relevance (Mohajeri et al. 2020). It contributes to education in schools and research on evolution education, specifically on situated learning, the cognitive load, and the clarification of misconceptions.

I. In line with current research findings (Beniermann 2019; Kuschmierz et al. 2020b), our students showed insufficient knowledge about evolution. This lack of knowledge was evident in the pre-test when students increasingly chose teleological misconceptions as answers in tasks on adaptation and natural selection (Beniermann 2019; Kuschmierz et al. 2020b; Fenner 2013; Lammert 2012). Additionally, when analysing the pre-test descriptive results, we found that students used more key concepts in trait gain tasks than trait loss (see 'Baseline description' section). In addition, our descriptive post-test results are consistent with the current state of research as they indicated that the average students used more key concepts and fewer misconceptions when reasoning about trait gain compared to the trait loss scenarios (Opfer et al. 2012; Ha and Nehm 2014; Nehm and Ha 2011). The pattern, which was already discovered in the historical development of biologists' knowledge of evolution, therefore also persisted throughout our study, regardless of the instructional context (Ha and Nehm 2014). Since misconceptions still account for a considerable amount of students' answers in the pre- and post-test, we would like to reiterate the need to find conducive instructional strategies for students.

II. To the best of our knowledge, previous studies have primarily investigated reasoning about natural selection using a cross-sectional study design (e.g., Ha and Nehm 2014; Nehm and Ha 2011; Federer et al. 2015). Thus, our study goes beyond previous research in that we not only looked at single events but also included prior situated learning. We found that learning instructions on trait gain or trait loss

contexts do not significantly affect the subsequent use of key concepts and misconceptions. Further granular research could shed light on the effects of situated learning on factors such as transferring knowledge from one situation to another. First, research should investigate whether students understand learning concepts by, for example, investigating students' task solving while learning ('source'). Second, it is crucial to determine whether students understand subsequent reasoning tasks such as post-tests or follow-up tests ('target'). If students understand both issues, knowledge transfer should be possible if other variables do not inhibit it (e.g., negative emotions or stress; Klepsch and Seufert 2020). Third, research should inspect whether the similarity of post-test tasks can explain the lack of effects regarding situated learning.

III. This study is one of the first to use the cognitive load questionnaire by Klepsch and Seufert (2020) to examine cognitive load in a natural classroom rather than a laboratory setting. While using the instrument, we found that students learning in different evolutionary contexts (trait gain vs. trait loss) showed significantly different intrinsic and extraneous cognitive load levels. Students find it cognitively less demanding to learn in contexts of trait gain than trait loss. As the extraneous and intrinsic load increased simultaneously, we recommend that researchers and educators should be aware of underlying contexts in learning instructions and their effects on the students' working memory capacity. It is essential to deliberately choose instructional contexts, material design, and methods to minimize disturbing cognitive load.

IV. In addition, we found no evidence that clarifying misconceptions about natural selection led students to use significantly more key concepts and fewer misconceptions in both trait gain and trait loss scenarios. However, since similar studies could already generate significant learning gains, we recommend conducting further research on situated learning by improving interventions with the generated knowledge and circumventing the presented limitations.

V. We detected that combining the evolutionary context of trait gain and the clarification of misconceptions promises the potential for reducing misconceptions. Students use significantly fewer misconceptions in trait gain scenarios when they receive a clarification than students who did not receive the extra support. We could not find the effect in trait loss contexts. Moreover, we only

found this effect regarding the general use of misconceptions. For now, the clarification of misconceptions does not appear to be a universal solution as an instructional strategy in natural selection. Therefore, gaining new insights into instructions, including a clarification of misconceptions in other evolutionary contexts, especially less cognitively demanding contexts (e.g., natural selection covering animals compared to plants), is a worthwhile direction for future research.

Limitations

The results of our study must be interpreted in light of their limitations. The generalizability is limited to the chosen sample or similar group compositions (Hedges 2013). The results refer primarily to German students at the upper secondary school level, necessitating further research that supports the findings with a perspective on international school students. In terms of the sample, the sample size was generally appropriate for educational research purposes, given that such a number of students can be used to determine the average effects of educational interventions (Colton et al. 2018; Hattie 2008). However, future intervention studies should increase the sample size to provide more statistical power, given the small effects found in this study.

It is advisable to collect information on how many lessons each student received on the basic topics concerning future study designs. This is important because many German states have no fixed timelines for teaching specific topics. The curricula for German schools are primarily organized so that they specify the topics to be taught and formulate clear objectives for teaching units (e.g., MSW NRW 2013b; KMK 2016). Therefore, based on the curricula, we could only ascertain in this study that the pupils must already have had some initial experience with the topic of natural selection. However, we could not determine what proportion of the overall evolution instruction our intervention accounted for.

