Science Education

WILEY

Science Education

RESEARCH ARTICLE OPEN ACCESS

Threshold Concepts and Concept Networks in Evolution Education: An Experimental Intervention Study

Helena Aptyka[1] | Daniela Fiedler[2] | Jörg Großschedl[1]

[1]Institute for Biology Education, Faculty of Mathematics and Natural Sciences, University of Cologne, Cologne, Germany | [2]Department of Science Education, University of Copenhagen, Copenhagen, Denmark

Correspondence: Jörg Großschedl (j.grossschedl@uni-koeln.de)

Received: 30 July 2024 | **Revised:** 12 February 2025 | **Accepted:** 13 April 2025

Funding: The authors received no specific funding for this work.

Keywords: curriculum | evolution education | intervention study | network analysis | secondary school | student conceptions | threshold concepts

ABSTRACT

This study analyzes the effect of different instructions on threshold concepts within material covering natural selection on students' use of concepts about evolution. Moreover, it examines students' use of concepts as interconnected networks when reasoning about natural selection and analyzes how these concepts relate to each other regarding their connections and centrality within the network. We recruited a sample of $N = 128$ German 10th-grade students to address these aspects and conducted an experimental intervention study with three groups. We assessed the effect of instruction on the threshold concepts of randomness and probability by teaching them either in contexts of biology, the original discipline of mathematics, or not at all, along with input on natural selection. We analyzed the number of students' key concepts, misconceptions, and threshold concepts by applying an inferential statistical and concept network approach. The results reveal that teaching threshold concepts in biological or mathematical contexts leads to significantly higher use of key concepts than in the control group. Furthermore, only teaching them in a biological context resulted in a significantly higher use of threshold concepts than in the control group. The network analyses reveal descriptive differences in the structures, significant correlations of concepts, and most central and influential concepts (i.e., key concepts: selection factors or limited resources, change in the population; misconceptions: teleology/need, anthropomorphic conception; threshold concept: probability). We discuss the valuable potential of explicitly teaching threshold concepts and utilizing network theory to enhance a nuanced understanding of how students organize and apply evolutionary concepts.

1 | Introduction

It has been established that knowledge about evolutionary processes such as natural selection is central to students' fundamental biology education because evolution represents the integral framework underpinning all areas of modern biology. A scientific grasp of evolution not only facilitates cross-disciplinary knowledge but also equips individuals with the ability to comprehend and make well-informed decisions regarding science-related societal concerns, such as the conservation of the environment (Barthlott et al. 2016; Carroll et al. 2014; National Research Council 2012; Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany 2020).

Despite its importance, many students lack scientific knowledge about evolution (e.g., Aptyka et al. 2022; Beniermann 2019; Kampourakis and Zogza 2009; Kelemen 2011; Kuschmierz, Meneganzin et al. 2020; Smith 2010; Southerland et al. 2001; To et al. 2017). This knowledge shortfall may stem from difficulties in understanding overarching, abstract concepts, known as threshold concepts. These concepts are fundamental to comprehending evolution (e.g., Fiedler et al. 2019; Fiedler et al. 2018; Fiedler et al. 2017; Göransson et al. 2020). However, they are not always effectively integrated into curricula or applied through instruction (Solís-Pinilla and Merino Rubilar 2024).

To address these educational challenges, researchers in biology education have emphasized the importance of threshold concepts. They highlight teaching these concepts to support students in comprehending scientific key concepts related to evolution (Batzli et al. 2016; Fiedler et al. 2019; Fiedler et al. 2017; Tibell and Harms 2017). Threshold concepts are currently rather intangibly and metaphorically described (Salwén 2021) as "portals" or liminal cognitive spaces crucial for opening new, more abstract perspectives and broadening the understanding of previously inaccessible concepts. Due to its peculiar and abstract nature, this type of concept is currently assumed by researchers to possess the unique characteristic of permanently altering individuals' perceptions once it is fully comprehended (Land et al. 2016; Mead and Scott 2010; Meyer and Land 2003, 2006b; Meyer et al. 2010; Tibell and Harms 2017). Understanding threshold concepts is positively associated with using key concepts and negatively associated with using misconceptions when reasoning about evolution. Moreover, it was found that not only misconceptions of evolution but also a lack of comprehension of threshold concepts can complicate acquiring scientific knowledge about evolution and biology, as it can perform as a subjective filter and thus impair the understanding of new information. It is also considered that a lack of understanding of threshold concepts potentially leads to misconceptions (Bressan 2002; Buiatti and Longo 2013; Garvin-Doxas and Klymkowsky 2008; Göransson et al. 2020; Green et al. 2017; Mead and Scott 2010; Tibell and Harms 2017; Wagner 2012). However, the limitations of threshold concept theory lie in the lack of empirical evidence for the explanatory function and effect of threshold concept mediation on concept change and cognitive architecture change (Salwén 2021). Therefore, approximations are needed to provide tangible empirical evidence to visualize this effect.

For the development of instructional interventions in evolution education (e.g., Aptyka et al. 2022; Aptyka and Großschedl 2019; Kampourakis and Zogza 2009), researchers and educators commonly relied on the conceptual change theory and its various terminologies, perspectives, interpretations, and operationalizations (Posner and Strike 1992; Posner et al. 1982; Strike and Posner 1982; Vosniadou and Ioannides 1998). Generally speaking, the conceptual change theory describes, explains, and enables predicting a change process in individuals' conceptual knowledge. However, the change process should not be understood in the narrower sense as a linear process of mere exchange of concepts, as this view is now considered inappropriate (Hartelt and Martens 2024). This process should rather be understood as a developmental process of growth and reconstruction of conceptual knowledge. By applying this theory as a basis for instruction, educators can facilitate the gradual enhancement of students' knowledge about evolution and the use of scientific concepts while reducing the use of misconceptions.

Instructional approaches operationalizing the theory have shown diverging results. Some studies on university or college students provide clear evidence of a significant increase in knowledge (Andrews et al. 2011; Nehm et al. 2022), abstract reasoning, and a decrease in the use of misconceptions (Nehm et al. 2022). Other studies disclose a less successful reduction of the use of misconceptions with no (Aptyka and Großschedl 2019) or diverging effects of instruction (Aptyka et al. 2022) or a significant number of misconceptions remaining for secondary school (e.g., Aptyka et al. 2022; Aptyka and Großschedl 2019; Kampourakis and Zogza 2009) and university students (e.g., Nehm and Reilly 2007). Even though the methodological approach and study design vary between the studies, the general picture is that the effect of instruction is ambiguous, and further research on effective instruction is required (Harms and Reiss 2019).

While acknowledging traditional ways of teaching about key concepts and addressing misconceptions, our objective is to additionally harness the potential of threshold concepts (Batzli et al. 2016; Fiedler et al. 2019; Fiedler et al. 2018; Fiedler et al. 2017; Göransson et al. 2020; Meyer and Land 2005; Nehm and Reilly 2007; Solís-Pinilla and Merino Rubilar 2024; Tibell and Harms 2017). Our first aim is to use the threshold concepts of randomness and probability as an example for such instructional materials. Therefore, we analyzed students' knowledge about evolution (in forms of their concept used in reasoning) after incorporating threshold concepts in instructional material focusing on evolution by natural selection. Hence, the following research question (RQ) guided our research:

RQ1 *How do groups of students differ in their use of key concepts, misconceptions, and threshold concepts after receiving different instruction on threshold concepts within material covering natural selection?*

Furthermore, previous research suggests that measuring and analyzing the number of concepts used (separated by concept types: key concepts and misconceptions) may only superficially capture students' conceptions of evolution when seeking to understand their complex compositions (e.g., National Research Council 2001, 2012; Nehm 2018). This type of research has provided invaluable insights into the general concept use (e.g., frequency, diversity, or consistency) across different items or samples, the context-sensitive use of concepts, significant group differences, and effect sizes of treatments (e.g., Aptyka et al. 2022; Aptyka and Großschedl 2019; Bishop and Anderson 1990; Colton et al. 2018; Großschedl et al. 2018; Gutowski et al. 2023; Nehm and Ha 2011; Nehm and Reilly 2007; Nehm and Schonfeld 2008; Opfer et al. 2012; Rachmatullah et al. 2018; Rector et al. 2013). Previous assessments also analyzed (partial) correlations between concepts, which allowed worthwhile conclusions to be drawn about the relational use of concepts (while ruling out the effects of other concepts on this bivariate relationship; Nehm and Ha 2011; Opfer et al. 2012). However, these approaches to analyzing conceptions about evolution encounter limitations as only minimal information can be derived that clarifies parameters specific to students' net-like knowledge structure or cognitive architecture. These structures refer to the interconnected, web-like organization of students' knowledge, where understanding is not linear or isolated but instead consists of multiple interrelated ideas and concepts forming a complex network (Koponen and Nousiainen 2014; Siew et al. 2019).