Furthermore, we recommend researchers attempting to replicate this study to critically evaluate the choice of research design, especially for the pre-test. Our aim with the pre-test was to collect baseline prior knowledge. We conducted the knowledge about evolution with the KAEVO 2.0 to minimize a potential pre-testing effect or pre-test sensitization as best as possible (e.g., Richland et al. 2009; Salkind 2010). Hence, in this study, the intention was to avoid learning and familiarity with the ACORNS in the pre-test and to be able to write the differences in the post-test of the intervention. However, one disadvantage of this design is that it did not allow us to calculate the differences between pre- and post-test

performance. We are aware that this could have undermined our results. Accordingly, to investigate differences in future studies, we suggest employing the same instrument to assess knowledge about evolution in the pre- and post-test. Thereby attention should be paid to possible confounders due to, for example, different surface features of tasks (e.g., Nehm and Ha 2011; Federer et al. 2015) and situatedness (e.g., Kirsh 2009; Reder et al. 1994). Alternatively, the above limitations could be minimized by using the Solomon four group design. This can help to screen the effect from pre-test to post-test (Solomon 1949). Nevertheless, the design has the disadvantage that it requires much effort and would have been disproportionate and uneconomical for the present study with the already existing four groups.

Moreover, according to Klepsch and Seufert (2020), researchers have mainly used the cognitive load questionnaire in studies with systematically varied variables. Consequently, the validity of this instrument was rarely investigated in classroom settings. Since the operationalization of cognitive load is highly complex, divergent results can still occur. For example, in findings where increasing germane load does not inevitably imply an increase in post-test performance or in cases where students cannot clearly distinguish between the intrinsic and extraneous load (Klepsch and Seufert 2020; Klepsch et al. 2017). Therefore, the role of individual cognitive load is still debated (Klepsch and Seufert 2021). Besides, the questionnaire consists of subjective ratings, which can vary among individuals, as everyone has a different memory capacity and perception. We recommend augmenting future studies with an additional objectively scored instrument to protect the cognitive load analysis from confounding subjectivity and strengthen arguments for convergent validity (Maxwell et al. 2017). In this regard, Kalyuga and Plass (2017) present several practical methods for measuring cognitive load (e.g., dual-task measures).

Additionally, all students had equal time to work on the intervention materials. As we aimed to consider the clarification as a complement rather than an alternative to teaching key concepts, we manipulated the factor clarification of misconceptions (no vs. yes) by integrating the clarification of misconceptions into the actual materials of the respective groups. They received four additional tasks. This procedure could have constituted a limitation as the other groups did not receive any extra tasks during the working time. An alternative would have been to provide the students without the clarification with other tasks about natural selection, but this would probably have caused advantageous knowledge about further contexts and key concept use. Off-topic tasks would have resulted in less time to engage with natural selection.

Thus, we anticipate that the alternative would have led to a stronger bias in the results than the current approach, in which we integrated four additional items into the intervention materials for part of the groups.

Lastly, we cannot exclude the possibility that a lack of transfer hindered the effects of situated learning. Future studies should investigate how the ability to transfer contextual knowledge about natural selection can be visualized and promoted in students (Kirsh 2009; Richey and Nokes-Malach 2015; Veenman et al. 2004; Hajian 2019).

Conclusion

Overall, we replicated and corroborated current research on secondary school students' knowledge about evolution. Our descriptive results indicated that students' knowledge about evolution is low and that most of the misconceptions they resort to are teleological in origin. Furthermore, descriptive results showed that students used fewer misconceptions in trait gain scenarios than in trait loss scenarios, both in the pre- and post-test. With regard to our hypotheses, the results did not reveal a significant effect of situated learning on later reasoning about natural selection. Furthermore, the findings revealed that students who learned in the intervention contexts of trait gain perceive lower intrinsic and extraneous load than those who learned in trait loss contexts. Additionally, the clarification of misconceptions showed no benefits when disregarding the instructional contexts. The same is true when considering the interaction effect of the 2×2 factorial design and its effect on key concepts and misconceptions, both in trait gain and trait loss scenarios of the post-test. Nevertheless, when considering the general use of key concepts and misconceptions, learning in trait gain contexts with an additional clarification of misconceptions can lead to significantly fewer misconceptions in later reasoning about natural selection. This effect is especially notable when compared to learning in trait gain contexts without the clarification of misconceptions.

Moreover, our contribution improves understanding of natural selection learning and advises teachers on conductive instructional strategies. Our results recommend that researchers and educators pay attention to the complex interplay of prior situated learning, differences in instructional contexts, effects of an explicit clarifying of misconceptions, and cognitive load. Paying attention can aid in developing instructional strategies with an appropriate design and allotted time. In addition, we endorse considering the situated learning approach as a valuable lens to interpret concept use. We also advocate broadening research to other contexts (e.g., animals and plants) or research in regular school environments that analyses the behavior of cognitive load and achieves learning success while learning with a clarification of misconceptions.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12052-022-00163-5.

> **Additional file 1.** Intervention materials for all four groups (German version). This file contains the intervention material for all four intervention groups of this study.
>
> **Additional file 2.** Key concepts and misconceptions of natural selection. This file provides definitions of the key concepts and misconceptions underlying this study.
>
> **Additional file 3.** Research design and formal model descriptions used to test the hypotheses. This file comprises the research design and formal model descriptions used to test the hypotheses.

Availability of data and materials
As the data were collected on high school students who were minors, the generated and analyzed dataset is not publicly available, but an aggregated dataset is available from the corresponding author JG upon reasonable request.

Declarations

Ethics approval and consent to participate
Throughout the study, we ensured compliance with school law agreements and adhered to ethical principles and guidelines for protecting human research subjects (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1979; Ministry of the Interior of the State of North Rhine-Westphalia, 2020). Participating students provided informed consent at the beginning of the study. Students used a personal four-digit code to mark materials of the three study phases that belong together to grant anonymity. Teachers and students received no reward for their participation.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Present Address: Institute for Biology Education, Faculty of Mathematics and Natural Sciences, University of Cologne, Herbert-Lewin-Straße 10, 50931 Cologne, Germany. [2]Present Address: IPN-Leibniz Institute for Science and Mathematics Education, Kiel, Germany.

References

Anderson JR, Reder LM, Simon HA. Situated learning and education. Educ Res. 1996;25(4):5–11. https://doi.org/10.3102/0013189X025004005.

Andrews TM, Leonard MJ, Colgrove CA, Kalinowski ST. Active learning not associated with student learning in a random sample of college biology courses. CBE Life Sci Educ. 2011;10(4):394–405. https://doi.org/10.1187/cbe.11-07-0061.

Aptyka H, Großschedl J. Fostering students' understanding of natural selection by teaching negative knowledge. In: Proceedings of the 1st International Conference on Innovation in Learning Instruction and Teacher Education (ILITE). Hanoi, Vietnam: University of Education Publishing House; 2019. p. 354–364.

Barsalou LW. Situated conceptualization. In: Cohen H, Lefebvre C, editors. Handbook of categorization in cognitive science. Amsterdam: Elsevier Science Ltd.; 2005. p. 619–50.

Barsalou LW. Situating concepts. In: Robbins P, Aydede M, editors. The Cambridge handbook of situated cognition. 1st ed. Cambridge: Cambridge University Press; 2009. p. 236–63.

Barsalou LW. Situated conceptualization: theory and applications. In: Coello Y, Fischer, MH, editors. Foundations of embodied cognition: perceptual and emotional embodiment. East Sussex, UK: Psychology Press; 2016. p. 11–37.

Bechtel W. Explanation: mechanism, modularity, and situated cognition. In: Robbins P, Aydede M, editors. Cambridge handbook of situated cognition. 1st ed. Cambridge: Cambridge University Press; 2009. p. 155–70.

Beggrow EP, Sbeglia GC. Do disciplinary contexts impact the learning of evolution? Assessing knowledge and misconceptions in anthropology and biology students. Evol Educ Outreach. 2019;12(1):1–17. https://doi.org/10.1186/s12052-018-0094-6.

Beniermann A. Evolution – von Akzeptanz und Zweifeln empirische Studien über Einstellungen zur Evolution und Bewusstsein [Evolution – about acceptance and doubts: empirical studies on attitudes towards evolution and consciousness]. Wiesbaden: Springer Spektrum; 2019. https://doi.org/10.1007/978-3-658-24105-6.

Beniermann, A., Kuschmierz, P., Pinxten, A., Aivelo, T., Bohlin, G. & Brennecke, JS. *et al*. Evolution Education Questionnaire on acceptance and knowledge (EEQ) – standardised and ready-to-use protocols to measure acceptance of evolution and knowledge about evolution in an international context, *Zenodo*, https://doi.org/10.5281/zenodo.4554742 (2021).