Interdisciplinary research suggests that knowledge is arranged in a cognitive network with concepts that interact in complex ways. Therefore, it is recommended to treat knowledge as an entangled network (Amin et al. 2014; BonJour 1985; diSessa and Sherin 1998; Koponen and Nousiainen 2014; Papo et al. 2014; Siew et al. 2019; Thagard 1992). It is imperative to rethink current ways of knowledge assessment to draw valid conclusions and achieve a more nuanced picture of students' conceptions of evolution (National Research Council 2001, 2012). One option is to expand previous research by using the approach of sophisticated network analysis. Analyzing the conceptions of evolution through networks can provide a new dimension and a more detailed layer of interpretability and understanding of students' interconnected conceptions. Inferences can be drawn about the global network characteristics or structures that relate to the network-like structure of students' conceptual knowledge about evolution. These inferences are not limited to statements about the connection between two concepts of evolution but extend previous findings with inferences about relationships between multiple concepts. For example, conclusions could be drawn about the number and strength of clusters of concepts about evolution. Furthermore, these analyses provide computational evidence for the centrality of concepts in the network (i.e., their influence or crucial key role in students' conceptions that connect other concepts) and the strength or weights of relationships between concepts (e.g., deeply connected concepts in students conceptions; Koponen and Nousiainen 2014; Siew et al. 2019). Accordingly, our second aim is to advance valuable prior research by examining students' use of concepts when reasoning about natural selection as interconnected components of concept networks. We posit the following research question:

RQ2 *How do students in different intervention groups form concept networks when reasoning about natural selection, and how do these concepts relate to each other regarding their interconnections and centrality within the network?*

2 | Theoretical Background

2.1 | Threshold Concepts in the Context of Evolution Education

The fields of education (Meyer and Land 2005; Perkins 2008), educational psychology (Entwistle 2008), natural sciences (Green et al. 2017), and recently also biology education have emphasized integrating threshold concepts into traditional teaching methods. Threshold concepts can be presented as a helpful substitute or supplement in dealing with key concepts and misconceptions, as they have the power to alter the understanding and perspective of a phenomenon (Batzli et al. 2016; Fiedler et al. 2019; Land et al. 2016; Tibell and Harms 2017). They act as conceptual gateways, allowing the adoption of a previously inaccessible alternative and unconventional perspective within a liminal space of uncertainty (Mead and Scott 2010; Meyer and Land 2003, 2005, 2006b). Additionally, they promote deeper learning and the development of higher-order cognitive skills (Entwistle 2008; Green et al. 2017; Perkins 2008). For example, threshold concepts such as randomness, probability, temporal scales, and spatial scales are closely associated with the context of evolution (Fiedler et al. 2019; Fiedler et al. 2018; Fiedler et al. 2017; Garvin-Doxas and Klymkowsky 2008; Göransson et al. 2020; Ross et al. 2010; Walck-Shannon et al. 2019). Especially, the correct comprehension of randomness and probability was argued to represent a vital basis for grasping natural selection (Fiedler et al. 2019). A case in point is that holding knowledge about the threshold concept of randomness can help students understand that random mutations on the gene level are one cause of variation. Knowledge about the threshold concept of probability can support students in comprehending the key concept of survival potential, meaning that better-adopted individuals have a higher chance of survival (Tibell and Harms 2017). Accordingly, threshold concepts can positively affect the use of key concepts (Fiedler et al. 2019; Kampourakis and Zogza 2009; Ross et al. 2010; Taylor 2006).

Furthermore, these threshold concepts provide a counterpoint to misconceptions about evolution, such as teleology/need, and may assist in identifying and negotiating misconceptions. To illustrate, evolution is regularly misinterpreted as an efficient goal- or purpose-oriented process (Kampourakis and Zogza 2009). Many individuals struggle to accept that complex structures and functions, such as the eyes of animals, have emerged through random and partly probable processes during evolution (Schoenemann and Clarkson 2023). However, if individuals understand that complex structures and functions result from collective random (e.g., mutations) and probable events and processes (e.g., differential survival or reproductive success), the use of misconceptions such as teleological ones could be dismantled. Accordingly, knowledge about the two threshold concepts can negatively relate to the use of misconceptions (Göransson et al. 2020; Kampourakis and Zogza 2009; Ross et al. 2010; Taylor 2006).

Notably, students can only benefit from threshold concepts when they fully understand them in context and effectively manage the so-called transformative threshold crossing process (Meyer and Land 2003; Walck-Shannon et al. 2019). This process unfolds a "previously inaccessible way of thinking about, understanding, interpreting, or viewing something without which the learner cannot progress" (Meyer and Land 2003, 3). The status quo on understanding such concepts reveals that many students struggle to fully grasp threshold concepts, such as randomness and probability, in mathematics and biology (Fiedler et al. 2019; Gregory 2009). Consequently, students also fail to adopt them when reasoning about natural selection (Fiedler et al. 2017). The struggle to grasp randomness and probability is especially problematic because these concepts originate from mathematics, and teaching them is traditionally only mandated in the mathematics curriculum (Common Core State Standards Initiative 2010; Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany 2022; Texas Education Agency 2023). Only rare positive examples recommend explicit interdisciplinary thematization (National Research Council 2013). Consequently, students' knowledge about threshold concepts may depict rather implicit, unconscious knowledge or abstract rules, respectively, tacit knowledge (Brock 2017) in disciplines other than mathematics, including biology. Furthermore, recent research suggests that knowledge about randomness and probability in mathematical and biological contexts constitutes two distinct knowledge domains that should enable a deeper scientific understanding of evolution (Fiedler et al. 2023; Tibell and Harms 2017). Therefore, the lack of explicit instruction on threshold concepts in biological contexts could leave students uncertain about the relevance of these concepts to biological contexts (Fiedler et al. 2017).

2.2 | Conceptual Change and Concept Networks of Evolution by Natural Selection

2.2.1 | The Consensus of Diverse Perspectives on Conceptual Change Theory

The research consensus on the conceptual change theory posits that knowledge is acquired in all kinds of everyday situations. This knowledge influences the acquisition of advanced knowledge. Hence, if prior knowledge of a context is relatively low, faulty, or incongruent with new information, conceptual change can be time-consuming and effortful. Concurrently, it is assumed that if a person already has either extensive tacit or explicit content-specific knowledge that is compatible with new information, this new information can be learned comparatively quickly (Özdemir and Clark 2007; Posner and Strike 1992; Posner et al. 1982; Strike and Posner 1982). This assumption also supports the positive effects of the tacit knowledge of threshold concepts, enabling a better understanding of key concepts (Tibell and Harms 2017).

2.2.2 | Debated Perspectives on Cognitive Processes Underlying Conceptual Change

Apart from the consensus on the conceptual change theory, researchers continue to debate some unresolved aspects of the cognitive processes underlying conceptual change, such as knowledge structure. The ambiguity surrounding the knowledge structure has given rise to delineating two predominant perspectives on the theoretical framework of conceptual change. These are evident in the knowledge-in-theory and knowledge-in-pieces perspectives (Özdemir and Clark 2007). The knowledge-in-theory perspective considers naïve knowledge to be developed in a coherent, highly structured, theory-like manner by processes such as a revolutionary replacement (e.g., Carey 1999; Gopnik and Wellman 1994; Ionnides and Vosniadou 2002; Wellman and Gelman 1992). The knowledge-in-pieces perspective describes that knowledge consists of quasi-independent fragmented elements of a connected conceptual network (e.g., diSessa 1993; diSessa et al. 2004; Harrison et al. 1999; Özdemir and Clark 2007).

The knowledge-in-pieces approach towards the knowledge structure in the conceptual change process aligns more closely with the network theory. The knowledge-in-pieces perspective proposes that conceptual change occurs through gradually revising, restructuring, refining, and reprioritizing (competing) concepts in a more extensive network. Throughout learning, individuals regularly face several stages of fragmented and transitional knowledge that may consist partly of coexisting conflicting concepts. Additionally, it has been established that specifically novice learners activate, use, and connect these knowledge elements context-sensitively (e.g., diSessa and Sherin 1998; Özdemir and Clark 2007). With extensive knowledge and higher expertise, individuals apply concepts more consistently with systematicity. Most potentially, they form an increasingly stable scientific knowledge that is more abstract and irrespective of the underlying context (e.g., Clark 2006; diSessa 1993; diSessa et al. 2004; Harrison et al. 1999; Özdemir and Clark 2007). In addition, findings from the educational research discourses have identified several key insights: the coexistence and interplay of contradictory key concepts and misconceptions (Aptyka et al. 2022; Opfer et al. 2012), the context-sensitive use of concepts (e.g., Nehm and Ha 2011; Opfer et al. 2012), the relational utilization of concepts (e.g., Nehm and Ha 2011; Opfer et al. 2012), and knowledge being cognitively structured in the form of a network (Amin et al. 2014; BonJour 1985; diSessa and Sherin 1998; Koponen and Nousiainen 2014; Papo et al. 2014; Siew et al. 2019; Thagard 1992). These findings are more congruent with the knowledge-in-pieces perspective.

Findings from the field of network science can complement the perspectives on knowledge structures within the conceptual change theory with more evidence. Research from this field outlines that the process of storing knowledge or concepts is complex and dynamic, with each person developing a unique information network through gradual cognitive processing (Amin et al. 2014; Boshuizen et al. 2020; diSessa and Sherin 1998; Koponen and Nousiainen 2014; Thagard 1992). Furthermore, it is stated that when several concepts are consistently used together, these intertwined connections demonstrate higher cognitive density and clustering of knowledge about a topic (Boshuizen et al. 2020; Koponen and Pehkonen 2010). Moreover, it was detected that if students use certain concepts frequently, this can lead to increased use of similar concepts, particularly semantically closely related ones (Anderson 1983; Collins and Loftus 1975). Thus, changing one concept can affect others because concepts in a network are intertwined (Koponen and Nousiainen 2014; Siew et al. 2019).

Insights into these knowledge structures also exist in research on evolution education. Research has shown that conceptual networks must contain at least three core concepts to enable scientific reasoning about natural selection in various contexts: variation (cause and presence of individual differences), heritability (inheritance of genetic variability), and individual fitness (e.g., survival potential; Federer et al. 2015; Nehm et al. 2012; Nehm and Ha 2011; Nehm et al. 2010; Nehm and Reilly 2007; Opfer et al. 2012). Moreover, initial exemplary studies analyzed the complexity of American university students' conceptions of evolution by employing net-like (partial) correlations of concepts. They revealed that each key concept tends to correlate positively with using other key concepts and negatively with misconceptions. In addition, using certain misconceptions correlated positively with other misconceptions (Nehm and Ha 2011; Opfer et al. 2012).

Given the complexity of conceptual change in light of concept networks, we propose that treating conceptions as concept networks can broaden our understanding of students' concept use and different structures of conceptions. We argue that the concept network approach provides a reasonable basis for the further development of evolution education.

3 | Methods

3.1 | Sample

We recruited 10th-grade level students from eight secondary schools in Germany, of which $N = 128$ participated throughout the whole study (G8 Gymnasium [8-year grammar school system]; equivalent to the third level of the International Standard Classification of Education [ISCED 3]; Eurydice 2021). We chose this sample because evolution is relevant to the students at this stage of their educational paths (Ministry for Schools and Further Education of the State of North Rhine-Westphalia 2022). The students were $M = 15.6$ years old ($SD = 0.8$), and the majority were female (65.1% female, 33.3% male, 1.6% nonbinary). Regarding their religious or denominational group, the sample was diverse, with the majority being Catholics (50.8%), followed by Muslims (15.9%), atheists (9.5%), Protestants (8.7%), members of the Christian Church (5.6%), members of the Orthodox Church (4.0%), and Hindus (0.8%). About 0.8% belonged to other denominations, and 4% of the participants refrained from answering this item. Students assessed their knowledge about evolution as moderate ($M = 3.3$, $SD = 1.4$) and reported a relatively high interest in biology ($M = 4.8$, $SD = 1.5$), as well as a moderate interest in mathematics ($M = 3.9$, $SD = 1.9$). They also showed good grades for biology ($M = 2.3$, $SD = 0.8$) and mathematics ($M = 2.6$, $SD = 1.0$) on a scale from 1 (*best possible grade*) to 6 (*worst possible grade*).

3.2 | Research Design and Procedures

The present study employed an experimental intervention design with two experimental groups and one control. The study was conducted in the students' regular environments to ensure ecological validity. However, the teachers were not involved in the instruction to minimize their influences on the study. We standardized the procedure by training the study instructors, providing informed consent forms to the students, and allocating equal time frames for completing the paper-and-pencil-based materials. The study included a pretest, an intervention, and a posttest, and it was carried out through three school visits, one visit per phase of the study (see Figure 1).

3.2.1 | Pretest

During the pretest, students had 45 min to answer questions related to their knowledge about evolution using the KAEVO 2.0 (part A and B) instrument (Beniermann 2019; Beniermann et al. 2021; Kuschmierz, Beniermann et al. 2020). Furthermore, to contextualize the overall results of this study, we assessed students' acceptance of evolution using the GAENE 2.1 instrument (Smith et al. 2016) and their personal religious faith using the PERF 2.0 (Beniermann 2019).

3.2.2 | Intervention

The intervention (including the opening and closure of the lesson) lasted 90 min and comprised four parts. All students were allocated equal time to engage with these parts.

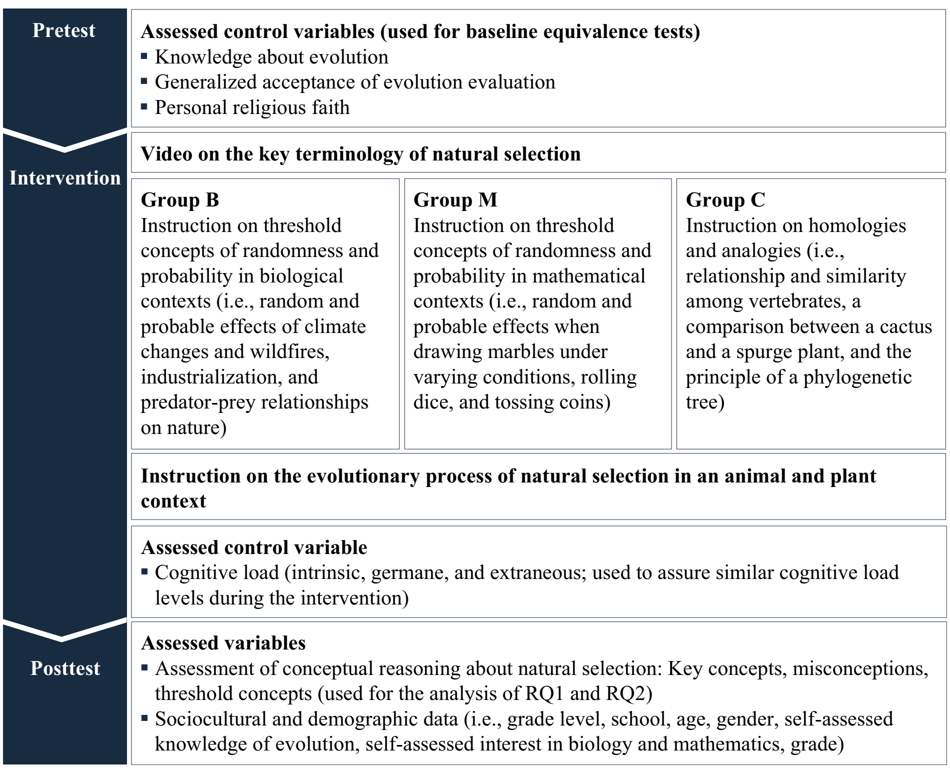


FIGURE 1 | Study design of the experimental intervention study.

In the *first part* (approximately 5 min), a 3-min video on the key terminology of natural selection (e.g., population, individuum, traits, generation, natural selection, the key concept, and misconception) was provided frontally to the students in each class and the students noted definitions. This step aimed to reduce the confounding effects of terminology barriers (Green et al. 2017).

The *second part* (approximately 35 min) involved randomly allocating the students to three intervention groups and distributing the materials. This second part of the intervention materials was manipulated for the three intervention groups and is therefore accountable for group effects.

The first group (B, $n = 45$) worked with materials addressing the threshold concepts of randomness and probability in biological contexts. For example, the students were initially provided with biologists' definitions of randomness and probability (Göransson et al. 2020; Mead and Scott 2010). Subsequently, they received information texts detailing the effects of climate change, wildfires, industrialization, and predator-prey relationships on nature. After that, the students were asked to apply their knowledge about randomness and probability to explain the extent to which these effects on nature could be considered random or probable (Esders et al. 2015).

The second group (M, $n = 43$) engaged with materials designed to address the two threshold concepts within mathematical contexts. For instance, students were initially presented with mathematicians' definitions of randomness and probability (Bennett 2011; Feller 1968). Subsequently, they were provided with information texts describing mathematical experiments, such as drawing marbles under varying conditions, rolling dice, and tossing coins. Finally, the students were required to apply their knowledge of randomness and probability to explain the outcomes of the experiments (based on Schöps 2021).

The third group (C, $n = 40$) served as the control group and received materials unrelated to natural selection that did not thematize randomness and probability. Instead, this group was provided with definitions of homologies and analogies. Following this, they read information texts on topics such as the relationship and similarity among vertebrates, a comparison between a cactus and a spurge plant, and the principle of a phylogenetic tree. After that, they had to apply their knowledge of homologies and analogies to determine whether specific traits of plants and animals were analogous or homologous (based on Becker et al. 2019; Neubrand 2017).

Overall, we applied the methodological approach of self-explanations for this tripartite section of the intervention to involve learners in independently creating explanations about the group-specific contexts. This approach is based on the idea that actively creating explanations deepens understanding and improves retention of information. Also, it can support the coherent linking of knowledge and recognize inaccurate ideas. These approaches were deliberately chosen, as they find support in extant research (Paas et al. 2003; Richey and Nokes-Malach 2015).

The *third part* of the intervention (approximately 40 min) was similar to the one of Aptyka et al. (2022). It focused on the evolutionary process of natural selection in an animal and plant context and was identical for all groups. To begin, we employed worked examples that comprised writing tasks and best-practice solutions to illustrate how experts approach and solve the tasks in group-specific contexts, giving learners a model to reproduce. These examples reduced the extraneous cognitive load during learning and promoted the recognition of one's knowledge gaps (Aptyka et al. 2022; Leppink and van den Heuvel 2015; Richey and Nokes-Malach 2015). Thematically, these worked examples were concerned with the evolutionary gain of a trait in an animal context and the loss of a trait in a plant context, emphasizing the key concepts of variation (cause and presence of variation), heritability of variation, selection pressure or limited resources, and survival potential. Next, examples of common misconceptions were presented, including anthropomorphic conception (intentional change and active adaptation), teleology/need, and use and disuse. Students were encouraged to identify and explain why these misconceptions were incorrect. Finally, we provided an informational text that corrected the misconceptions and clarified why they were inaccurate.

Immediately afterward, in the *fourth part*, we used the cognitive load questionnaire by Klepsch et al. (2017) to check whether the level of cognitive load induced by the intervention materials was comparable between the three groups.

3.2.3 | Posttest

In the posttest, we assessed students' reasoning about natural selection using two open-ended ACORNS items (Nehm et al. 2012; Opfer et al. 2012) and collected information on the students' sociocultural and demographic variables such as grade level, school, age, gender, denominational group, self-assessed knowledge of evolution, self-assessed interest in biology and mathematics, and grade (see Table 1; Nehm et al. 2012).

3.3 | Instruments

We used a rigorous approach to select and apply only validated measures (see Table 2). We aimed to increase the reliability and validity of our results and provide robust evidence.

3.3.1 | Knowledge About Evolution

We chose the KAEVO 2.0 (part A and B) instrument (Beniermann 2019; Beniermann et al. 2021; Kuschmierz, Beniermann et al. 2020) at the time of pretesting, as it enables an economical query of prior knowledge about evolution and minimize the effects of pretest learning. Another benefit is that the surveyed knowledge aligns closely with the posttest assessment of contextual reasoning about natural selection. Both are designed to evaluate understanding of natural selection concepts (e.g., variation) through scenarios involving animals and plants, focusing on the gain or loss of traits (Nehm et al. 2012). For instance, the items survey students' knowledge about how a specific trait of a plant or animal evolved, presenting multiple single-choice options that include one key concept alongside several misconceptions, requiring students to identify the correct answer (Kuschmierz, Beniermann

TABLE 1 | Description of the key concepts, misconceptions, and threshold concepts scored in this study.