Bishop BA, Anderson CW. Student conceptions of natural selection and its role in evolution. J Res Sci Teach. 1990;27(5):415–27. https://doi.org/10.1002/tea.3660270503.

Brown JS, Collins A, Duguid P. Situated cognition and the culture of learning. Educ Res. 1989;18(1):32–42. https://doi.org/10.3102/0013189X018001032.

Carroll SP, Jørgensen PS, Kinnison MT, Bergstrom CT, Denison RF, Gluckman P, et al. Applying evolutionary biology to address global challenges. Science. 2014;346(6207):1245993. https://doi.org/10.1126/science.1245993.

Chandler P, Sweller J. Cognitive load theory and the format of instruction. Cogn Instr. 1991;8(4):293–332. https://doi.org/10.1207/s1532690xci0804_2.

Choi J-I, Hannafin M. Situated cognition and learning environments: roles, structures, and implications for design. Educ Technol Res Dev. 1995;43(2):53–69. https://doi.org/10.1007/BF02300472.

Colton J, Sbeglia G, Finch S, Nehm RH. A quasi-experimental study of short- and long-term learning of evolution in misconception-focused classes. In: Paper presented at the American Educational Research Association International Conference (AERA). New York, NY; 2018.

CONSORT. The CONSORT flow diagram. 2010. http://www.consort-statement.org/consort-statement/flow-diagram. Accessed 07 Apr 2022.

Cooper G. Research into cognitive load theory and instructional design at UNSW. 1998. https://web.archive.org/web/20070830131437/http://education.arts.unsw.edu.au/staff/sweller/clt/index.html. Accessed 07 Apr 2022.

De Jong T. Cognitive load theory, educational research, and instructional design: some food for thought. Instr Sci. 2010;38(2):105–34. https://doi.org/10.1007/s11251-009-9110-0.

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

Eurydice. German Overview: national specificities of the education system. 2021. https://eacea.ec.europa.eu/national-policies/eurydice/content/germany_en. Accessed 07 Apr 2022.

Evans EM. Cognitive and contextual factors in the emergence of diverse belief systems: creation versus evolution. Cogn Psychol. 2001;42(3):217–66. https://doi.org/10.1006/cogp.2001.0749.

Evans EM. Conceptual change and evolutionary biology: a developmental analysis. In: Vosniadou S, editor. International handbook of research on conceptual change. New York: Routledge; 2008. p. 263–94. http://www-personal.umich.edu/~evansem/IHCC-Evans-Chapter.pdf. Accessed 07 Apr 2022.

Federer MR, Nehm RH, Opfer JE, Pearl D. Using a constructed-response instrument to explore the effects of item position and item features on the assessment of students' written scientific explanations. Res Sci Educ. 2015;45(4):527–53. https://doi.org/10.1007/s11165-014-9435-9.

Fenner A. Schülervorstellungen zur Evolutionstheorie, Konzeption und Evaluation von Unterricht zur Anpassung durch Selektion [Student conceptions of evolutionary theory, design and evaluation of lessons on adaptation by selection] [Dissertation]. Gießen, Germany: Justus-Liebig-University; 2013.

Gartmeier M, Bauer J, Gruber H, Heid H. Negative knowledge: understanding professional learning and expertise. Vocat Learn. 2008;1(2):87–103. https://doi.org/10.1007/s12186-008-9006-1.

Goel L, Johnson N, Junglas I, Ives B. Situated learning: conceptualization and measurement. Decis Sci J Innov Educ. 2010;8(1):215–40. https://doi.org/10.1111/j.1540-4609.2009.00252.x.

Gregory TR. Understanding natural selection: essential concepts and common misconceptions. Evol Educ Outreach. 2009;2(2):156–75. https://doi.org/10.1007/s12052-009-0128-1.

Großschedl J, Konnemann C, Basel N. Pre-service biology teachers' acceptance of evolutionary theory and their preference for its teaching. Evol Educ Outreach. 2014;7(1):18. https://doi.org/10.1186/s12052-014-0018-z.

Großschedl J, Seredszus F, Harms U. Angehende Biologielehrkräfte: evolutionsbezogenes Wissen und Akzeptanz der Evolutionstheorie [Pre-service biology teachers: knowledge and acceptance of evolutionary theory]. Zeitschrift für Didaktik der Naturwissenschaften (ZfDN). 2018;24(1):51–70. https://doi.org/10.1007/s40573-018-0072-0.

Ha M, Nehm RH. Darwin's difficulties and students' struggles with trait loss: cognitive-historical parallelisms in evolutionary explanation. Sci Educ. 2014;23(5):1051–74. https://doi.org/10.1007/s11191-013-9626-1.

Ha MS, Lee JK, Cha HY. A cross-sectional study of students' conceptions on evolution and characteristics of concept formation about it in terms of the subjects: human, animals and plants. J Korean Assoc Sci Educ. 2006;26(7):813–25.

Hajian S. Transfer of learning and teaching: a review of transfer theories and effective instructional practices. IAFOR J Educ. 2019;7(1):93–111. https://doi.org/10.22492/ije.7.1.06.

Hattie J. Visible learning: a synthesis of over 800 meta-analyses relating to achievement. 1st ed. London: Routledge; 2008. p. 392. https://doi.org/10.4324/9780203887332.

Hedges VL. Recommendations for practice: justifying claims of generalizability. Educ Psychol Rev. 2013;25(3):331–7. https://doi.org/10.1007/s10648-013-9239-x.

Heemsoth T, Heinze A. Secondary school students learning from reflections on the rationale behind self-made errors: a field experiment. J Exp Educ. 2016;84(1):98–118. https://doi.org/10.1080/00220973.2014.963215.

Hendricks CC. Teaching causal reasoning through cognitive apprenticeship: what are results from situated learning? J Educ Res. 2001;94(5):302–11. https://doi.org/10.1080/00220670109598766.

Johnson-Laird PN. Mental models in cognitive science. Cogn Sci. 1980;4(1):71–115. https://doi.org/10.1016/S0364-0213(81)80005-5.

Kalyuga S, Plass JL. Cognitive load as a local characteristic of cognitive processes implications for measurement approaches. In: Zheng RZ,

Aptyka *et al. Evolution: Education and Outreach*   (2022) 15:5

Page 20 of 21

editor. Cognitive load measurement and application. 1st ed. New York: Routledge; 2017. p. 59–74. https://doi.org/10.4324/9781315296258-5.

Kampourakis K, Zogza V. Preliminary evolutionary explanations: a basic framework for conceptual change and explanatory coherence in evolution. Sci Educ. 2009;18(10):1313–40. https://doi.org/10.1007/s11191-008-9171-5.

Kampourakis K, Palaiokrassa E, Papadopoulou M, Pavlidi V, Argyropoulou M. Children's intuitive teleology: shifting the focus of evolution education research. Evol Educ Outreach. 2012;5(2):279–91. https://doi.org/10.1007/s12052-012-0393-2.

Kirsh D. Problem solving and situated cognition. In: Robbins P, Aydede M, editors. The Cambridge handbook of situated cognition. Cambridge: Cambridge University Press; 2009. p. 264–306. https://doi.org/10.1017/CBO9780511816826.015.

Klahr D, Robinson M. Formal assessment of problem-solving and planning processes in preschool children. Cogn Psychol. 1981;13(1):113–48. https://doi.org/10.1016/0010-0285(81)90006-2.

Klepsch M, Seufert T. Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. Instr Sci. 2020;48(1):45–77. https://doi.org/10.1007/s11251-020-09502-9.

Klepsch M, Seufert T. Making an effort versus experiencing load. Front Educ. 2021. https://doi.org/10.3389/feduc.2021.645284.

Klepsch M, Schmitz F, Seufert T. Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. Front Psychol. 2017;8(NOV):1–18. https://doi.org/10.3389/fpsyg.2017.01997.

Kuschmierz P, Meneganzin A, Pinxten R, Pievani T, Cvetković D, Mavrikaki E, et al. Towards common ground in measuring acceptance of evolution and knowledge about evolution across Europe: a systematic review of the state of research. Evol Educ Outreach. 2020a;13(1):18. https://doi.org/10.1186/s12052-020-00132-w.

Kuschmierz P, Beniermann A, Graf D. Development and evaluation of the knowledge about evolution 2.0 instrument (KAEVO 2.0). Int J Sci Educ. 2020b;42(15):2601–29. https://doi.org/10.1080/09500693.2020.1822561.

Lammert N. Akzeptanz, Vorstellungen und Wissen von Schülerinnen und Schülern der Sekundarstufe I zu Evolution und Wissenschaft [Acceptance, perceptions and knowledge of lower secondary school students on evolution and science] [Dissertation]. Dortmund: Technical University Dortmund; 2012.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74. https://doi.org/10.2307/2529310.

Lave J, Wenger E. Situated learning: legitimate peripheral participation. Cambridge: Cambridge University Press; 1991. p. 1–138.