Concept	Description
Key concepts	
K1: Causes of variation[a]	Genetic predisposition or changes as a cause of variation (i.e., mutation and recombination)
K2: Presence of variation[a]	Individuals are unique (i.e., both in genotype and phenotype)
K3: Survival potential[a]	Individuals' different potential or chance of survival
K4: Selection factors or limited resources[a]	Abiotic and biotic selection factors indirectly affect the change in living beings
K5: Different survival of individuals	Actual survival of individuals and the resulting opportunity to reproduce
K6: Changes in the distribution of variation	Change in the frequencies of different genotypes in a population over time
K7: Heritability of variation[a]	Inheritance of traits and genetic variability from one generation to its offspring
K8: Differential reproductive success	High reproductive success in "well-adapted" individuals and low reproductive success in "mal-adapted" individuals
Misconceptions	
M1: Teleology/need	Goal or need as the driving force for evolutionary change
M2: Use and disuse[a]	The Lamarckian idea that the use or disuse of traits causes automatic change or the emergence of new traits
M3: Anthropomorphic conception[a]	Attribution of human characteristics to the objects of natural selection (individuals) or to the process itself
M4: Typological conception[a]	A species is understood as a unit of living beings that all have a set of fixed traits
M5: Gentle inheritance	Traits that living beings have learned or acquired during their own life can be directly passed on to their offspring
M6: Event vs. process	Natural selection is understood as an event and not as a permanent process
M7: Origin vs. sorting out variation	No distinction is made between the origin and the sorting out of variation by natural selection, which partly leads to the assumption that variation arises from the environment or natural selection
Threshold concepts	
T1: Randomness[b]	An event or a process is random, uncommon, and unpredictable
T2: Probability[b]	An event or a process is, to some degree, predictable
T3: Temporal scales	Time indicators are mentioned (e.g., millions of years, generations, years, days)
T4: Spatial scales	Different levels are mentioned (e.g., species, populations, individuals, genes)

Note: Concepts and descriptions are based on previous work (Aptyka et al. 2022; Göransson et al. 2020; Großschedl et al. 2018; Moharreri et al. 2014; Nehm et al. 2010).
[a]These concepts were part of all intervention groups.
[b]These concepts were only addressed in the groups B and M.

et al. 2020). Another argument for utilizing this instrument is that it has undergone validation for content validity and internal structure (Gutowski et al. 2023; Kuschmierz, Beniermann et al. 2020).

3.3.2 | Generalized Acceptance of Evolution Evaluation

We assess the generalized acceptance of evolution evaluation using measure GAENE 2.1, including items such as: "Evolutionary theory applies to all plants and animals, including humans." (Smith et al. 2016). The variable generalized acceptance of evolution served as a control variable for the baseline equivalence test, as it is positively associated with knowledge about evolution (Aptyka et al. 2022; Fiedler et al. 2023; Kuschmierz, Meneganzin et al. 2020). This instrument was subject to an extensive validation process. The validation included: involving highly qualified expert reviews, applying the measure to different target groups (Smith et al. 2016), confirming the statistical one-dimensionality (Metzger et al. 2018; Sbeglia and Nehm 2018; Smith et al. 2016), testing the appropriateness using the criteria of Rasch (Manikas et al. 2023; Rachmatullah et al. 2018; Sbeglia and Nehm 2018), showing convergent validity with similar measures, and detecting acceptable item and person reliabilities (Metzger et al. 2018; Sbeglia and Nehm 2018).

3.3.3 | Personal Religious Faith

The PERF 2.0 scale was designed to evaluate the extent of students' personal religious faith (e.g., monotheistic beliefs and

TABLE 2 | Overview of the measures used in this study.

Measure	Description	Sample	Items	*M* (*SD*)	α/ω[b]
KAEVO 2.0 (part A and B) (Kuschmierz, Beniermann et al. 2020)	Knowledge about evolution	Secondary school students (grades 10, 11, 12; biology as a major or minor)	22/24 items: 11 items (part A), 11 items (part B), multiple-choice, true/false statements	 3.78 (2.02) 7.34 (2.37)	 0.55/0.58 0.63/0.65
GAENE 2.1 (Smith et al. 2016)	Generalized acceptance of evolution evaluation	University and secondary school students	13 Items, 5-point Likert scale[a]	3.75 (0.55)	0.76/0.76
PERF 2.0 (Beniermann 2019)	Personal religious faith	heterogeneous and multicultural samples (from 7th grade)	Ten Items, 5-point Likert scale[a]	2.69 (1.31)	0.96/0.96
Cognitive load questionnaire (Klepsch et al. 2017)	Intrinsic, germane, and extraneous cognitive load (ICL, GCL, ECL)	University students	Two items ICL, two items GCL, three items ECL, 5-point Likert scale[a]	4.10 (1.16) 4.30 (1.45) 4.84 (1.55)	0.69/0.69 0.76/0.76 0.74/0.75
ACORNS (Nehm et al. 2012)	Assessment of conceptual reasoning about natural selection; coded by the refined coding manual of Göransson et al. (2020)	University and secondary school students	Two open-response items; coded: Eight key concepts, seven misconceptions, four threshold concepts	 3.91 (3.59) 1.98 (2.33) 1.22 (1.40)	 0.93[c] 0.86[c] 0.93[c]

[a]For the items rated on a Likert scale; a high score indicates a strong expression of the characteristic in question.
[b]We used the internal consistency indices of Cronbach´s alpha (α) and McDonald's omega (ω) to evaluate the reliability of the measures. Cronbach´s α has the advantage of being extensively used, allowing us to compare our results to the ones of earlier studies. McDonald's ω has the advantage of relying on fewer strict assumptions (e.g., tau-equivalence) and more realistic ones about the items of a construct (e.g., the constant variance of true values and tolerated change of error variances; Hayes and Coutts 2020; Ravinder and Saraswathi 2020).
[c]As indices for the reliability of the coded ACORNS items, we used the interrater reliability calculated by an unweighted Cohen´s κ (Cohen 1960; Landis and Koch 1977).

associated behaviors) and included items such as: "I feel most fulfilled when I am in a close connection with God." (Beniermann 2019). Assessing this variable as a control variable can be valuable, as previous studies suggest mixed results regarding its association with knowledge about evolution, indicating either no or a small negative relationship (Aptyka et al. 2022; Deniz et al. 2008; Fiedler et al. 2023; Göransson et al. 2020; Kuschmierz, Meneganzin et al. 2020; Park 2007). Experts from disciplines such as theology, philosophy, sociology, and psychology reviewed the instrument to ensure its content validity (Aptyka et al. 2022; Beniermann 2019).

3.3.4 | Intrinsic, Germane, Extraneous Cognitive Load

The intervention-induced cognitive load was measured using the cognitive load questionnaire and items such as "For this task, I had to think intensively what things meant." (Klepsch et al. 2017; Klepsch and Seufert 2020). The tripartite scale allows analyzing intrinsic (i.e., mental effort induced by the complexity of the material), germane (i.e., cognitive load devoted to meaningful learning in forms of constructing and organizing knowledge), and extraneous (i.e., demanded cognitive resources through poorly designed instructional materials and distractors) cognitive load separately. The instrument was validated concerning substantiative and content validity, and the internal structure was evidenced by expert ratings, confirmatory factor analysis, and acceptable reliability scores in previous studies (Klepsch et al. 2017; Klepsch and Seufert 2020).

3.3.5 | Assessment of Conceptual Reasoning About Natural Selection

We chose the ACORNS instrument as it is regularly used to gain insights into students' reasoning about natural selection and it was validated via convergent validity. This instrument provides a repertoire of different isomorphic items that maintain the same structure while varying the surface features. Specifically, the two chosen tasks differed regarding the taxon (animal vs. plant) and the polarity of trait change (trait gain vs. trait loss). Each task was framed with the prompt: "How would biologists explain how a living (taxon) species with [/lacking] (trait) evolved from an ancestral (taxon) species that lacked [/had] (trait)?" (Nehm et al. 2012). Furthermore, due to the open response format, students could generate individual responses, which could be coded concerning the conceptualization of evolution afterward. Hence, to evaluate the answers to the two open-response items, we used the refined coding manual of Göransson et al. (2020), which ultimately determined the selection of the coded concepts. This coding manual was used to code key concepts, misconceptions, and threshold concepts. To ensure the reliability of the codings (Cohen 1960), we employed two independent raters who analyzed and quantified the use of key concepts, misconceptions, and threshold concepts. An overview of the concepts and descriptions is provided in Table 1. After coding these concepts, we calculated the interrater reliability and obtained an almost perfect interrater agreement of (unweighted) Cohen's Kappa, $\kappa > 0.80$ (Cohen 1960; Landis and Koch 1977). To underpin the validity of the new coding manual, we additionally coded four of the concepts with the established scoring guide for open response instrument (ORI) and evolutionary gain and loss (EGALT) by Nehm et al. (2010). We correlated the concepts that semantically represent similar constructs and found that they were also statistically similar (see Results Supporting Information S2: Table 1). This finding indicates the convergent validity of the new coding manual. The manual satisfies the quality criteria of accuracy, exclusivity, and completeness (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education 2014).

3.4 | Data Analysis

3.4.1 | Data Preparation and Baseline Equivalence Tests

We analyzed the data using RStudio version 4.2.0 (R Core Team 2022). Initially, we assessed the data for missing values and checked preconditions (e.g., normal distribution, homogeneity of variances, and the variance-covariance matrix). We then examined our sample using descriptive statistics and validated the baseline equivalence of the three groups using inferential statistical analyses of variance (ANOVAs). To correctly reject the alternative hypothesis for the baseline tests with higher test strength and a lower β-error probability (false negative), we set the significance level to 10%. We set the significance level at 5% for all other null hypotheses tests.