Leppink J, van den Heuvel A. The evolution of cognitive load theory and its application to medical education. Perspect Med Educ. 2015;4(3):119–27. https://doi.org/10.1007/s40037-015-0192-x.

Limón M. On the cognitive conflict as an instructional strategy for conceptual change: a critical appraisal. Learn Instr. 2001;11(4–5):357–80. https://doi.org/10.1016/S0959-4752(00)00037-2.

Maxwell SE, Delaney HD, Kelley K. Designing experiments and analyzing data. 3rd ed. New York: Routledge; 2017. https://doi.org/10.4324/9781315642956.

Mayer RE. Rote versus meaningful learning. Theory Pract. 2002;41(4):226–32. https://doi.org/10.1207/s15430421tip4104_4.

Ministry for School and Further Education of North Rhine-Westphalia. Kernlehrplan für die Realschule in Nordrhein-Westfalen: Biologie [Core curriculum for the Realschule in North Rhine-Westphalia: biology]. 1st ed. Düsseldorf; 2011a. https://www.schulentwicklung.nrw.de/lehrplaene/upload/klp_SI/RS/Biologie/RS_Biologie_Endfassung.pdf. Accessed 07 Apr 2022.

Ministry for School and Further Education of North Rhine-Westphalia. Kernlehrplan für die Hauptschule in Nordrhein-Westfalen: Lernbereich Naturwissenschaften Biologie, Chemie, Physik [Core curriculum for the Hauptschule in North Rhine-Westphalia: learning area natural sciences biology, chemistry, physics]. 1st ed. Düsseldorf; 2011b. https://www.schulentwicklung.nrw.de/lehrplaene/upload/lehrplaene_download/hauptschule/NW_HS_KLP_Endfassung.pdf. Accessed 07 Apr 2022.

Ministry for School and Further Education of North Rhine-Westphalia. Kernlehrplan für das Abendgymnasium und Kolleg in Nordrhein-Westfalen: Biologie [Core curriculum for the evening high school and college in North Rhine-Westphalia: biology]. Düsseldorf; 2015. https://www.schulentwicklung.nrw.de/lehrplaene/lehrplan/148/KLP_WbK_BI.pdf. Accessed 07 Apr 2022.

Ministry for School and Education of the State of North Rhine-Westphalia. Kernlehrplan für das Gymnasium – Sekundarstufe I in Nordrhein-Westfalen: Biologie [Core curriculum for the Gymnasium – secondary level I in North Rhine-Westphalia: biology]. 1st ed. Düsseldorf; 2019. https://www.schulentwicklung.nrw.de/lehrplaene/lehrplan/197/g9_bi_klp_3413_2019_06_23.pdf. Accessed 07 Apr 2022.

Ministry for School and Further Education of the State of North Rhine-Westphalia. Kernlehrplan für die Sekundarstufe II Gymnasium/Gesamtschule in Nordrhein-Westfalen: Biologie [Core curriculum for the secondary level II gymnasium/comprehensive school in North Rhine-Westphalia: biology]. 1st ed. Düsseldorf; 2013a. https://www.schulentwicklung.nrw.de/lehrplaene/upload/klp_SII/bi/GOSt_Biologie_Endfassung.pdf. Accessed 07 Apr 2022.

Ministry for School and Further Education of the State of North Rhine-Westphalia. Kernlehrplan für die Gesamtschule – Sekundarstufe I in Nordrhein-Westfalen: Naturwissenschaften Biologie, Chemie, Physik [Core curriculum for the comprehensive school – secondary level I in North Rhine-Westphalia: natural sciences biology, chemistry, physics]. Düsseldorf; 2013b. https://www.schulentwicklung.nrw.de/lehrplaene/lehrplan/130/KLP_GE_NW.pdf. Accessed 07 Apr 2022.

Ministry of the Interior of the State of North Rhine-Westphalia. Schulgesetz für das Land Nordrhein-Westfalen (Schulgesetz NRW – SchulG) [School law for the state of North Rhine-Westphalia]. Germany; 2020. https://bass.schul-welt.de/pdf/6043.pdf?20210402121340. Accessed 07 Apr 2022.

Mohajeri K, Mesgari M, Lee AS. When statistical significance is not enough: investigating relevance, practical significance, and statistical significance. MIS Q Manag Inf Syst. 2020;44(2):525–59. https://doi.org/10.25300/MISQ/2020/13932.

Moreno R, Park B. Cognitive load theory: historical development and relation to other theories. In: Plass JL, Moreno R, Brunken R, editors. Cognitive load theory. Cambridge: Cambridge University Press; 2010. p. 9–28. https://doi.org/10.1017/CBO9780511844744.003.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report: ethical principles and guidelines for the protection of human subjects of research. 1979.

National Research Council [NRC]. Thinking evolutionarily: evolution education across the life sciences: summary of a convocation. In: Labov JB, editor. Washington, D.C.: National Academies Press; 2012. https://doi.org/10.17226/13403.

Nehm RH. Evolution. In: Kampourakis K and Reiss MJ, editors. Teaching biology in schools. New York, NY: Routledge; 2018. https://doi.org/10.4324/9781315110158.

Nehm RH, Ha M. Item feature effects in evolution assessment. J Res Sci Teach. 2011;48(3):237–56. https://doi.org/10.1002/tea.20400.

Nehm RH, Reilly L. Biology majors' knowledge and misconceptions of natural selection. Bioscience. 2007;57(3):263–72. https://doi.org/10.1641/B570311.

Nehm RH, Beggrow EP, Opfer JE, Ha M. Reasoning about natural selection: diagnosing contextual competency using the ACORNS instrument. Am Biol Teach. 2012;74(2):92–8. https://doi.org/10.1525/abt.2012.74.2.6.

Nehm RH, Ha M, Rector M, Opfer JE, Perrin L, Ridgway J, et al. Scoring guide for the open response instrument (ORI) and evolutionary gain and loss test (EGALT). In: Technical Report of National Science Foundation REESE Project 0909999; 2010. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.471.7384&rep=rep1&type=pdf. Accessed 07 Apr 2022.

Nelson CE. Teaching evolution (and all of biology) more effectively: strategies for engagement, critical reasoning, and confronting misconceptions. Integr Comp Biol. 2008;48(2):213–25. https://doi.org/10.1093/icb/icn027.

Opfer JE, Nehm RH, Ha M. Cognitive foundations for science assessment design: knowing what students know about evolution. J Res Sci Teach. 2012;49(6):744–77. https://doi.org/10.1002/tea.21028.

Orru G, Longo L. The Evolution of Cognitive load theory and the measurement of its intrinsic, extraneous and germane loads: a review. 2019. p. 23–48. https://doi.org/10.1007/978-3-030-14273-5_3.

Oser FK, Näpflin C, Hofer C, Aerni P. Towards a theory of negative knowledge (NK): almost-mistakes as drivers of episodic memory amplification. 2012. p. 53–70. https://doi.org/10.1007/978-90-481-3941-5_4.

Otto HE, Mandorli F. A framework for negative knowledge to support hybrid geometric modeling education for product engineering. J Comput Des Eng. 2018;5(1):80–93. https://doi.org/10.1016/j.jcde.2017.11.006.

Paas F, Ayres P. Cognitive load theory: a broader view on the role of memory in learning and education. Educ Psychol Rev. 2014;26(2):191–5. https://doi.org/10.1007/s10648-014-9263-5.

Paas F, Renkl A, Sweller J. Cognitive load theory and instructional design: recent developments. Educ Psychol. 2003;38(1):1–4. https://doi.org/10.1207/S15326985EP3801_1.

Park HJ. Components of conceptual ecologies. Res Sci Educ. 2007;37(2):217–37. https://doi.org/10.1007/s11165-006-9023-8.

Poppenk J, Köhler S, Moscovitch M. Revisiting the novelty effect: when familiarity, not novelty, enhances memory. J Exp Psychol Learn Mem Cogn. 2010;36(5):1321–30. https://doi.org/10.1037/a0019900.

Posner GJ, Strike KA, Hewson PW, Gertzog WA. Accommodation of a scientific conception: toward a theory of conceptual change. Sci Educ. 1982;66(2):211–27. https://doi.org/10.1002/sce.3730660207.

Rachmatullah A, Nehm RH, Roshayanti F, Ha M. Evolution education in Indonesia: pre-service biology teachers' knowledge, reasoning models, and acceptance of evolution. In: Deniz H, Borgerding L, editors. Evolution education around the globe. Cham: Springer International Publishing; 2018. p. 335–55. https://doi.org/10.1007/978-3-319-90939-4_18.

Rea-Ramírez MA, Clement JJ. In search of dissonance: the evolution of dissonance in conceptual change theory. In: Proceedings of the National Association for Research in Science Teaching (NARST). San Diego, CA; 1998. p. 1–29.

Rector MA, Nehm RH, Pearl D. Learning the language of evolution: lexical ambiguity and word meaning in student explanations. Res Sci Educ. 2013;43(3):1107–33. https://doi.org/10.1007/s11165-012-9296-z.