3.4.2 | Method for Analyzing Group Differences in Concept Use (RQ1)

We conducted three ANCOVAs with post hoc tests using Hommel (1988) adjustment to address our first RQ. This analysis approach enables us to examine the differences in the sum scores of key concepts, misconceptions, and threshold concepts among the three intervention groups (B, M, and C) and compare the results to similar studies on concept use (e.g., Aptyka et al. 2022; Nehm and Ha 2011; Opfer et al. 2012). What is more is that the selected statistical approach responds well to the structure of our data and offers several benefits owing to its step-up procedure involving sequential testing of ordered hypotheses (e.g., robustness against different sample sizes, control of α-error probability [false positive], and avoidance of unnecessary overcorrection; Hommel 1988; Lüpsen 2019). To minimize errors and increase the test power, we included the pretest knowledge about evolution (KAEVO 2.0 scores [part A and B]) as covariates, as suggested by Maxwell et al. (2017).

3.4.3 | Method for Analyzing Group Differences in Concept Networks (RQ2)

For our second RQ, we analyzed the use of concepts among the different intervention groups for the two ACORNS items using three undirected, weighted concept networks. We followed a two-step process, beginning with general analysis routines and proceeding to analysis-specific ones.

The *general analysis routines* involved preparing the data for analysis and assessing the adequacy of our networks using accuracy and stability criteria. We estimated the accuracy of the edge weights by comparing the confidence intervals (CIs) of the edge weights of the samples and the bootstrapped samples for all groups. The stability of the networks was assessed by correlating the centrality index of the original samples' strengths with that of a subsample of 70% of the original one (Epskamp et al. 2018; Wickham 2016).

The *analysis-specific routines* included formalizing the network using established recommendations (see Table 3; Burger et al. 2022; Luke 2015).

First, we defined the network components, such as nodes, edges, and attributes. We employed the R packages *statnet* (Handcock et al. 2018; Hunter et al. 2008; Statnet Development Team 2003–2024), *intergraph* (Bojanowski 2023), *igraph* (Csárdi and Nepusz 2006; Csárdi et al. 2025), and *sna* (Butts 2024) to disclose basic network indices such as the edge size, density, components, diameters, and clustering (Burger et al. 2022; Luke 2015). After that, we visualized the networks using the Fruchterman-Reingold algorithm (Fruchterman and Reingold 1991).

Moreover, we analyzed the strength and direction of relationships between each concept and the others, using the number of concept pairs and the partial Spearman's rank correlation coefficients. These indices allowed for identifying dependencies in students' concept use. We also examined the centrality of different network concepts using the centrality indices of degree, closeness, betweenness, eigenvector, and overall centralization. The degree represents the number of adjacent edges to the node and specifies which concept in the overall conception of evolution is most central. The closeness indicates the proximity of a node relative to others and informs about the centrality in terms of accessibility or closeness of one concept to the others. The betweenness centrality imparts the frequency with which a node represents the shortest path between two others. Accordingly, it highlights a concept that functions as a bridge between other concepts or clusters of concepts. The eigenvector highlights a node's impact based on its connections to other critical ones. Correspondingly, it can inform about concepts having influential neighboring concepts. The overall centralization score discloses the centrality of the overall network and can indicate the degree of overall concentration of students' concepts in the network (Bonacich 1972; Burger et al. 2022; Dablander and Hinne 2019; Koschützki et al. 2005; Siew et al. 2019). As the centrality indices serve comparable functions, we employed degree centrality to provide a concise summary of the findings while also supplementing our analyses with the results of other indices for a more transparent and comprehensive perspective. To provide an extensive understanding of the complex procedures used in the network analyses, we offer additional details on the methods and intermediate results in the supplemental materials (see Methods and Analyses, Supporting Information). The network approach allows us to analyze students' evolutionary concepts as interrelated components of conceptual networks rather than viewing concepts as isolated entities. This approach captures the natural cognitive representation of concepts comparatively well.

4 | Findings

4.1 | Baseline Equivalence Tests

We conducted baseline equivalence tests to ensure that the three groups were comparable at the beginning of the study. ANOVAs were performed for the variables: knowledge about evolution, acceptance of evolution, religiosity, and the subscales of intrinsic, germane, and extraneous cognitive load. Our findings confirm the desired null hypothesis that the groups did not significantly differ from each other ($p > 0.10$).

4.2 | The Use of Key Concepts, Misconceptions, and Threshold Concepts by Secondary School Students

When examining the frequency of students' concept use across the two ACORNS items in the posttest, the results indicate that students used diverse key concepts, misconceptions, and threshold concepts when reasoning about natural selection. Figure 2 visualizes these results and shows that although threshold concepts were suspected to play an overarching role, students used them less frequently than key concepts and misconceptions.

4.3 | Effects of Instruction on the Use of Key Concepts, Misconceptions, and Threshold Concepts When Reasoning About Natural Selection

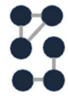
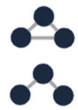
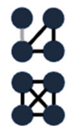
To address our first RQ, we tested whether the three intervention groups differed in their use of concepts when reasoning about natural selection after learning. The three ANCOVAs for the three dependent variables reveal significant differences between the groups regarding the use of key concepts, $F(1, 124) = 6.89$, $p = 0.009$, $\eta^2 = 0.05$, no significant results regarding the misconceptions, $F(1, 124) = 0.78$, $p = 0.378$, and a significant difference regarding the threshold concepts, $F(1, 124) = 6.15$, $p = 0.014$, $\eta^2 = 0.04$ (see Figure 3). Post hoc tests using Hommel (1988) adjustment showed that students in groups B and M did not differ significantly in the number of key concepts used, $p_{adj.} = 0.839$ ($M_{Diff} = 0.15$, 95% CI [−1.26, 1.56]). However, students who learned in group B, $p_{adj.} = 0.024$ ($M_{Diff} = 1.93$, 95% CI [0.50, 3.36]; $d_{Cohen} = 0.53$), and those who learned in group M, $p_{adj.} = 0.016$ ($M_{Diff} = 1.78$, 95% CI [0.34, 3.23]; $d_{Cohen} = 0.50$), used significantly more key concepts than students who learned in group C. The post hoc results regarding the use of threshold concepts revealed no significant differences between groups B and M, $p_{adj.} = 0.450$ ($M_{Diff} = 0.22$, 95% CI [−0.35., 0.79]), significant differences between groups B and C, $p_{adj.} = 0.014$ ($M_{Diff} = 0.72$, 95% CI [0.15, 1.30]; $d_{Cohen} = 0.51$), and no significant differences between groups M and C, $p_{adj.} = 0.088$ ($M_{Diff} = 0.51$, 95% CI [−0.08., 1.72]). Overall, the results show that students use more key concepts after learning about the threshold concepts of randomness and probability in biological

TABLE 3 | Description of the basic network terms.

Term	Description	Example
Node	*General:* ▪ Vertex that represents fundamental units or elements within the network *Study-specific:* ▪ The specific concept of evolution (e.g., causes of variation)	 Concept
Edge	*General:* ▪ The connections (arcs or relationships) between two nodes *Study-specific:* ▪ The connection of two concepts of evolution (e.g., causes of variation and changes in the distribution of variation)	Connection of two concepts
Attributes	*General (node attribute of name):* ▪ The node designation or classification *Study-specific (node attribute name):* ▪ The three concept types of key concepts, misconceptions, and threshold concepts, (colored) and the specific concept designation (abbreviation) *General (edge attribute of counted edges):* ▪ The number of the counted edges between all nodes as absolute edge weights *Study-specific (edge attribute of counted edges):* ▪ The number of the counted connection of two concepts as absolute edge weights *General (edge attribute of connection strength):* ▪ The strength of the connection between nodes *Study-specific (edge attribute of connection):* ▪ The strength of the connection in forms of the nonparametric partial Spearman's rank correlation between the concepts of evolution	 Key concept Misconception Threshold concept  Rare connections of two concepts Frequent connections of two concepts $r_S < -.3$  $r_S > .3$
Size of nodes/ edges	*General:* ▪ The number of nodes and edges that were present in the network *Study-specific:* ▪ The number of concepts and connection of two concepts about evolution in the network ▪ It represents the richness and complexity of the connections	 4 concepts 4 connections of two concepts
Density	*General:* ▪ Ratio of potential relationships (score between 0 and 1), meaning the ratio of the observable edges to the total number of possible edges *Study-specific:* ▪ Ratio of potential connections of concepts (score between 0 and 1), meaning the ratio of the observable connections to the total number of possible connections ▪ It reveals how connected all concepts are, respectively, the cohesion	 4 observable connections of two concepts 6 potential connections of two concepts

(Continues)

TABLE 3 | (Continued)

Term	Description	Example
Components	*General:* ▪ Separate clusters in a network *Study-specific:* ▪ Separate clusters of concepts about evolution in a network ▪ It specifies the number of separate subgroups and, therefore, enable the identification of isolated clusters of concepts	 2 components
Diameter	*General:* ▪ Compactness or the longest path from one node to another *Study-specific:* ▪ Compactness or the longest path from one concept about evolution to another ▪ It indicates the maximum distance between all concepts	 Diameter of 1 Diameter of 2
Clustering	*General:* ▪ Proportion of closed triangles between nodes or tendency to transitivity (score between 0 and 1) *Study-specific:* ▪ Proportion of closed triangles between concepts about evolution or tendency to transitivity (score between 0 and 1) ▪ It can point to the presence of a concept clusters	 Lower score of clustering Higher score of clustering

Note: The description of basic network terms was developed based on Burger et al. (2022) and (Luke 2015).