Reder L, Klatzky RL, Klatzky RL. The effect of context on training: is learning situated? Pittsburgh, United States; 1994.

Remembering SJ. In: Robbins P, Aydede M, editors. The Cambridge handbook of situated cognition. Cambridge: Cambridge University Press; 2008. p. 217–35. https://doi.org/10.1017/CBO9780511816826.013.

Renkl A. The worked-out examples principle in multimedia learning. In: Mayer RE, editor. The Cambridge handbook of multimedia learning. Cambridge: Cambridge University Press; 2005. p. 229–46. https://doi.org/10.1017/CBO9780511816819.016.

Richey JE, Nokes-Malach TJ. Comparing four instructional techniques for promoting robust knowledge. Educ Psychol Rev. 2015;27(1):181–218. https://doi.org/10.1007/s10648-014-9268-0.

Richland LE, Kornell N, Kao LS. The pretesting effect: do unsuccessful retrieval attempts enhance learning? J Exp Psychol Appl. 2009;15(3):243–57. https://doi.org/10.1037/a0016496.

Sadler TD. Situated learning in science education: socio-scientific issues as contexts for practice. Stud Sci Educ. 2009;45(1):1–42. https://doi.org/10.1080/03057260802681839.

Salkind N. Pretest sensitization. In: Salkind N, editor. Encyclopedia of research design. Thousand Oaks: SAGE Publications, Inc.; 2010. https://doi.org/10.4135/9781412961288.n332.

Schaffernicht M. Detecting and monitoring change in models. Syst Dyn Rev. 2006;22(1):73–88. https://doi.org/10.1002/sdr.329.

Schneeweiß N, Gropengießer H. Organising levels of organisation for biology education: a systematic review of literature. Educ Sci. 2019;9(3):207. https://doi.org/10.3390/educsci9030207.

Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the federal republic of Germany. The education system in the federal republic of Germany 2016/2017. Bonn; 2019. https://www.kmk.org/fileadmin/Dateien/pdf/Eurydice/Bildungswesen-engl-pdfs/dossier_en_ebook.pdf. Accessed 07 Apr 2022.

Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the federal republic of Germany. Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife [Educational standards in biology for the general university entrance qualification]. Berlin: Carl Link; 2020. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020/2020_06_18-BildungsstandardsAHR_Biologie.pdf. Accessed 07 Apr 2022.

Shen Z, Zhang L, Xiao X, Li R, Liang R. Icon familiarity affects the performance of complex cognitive tasks. i-perception. 2020;11(2):2041669520910167. https://doi.org/10.1177/2041669520910167.

Sinatra GM, Kienhues D, Hofer BK. Addressing challenges to public understanding of science: epistemic cognition, motivated reasoning, and conceptual change. Educ Psychol. 2014;49(2):123–38. https://doi.org/10.1080/00461520.2014.916216.

Smith MU. Current status of research in teaching and learning evolution: II. Pedagogical issues. Sci Educ. 2010;19(6–8):539–71. https://doi.org/10.1007/s11191-009-9216-4.

Solomon RL. An extension of control group design. Psychol Bull. 1949;46(2):137–50. https://doi.org/10.1037/h0062958.

Stanovich KE. Implicit philosophies of mind: the dualism scale and its relation to religiosity and belief in extrasensory perception. J Psychol Interdiscip Appl. 1989;123(1):5–23. https://doi.org/10.1080/00223980.1989.10542958.

Sweller J. Cognitive load theory. In: Ross BH, editor. Psychology of learning and motivation—advances in research and theory. Cambridge: Academic Press; 2011. p. 37–76. https://doi.org/10.1016/B978-0-12-387691-1.00002-8.

Sweller J, Chandler P. Why some material is difficult to learn. Cogn Instr. 1994;12(3):185–233. https://doi.org/10.1207/s1532690xci1203_1.

Taber KS. Knowledge, beliefs and pedagogy: how the nature of science should inform the aims of science education (and not just when teaching evolution). Cult Stud Sci Educ. 2017;12(1):81–91. https://doi.org/10.1007/s11422-016-9750-8.

Tulis M, Steuer G, Dresel M. Learning from errors: a model of individual processes. Front Learn Res. 2016;4(4):12–26. https://doi.org/10.14786/flr.v4i2.168.

Veenman MVJ, Wilhelm P, Beishuizen JJ. The relation between intellectual and metacognitive skills from a developmental perspective. Learn Instr. 2004;14(1):89–109. https://doi.org/10.1016/j.learninstruc.2003.10.004.

Publisher's Note