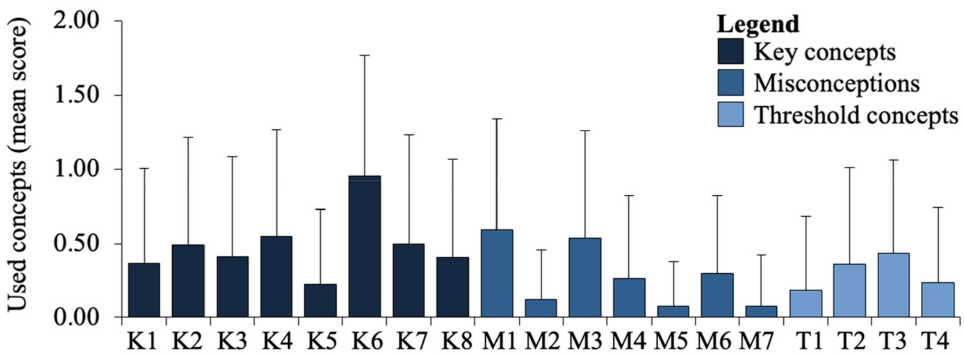


FIGURE 2 | Students' use of concepts when reasoning about natural selection in the two ACORNS items in the posttest. Used concepts (mean): $max = 2$ (absence/presence of concepts in the two ACORNS items). K1: causes of variation, K2: presence of variation, K3: survival potential, K4: selection factors or limited resources, K5: different survival of individuals, K6: changes in the distribution of variation, K7: heritability of variation, K8: differential reproductive success, M1: teleology/need, M2: use and disuse, M3: anthropomorphic conception, M4: typological conception, M5: gentle inheritance, M6: event versus process, M7: origin versus sorting out variation, T1: randomness, T2: probability, T3: temporal scales, T4: spatial scales.

(group B) and mathematical (group M) contexts and natural selection compared to students in the control group C. Students use more threshold concepts after learning about the concepts of randomness and probability in biological contexts (group B) compared to students in the control group C. Our results additionally suggest that the use of misconceptions in explaining natural selection after the intervention does not differ significantly between groups.

4.4 | Explorative Network Analysis on the Use of Concepts

4.4.1 | Accuracy and Stability of the Networks

We carried out the general analysis routines by inspecting the accuracy and stability of the networks. The accuracy analyses of the edge weights for all groups show wide CIs and partly

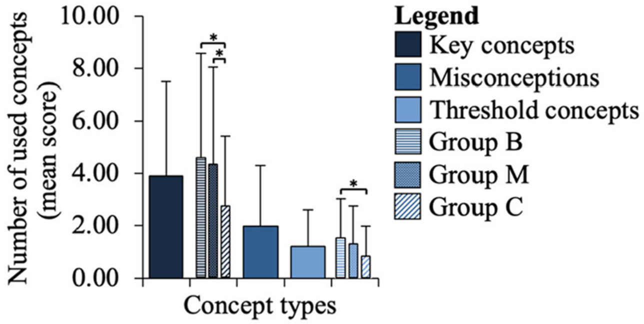


FIGURE 3 | Differences in concept use between the intervention groups in the two ACORNS items in the posttest. The figure shows the key concepts (*max* = 16), misconceptions (*max* = 14), and threshold concepts (*max* = 8) that students from different intervention groups use when responding to the two ACORNS items ($M \pm SD$). $^*p < 0.05$.

TABLE 4 | Basic network indices.

Basic description	Group B	Group M	Group C
Size of nodes/ edges	19/ 152 (992)[a]	19/ 127 (725)[a]	19/ 128 (413)[a]
Density	0.89	0.75	0.75
Components	1	1	1
Diameter	2	2	2
Clustering	0.91	0.84	0.81

Note: Nodes = concepts that act as vertices in the model, edges = connections (arcs or relationships) between nodes, density = ratio of potential relationships (score between 0 and 1), meaning the ratio of the observable edges to the total number of possible edges, components = subgroups in a network, diameter = compactness or the longest path from one vertex to another, clustering = proportion of closed triangles between vertices or tendency to transitivity (score between 0 and 1),
[a]All edges, the duplicates inclusive.

differing orders of the edge weights of the original and the bootstrapped samples (see Results Supporting Information S1: Figure 1). Furthermore, the results on the stability of the strength reveal a good level of stability for group B (CS-coefficient = 0.65) and excellent stability parameters for groups M and C (CS-coefficient > 0.70; see Results Supporting Information S1: Figure 1). The results of the edge weights indicate a lack of accuracy and should be analyzed cautiously. The correlation stability coefficients reveal good to excellent stability (Epskamp et al. 2018).

4.4.2 | Basic Description, Visualization, and Comparison of the Networks

As a first step of the analysis-specific routines, we were interested in comparing the basic network indices of the three networks. The networks had similar numbers of nodes ($n = 19$, concepts), and each network formed a single component with a diameter of 2 (see Table 4). However, notable descriptive differences were found between the networks, particularly in comparison with group B. This network had more possible edges and a higher frequency of edge usage than the networks of the other two groups. Moreover, group B's network density was 14% higher than those of the other two groups, and its clustering coefficient was 7% and 10% higher than groups M and C, respectively.

The next step of the analysis-specific routines was to visualize the networks. Figure 4 shows that each network includes eight key concepts (K1–K8), seven misconceptions (M1–M7), and four threshold concepts (T1–T4; for a more extensive representation, see Results Supporting Information S1: Figures 3–5). Also, it discloses the basic network indices, including the network density or overall cohesion, characterized by the linkage of nodes along the edges. The gray lines in this figure represent the individually weighted edges, with thicker lines indicating higher absolute edge weights and thinner lines representing smaller ones (for exact edge weights, see Results Supporting Information S2: Tables 2–4). The weights in the network of group B range from 0 (e.g., for *differential reproductive success [K8]* and *gentle inheritance [M5]*) to 23 (e.g., for the *presence of variation [K2]* and *changes in the distribution of variation [K6]*). Those in the network of group M range from 0 (e.g., for *differential reproductive success [K8]* and *gentle inheritance [M5]*) to 21 (e.g., for *changes in the distribution of variation [K6]* and *heritability of variation [K7]*). The network of group C shows the lowest number of edge weights, from 0 (e.g., for *differential reproductive success [K8]* and *gentle inheritance [M5]*) to 17 (e.g., for *teleology/need [M1]* and *anthropomorphic conception [M3]*). Overall, the highest weights are found for group B. This group used the most concept pairs after receiving instructions on threshold concepts in biological contexts and natural selection. Hence, students' concept networks regarding evolution by natural selection were the most complex.

The networks in Figure 4 also include partial Spearman's rank correlation coefficients, which show the strength and direction of the dependencies or relationships between concepts. Notably, these partial correlation coefficients already factor out the effect of all other concepts on a relationship. The strength and direction of the relationships are represented by the shading of the edges (for exact correlation coefficients, see Results Supporting Information S2: Tables 5–7). For group B, the results show several significant correlations, such as the positive relationship between the use of the key concepts of *causes of variation (K1)* and *changes in the distribution of variation (K6*; $r = 0.38$, $p < 0.05$). Additionally, most misconceptions, such as *teleology/need (M1)* and *anthropomorphic conception* (*M3*; $r = 0.38$, $p < 0.01$), are positively correlated with each other. Key concepts and misconceptions mostly exhibit a negative relationship (e.g., *causes of variation [K1]* and *teleology/need* [*M1*; $r = -0.40$, $p < 0.05$] and *changes in the distribution of variation [K6]* and *origin versus sorting out variation* [*M7*; $r = -0.46$, $p < 0.01$]). The use of threshold concepts in this group yields mixed results, with concept *randomness (T1)* taught in the intervention not significantly associated with the use of other concepts. Conversely, concept *probability (T2)* correlates moderately to strongly positively with concepts *survival potential* (*K3*; $r = 0.48$, $p < 0.05$), *gentle inheritance* (*M5*; $r = 0.39$, $p < 0.05$), and *origin versus sorting out variation* (*M7*; $r = 0.50$, $p < 0.01$), and moderately negatively with concept *differential reproductive success* (*K8*; $r = -0.38$, $p < 0.05$).

In contrast, the group M exhibits fewer significant correlations than the other groups. Positive correlations exist between concepts

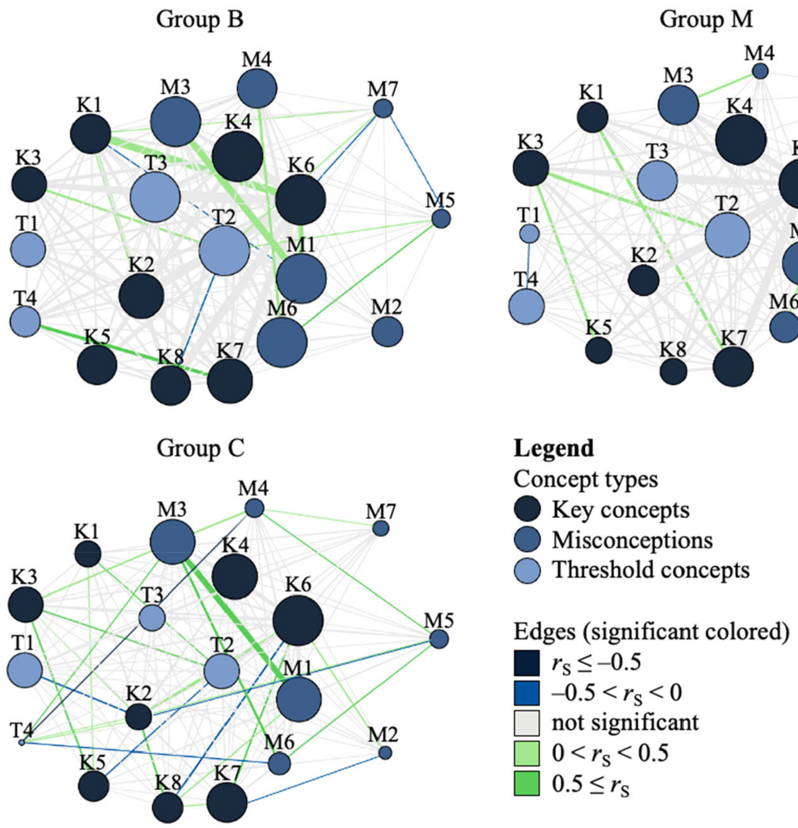


FIGURE 4 | Network showing the concept use of all intervention groups. The thickness of the lines represents the edge weights. The size of the nodes depicts the degree. K1: causes of variation, K2: presence of variation, K3: survival potential, K4: selection factors or limited resources, K5: different survival of individuals, K6: changes in the distribution of variation, K7: heritability of variation, K8: differential reproductive success, M1: teleology/need, M2: use and disuse, M3: anthropomorphic conception, M4: typological conception, M5: gentle inheritance, M6: event versus process, M7: origin versus sorting out variation, T1: randomness, T2: probability, T3: temporal scales, T4: spatial scales.

of the same type (e.g., key concepts: *causes of variation [K1]* and *heritability of variation* [*K7*; $r = 0.42$, $p < 0.05$], *survival potential [K3]* and *different survival of individuals* [*K5*; $r = 0.48$, $p < 0.05$]; misconceptions: *anthropomorphic conception [M3]* and *typological conception* [*M4*; $r = 0.45$, $p < 0.05$]). Furthermore, the taught threshold concept of *probability (T2)* and key concept of *survival potential* (*K3*; $r = 0.44$, $p < 0.05$) exhibit a moderate correlation. The threshold concepts of *randomness (T1)* and *spatial scales (T4)* correlated negatively ($r = -0.39$, $p < 0.05$).

Group C displays the highest number of significant correlations, including positive correlations between key concepts of *presence of variation (K2)* and *differential reproductive success* (*K8*; $r = 0.59$, $p < 0.01$) and *survival potential (K3)* and *different survival of individuals* (*K5*; $r = 0.63$, $p < 0.01$). This group also has an exceptionally high number of correlations between misconceptions, such as the positive correlations between *teleology/need (M1)* and *anthropomorphic conception* (*M3*; $r = 0.51$, $p < 0.05$), *anthropomorphic conception (M3)* and *typological conception* (*M4*; $r = 0.48$, $p < 0.05$), *anthropomorphic conception (M3)* and *event versus process* (*M6*; $r = 0.68$, $p < 0.01$), *typological conception (M4)* and *origin versus sorting out variation* (*M7*; $r = 0.44$, $p < 0.05$), and *gentle inheritance (M5)* and *event versus process* (*M6*; $r = 0.59$, $p < 0.01$). Regarding threshold concepts, *probability (T2)* and *causes of variation* (*K1*; $r = 0.54$, $p < 0.01$) and *probability (T2)* and *survival potential* (*K3*; $r = 0.54$, $p < 0.01$) show significant correlations. Numerous irregular correlations exist, such as the positive correlations between the key concepts and misconceptions of *survival potential (K3)* and *anthropomorphic conception* (*M3*; $r = 0.49$, $p < 0.05$), *changes in the distribution of variation (K6)* and *use and disuse* (*M2*; $r = 0.46$, $p < 0.05$), and *differential reproductive success (K8)* and *teleology/need* (*M1*; $r = 0.45$, $p < 0.05$). Furthermore, some threshold concepts correlate negatively with key concepts, such as, *randomness (T1)* with *presence of variation* (*K2*; $r = -0.43$, $p < 0.05$), *probability (T2)* with *different survival of individuals* (*K5*; $r = -0.44$, $p < 0.05$), and *spatial scales (T4)* with *survival potential* (*K3*; $r = -0.62$, $p < 0.01$) and positively with misconceptions, such as, *temporal scales (T3)* with *use and disuse* (*M2*; $r = 0.43$, $p < 0.05$), *spatial scales (T4)* with *anthropomorphic conception* (*M3*; $r = 0.56$, $p < 0.01$), and *spatial scales (T4)* with *gentle inheritance* (*M5*; $r = 0.55$, $p < 0.01$). In sum, the group B exhibited a moderate number of correlations with relatively consistent directions, the group M displayed a low number of correlations with consistent directions, and the group C revealed a high number of somewhat inconsistent correlations.

4.4.3 | Networks' Centrality Indices of Degrees, Closeness, Betweenness, and Eigenvector

The analyses of centrality indices provide insights into concepts that are most and least central across the three groups (see Results Supporting Information S2: Tables 8–10). When comparing the degree of centrality between the groups, the results show that the key concepts of *selection factors or limited resources (K4)* and *changes in the distribution of variation (K6)*, misconceptions of *teleology/need (M1)*, and *anthropomorphic conception (M3)*, and threshold concept of *probability (T2)* have the highest degree of centrality (16–18) across all groups. These findings indicate a consistency regarding the most influential connections of these concepts to others. The least central concepts vary across the groups. For instance, for group B, the concepts of *spatial scales* (*T4*; degree of 14), *randomness* (*T1*; degree of 15), and *survival potential* (*K3*; degree of 15) show relatively low centrality scores. For group M, the concepts of *randomness* (*T1*; degree of 11), *differential reproductive success* (*K8*; degree of 13), and *different survival of individuals* (K5; degree of 13) demonstrate low centrality scores. For group C, the concepts of *spatial scales* (*T4*; degree of 6), *temporal scales* (*T3*; degree of 13), and *presence of variation* (*K2*; degree of 13) indicate the lowest centrality scores. The alternative centrality indices of closeness, betweenness, and eigenvector mostly corroborate the findings of the degree of centrality of each group. When comparing the centralization scores between the groups, the degree and closeness centralization indices suggest differences between group B and the other two groups. The score of group B (degree centralization = 0.12, closeness centralization = 0.20) compared to group M (degree centralization = 0.27, closeness centralization = 0.39) and group C (degree centralization = 0.28, closeness centralization= 0.41) exhibits that several concepts have approximately the same number of relationships to others and are similarly accessible. These findings point to a highly connected, more balanced network of group B with diverse concepts compared to the other groups. The betweenness centralizations of all groups reveal that no central concepts connect all others (centralization index = 0.01–0.03), and the eigenvector centralization unveils that several concepts in the network have similarly influential neighboring concepts (eigenvector centralization = 0.04–0.09). Therefore, the most central concepts across all groups are the key concepts of *selection factors or limited resources (K4)* and *changes in the distribution of variation (K6)*, misconceptions of *teleology/need (M1)* and *anthropomorphic conception (M3)*, and threshold concept of *probability (T2)*. The most balanced, decentralized network is that of group B.

5 | Discussion

The importance of comprehensive knowledge about evolution has been established as an educational objective for students' basic biological education (National Research Council 2012; Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany 2020). Despite this objective, secondary school students struggle to acquire a scientific conception of evolution by natural selection and often exhibit numerous misconceptions (e.g., Aptyka et al. 2022; Beniermann 2019; Kuschmierz, Meneganzin et al. 2020). To address this issue, previous research has suggested that incorporating threshold concepts of randomness and probability into instruction could provide access to a more profound and holistic knowledge about evolution (Fiedler et al. 2019; Garvin-Doxas and Klymkowsky 2008; Göransson et al. 2020; Meyer and Land 2003; Ross et al. 2010; Walck-Shannon et al. 2019). Moreover, researchers have advocated treating concepts as entities within a concept network to understand the use and interrelationships of concepts at a deeper level (Amin et al. 2014; BonJour 1985; diSessa and Sherin 1998; Koponen and Nousiainen 2014; Papo et al. 2014; Siew et al. 2019; Thagard 1992). In this study, we investigated whether students differ in using key concepts, misconceptions, and threshold concepts when reasoning about natural selection after three instructional conditions. We also visualized the use of concepts through network analyses and examined how these networks were composed for each group, how concepts are related within a network, and which concepts were central or decentral in the networks.

5.1 | Instruction About Randomness and Probability in Biological Contexts Can Foster the Use of Key and Threshold Concepts

Our results on the first RQ (cf. Table 5) suggest that instructions combining information on the threshold concepts of randomness and probability in biological (group B) or mathematical (group M) contexts and natural selection resulted in a significantly stronger use of key concepts than when only receiving instruction on natural selection (group C). Furthermore, students who learned about randomness and probability in an evolutionary context and natural selection (group B) used significantly more threshold concepts than those who received no instruction on threshold concepts and only learned about natural selection (group C). These findings align with the theoretical approach, indicating that incorporating threshold concepts can expand students' knowledge about science, including the biological context of evolution (e.g., Land et al. 2016; Meyer and Land 2003). More explicitly, our findings provide some of the first empirical evidence for the effectiveness of incorporating the threshold concepts of randomness and probability into instruction on evolution at the secondary school level, supporting previous theoretical assumptions in the research field of evolution education (Fiedler et al. 2019; Garvin-Doxas and Klymkowsky 2008; Göransson et al. 2020; Kampourakis and Zogza 2009; Ross et al. 2010; Walck-Shannon et al. 2019). Additionally, we extend the findings of Fiedler et al. (2018), who suggested that biological and mathematical knowledge of randomness and probability represent distinct knowledge domains that predict knowledge about evolution and enable a more profound scientific knowledge. Overall, these results hold important implications for educators who aim to enhance education on key concepts of evolution. Specifically, these findings suggest that teaching threshold concepts such as randomness and probability in mathematics or biology classes before or during evolution may potentially promote a more profound knowledge about evolution among secondary school students. The effect of the intervention on learning about evolution by natural selection can be classified as moderate after only a short period of instruction. Noteworthy is that the effect is higher than the average effect of interventions (cf. Hattie 2008). Therefore, we recommend prioritizing teaching these threshold concepts in mathematics and biology classes when evaluating and adapting the curriculum.

TABLE 5 | Key findings and implications for science education.

RQ	Key findings	Implications for science education
1	***How do groups of students differ in their use of key concepts, misconceptions, and threshold concepts after receiving different instruction on threshold concepts within material covering natural selection?***	
	Group differences in the use of key concepts, misconceptions, and threshold concepts	
	Students use more key concepts after learning about the threshold concepts of randomness and probability in biological (group B) and mathematical (group M) contexts and natural selection compared to students in the control group C.	The clarification of threshold concepts enhances the use of key concepts. Hence, science education could prioritize an interdisciplinary integration of threshold concepts into curricula to deepen students' knowledge.
	Students used more threshold concepts after learning about the concepts of randomness and probability in biological contexts (group B) compared to students in the control group C.	The clarification of threshold concepts within contexts closely aligned with the targeted instructional context has the potential to significantly enhance students' application and integration of these concepts.
	The use of misconceptions in explaining natural selection after the intervention does not differ significantly between groups.	Misconceptions are resilient to short-term interventions on threshold concepts and natural selection. The need pertains to focus instructional efforts in science education to directly address and correct these misconceptions, using more targeted strategies.
2	***How do students in different intervention groups form concept networks when reasoning about natural selection, and how are these concepts relate to each other regarding their interconnections and centrality within the network?***	
	Size of nodes and edges	
	All networks showed that the students used all 19 concepts, however the network of group B had more possible edges and a higher frequency of edge usage than the networks of the other two groups. *Density* The density of group B's network was 14% higher than those of the other two groups. *Components and diameter* Each network formed a single component with a diameter of 2. *Clustering* The group B's clustering coefficient was 7% and 10% higher than those of groups M and C.	The thematization of the threshold concept in science lessons can lead to comparably higher network complexity and a stronger cognitive interconnectedness or clustering among concepts. The broad range of used concepts shows that students are generally capable of engaging with diverse concepts when provided with appropriate instructional support.
	Edge weights	
	The highest edge weights are found for group B. *Connection strengths* The group B shows a moderate number of correlations with relatively consistent directions, the group M displays a low number of correlations with consistent directions, and the group C reveals a high number of rather inconsistent correlations.	Instructions on threshold concepts in biological contexts and natural selection can foster the use of interconnecting of concepts, consistent reasoning, and the complexity of concept networks. Interdisciplinary instruction on threshold concepts, such as those in mathematical contexts and natural selection, may require complementary strategies to fully engage students in making meaningful conceptual connections. The high number of inconsistent correlations in group C reflects fragmented understanding, underscoring the need for targeted interventions that do not only address natural selection but basic tacit knowledge (e.g., threshold concepts).
	Centrality (degree, closeness, betweenness, and eigenvector)	
	The key concepts of selection factors or limited resources and changes in the distribution of variation, misconceptions of teleology/need and anthropomorphic conception, and threshold concept of probability have the highest degree of centrality across all groups.	There are stable concepts across the groups that stand for the most influential concepts and are therefore most often associated or co-occurring with the other concepts. Teaching that emphasizes the most influential key concepts, misconceptions, and threshold concepts (generally or individually) could be beneficial for students.

(Continues)

TABLE 5 | (Continued)

RQ	Key findings	Implications for science education
	Centralization (degree, closeness)	
	The network of group B reveals several concepts with a similar number of connections and comparable accessibility, indicating a highly connected and balanced and decentral structure, which suggests that a diverse range of concepts is being utilized more evenly in group B compared to the other groups.	Instructions on threshold concepts in biological contexts and natural selection can promote conceptual diversity; increase the accessibility of knowledge; foster a more decentralized and distributed cognitive model, where knowledge is not clustered around a few dominant ideas but is instead dispersed across multiple concepts; and enhance the potential for more robust knowledge and the interrelationship of concepts.
	Overarching findings	
	Benefits of concept networks and perspectives on change theory	
	These findings suggest that analyzing conceptions of evolution as interconnected concept networks offers a valuable tool for enriching traditional instruction. Additionally, this approach provides empirical evidence that could contribute to the ongoing and unresolved debates and different perspectives on the conceptual change theory that have persisted for over four decades. The findings related to the network analysis imply that the perspectives of knowledge-in-theory and knowledge-in-pieces on conceptual change could be part of a unified continuum, emphasizing the importance of refining and combining these viewpoints to enhance the theory's overall strength. Furthermore, applying network theory offers a deeper understanding of how students use and structure concepts, potentially shedding light on the cognitive processes behind conceptual change and guiding future research in this area.	

Furthermore, our study revealed no significant differences in the use of misconceptions among students of different interventions, as we could not provide evidence for the theoretical approaches regarding the advantages of threshold concepts for reducing students' use of misconceptions (Göransson et al. 2020; Kampourakis and Zogza 2009; Ross et al. 2010; Taylor 2006). Hence, our empirical results indicate that our approach is less effective in minimizing the use of misconceptions than some of the previous studies (e.g., Aptyka et al. 2022; Nehm et al. 2022). One possible explanation is that instructional interventions may have been insufficient to initiate a reduction of misconceptions (such as *teleology/need*) in our study (Aptyka et al. 2022; Duit and Treagust 2003; Gregory 2009; Southerland et al. 2001). Students may not have fully consolidated and comprehended the key and threshold concepts. Consequently, they could not benefit from the positive effects of comprehending these scientific concepts (Batzli et al. 2016; Walck-Shannon et al. 2019). In cases where students still used misconceptions, they may have perceived them to be the most helpful concepts for reasoning about natural selection (Hartelt and Martens 2024). Walck-Shannon et al. (2019) suggested that fully grasping threshold concepts can vary in the amount of time and effort required and may produce varying degrees of effectiveness on the overall increase of knowledge about evolution. Hence, our study highlights the need for further research to investigate how quickly educators can promote fully grasping threshold concepts in different individuals. Additionally, it should be studied at what level of expertise the knowledge of threshold concepts can help reduce the use of misconceptions. Future studies could consider the recommendations of Nehm et al. (2022) and investigate the effects of different magnitudes of threshold concept interventions.

Another possible explanation for the missing intervention effect on misconceptions is that students may have acquired new key and threshold concepts but were unaware that they contradicted their pre-existing misconceptions. This lack of awareness can hinder conceptual change (Gartmeier et al. 2008; Kampourakis and Zogza 2009; Limón 2001; Nelson 2008; Oser et al. 2012; Rea-Ramírez and Clement 1998; Tulis et al. 2016). As previous studies achieved a significant minimization of misconceptions about evolution, for instance, by leveraging a misconception-focused instruction (e.g., Aptyka et al. 2022; Nehm et al. 2022), we hypothesize that students may have benefited from a more substantial clarification of misconceptions.

5.2 | Teaching About Randomness and Probability in Biological Contexts Promotes the Integration of Concepts and the Development of Complex and Dense Concept Networks

Our network approach and the perspective that concepts are structured in concept networks led to innovative and profound results regarding our second RQ (cf. Table 5) that significantly complement the implications of previous studies. The basic network structures of the three intervention groups yielded similar results regarding the number of concepts used, the components, and the diameters. These results indicate that the three groups used all 19 concepts to describe evolution and showed no isolated clusters of concepts. However, we observed differences in the number of edges, density, and clustering indices. The group B exhibited the highest values, followed by the groups M and C. The results imply that the network of group B is more complex, richer in concepts and conceptual relationships, and more diverse than the other networks. Interdisciplinary research has shown that such structures indicate strong cognitive interconnectedness among concepts (Koponen and Pehkonen 2010; Papo et al. 2014; Siew et al. 2019). The analyses of the relationships between concepts via the edge weights highlight the more vital connectedness of concepts in the network of group B. This group showed the most connections, followed by groups M and C. Using the network approach, we can complement current findings on the effects of threshold concepts on knowledge and learning (cf. Mead and Scott 2010; Meyer and Land 2003; Tibell and Harms 2017), approach the clarification of the explanatory role of threshold concepts (Salwén 2021), and derive the implication that the input about the threshold concepts of randomness and probability, particularly in biological contexts, can lead to comparably higher network complexity and a more substantial cognitive interconnectedness among concepts.

5.3 | The Correlation Patterns Observed in the Groups Indicate Divergences in the Way Students Integrate Concepts

We identified several significant correlations between the concepts used. For groups B and M, we found correlation tendencies showing that key concepts correlate positively with other key and threshold concepts, key concepts correlate negatively with misconceptions, and misconceptions relate positively to others. These findings are consistent with correlation trends in other studies (Nehm and Ha 2011; Opfer et al. 2012).

Group C showed the most arbitrary correlations between different concept types. Students used scientific (i.e., key and threshold concepts) and unscientific concepts (i.e., misconceptions) in a mixture, such that contradictory concepts coexisted in students' reasoning about evolution. By contrasting our findings with prior research, it becomes evident that a considerable number of secondary (Hartelt and Martens 2024) and university (e.g., Nehm and Ha 2011; Opfer et al. 2012) students, as well as preservice teachers (e.g., Aptyka and Großschedl 2022; Großschedl et al. 2018), possess coexisting scientific and unscientific concepts, which they jointly use when reasoning about evolution. In our study, the inconsistent and arbitrary use of concepts within group C indicates that concepts that should inherently oppose each other are seemingly not perceived as contradictory by the students. Consequently, we infer that Group C maintained a limited knowledge about evolution even after the intervention. Students in this group may be categorized as novice learners (Aptyka et al. 2022; Evans 2001; Opfer et al. 2012).

Although we identified positive effects of instructions on randomness and probability on the subsequent use of concepts in the results on the first RQ, the correlative results are not straightforward. Unexpectedly, the groups showed mixed correlation tendencies for the relationships of threshold concepts with others. Contrary to theoretical assumptions, the threshold concepts did not exclusively relate positively to key concepts and negatively to misconceptions (Göransson et al. 2020; Ross et al. 2010).

A first explanation for the arbitrariness may be that the groups used the threshold concepts in a preliminary, undifferentiated form. Considering the theory of threshold concepts, the diverse mixture of concepts could be a typical step in the learning process, as students were at different stages of liminality, oscillating and reconstructing their understanding of the concepts (Meyer and Land 2005, 2006a, 2006b). Novice learners may inadvertently mix concepts and only after gaining more knowledge about evolution distinguish between key concepts, threshold concepts, and misconceptions when explaining evolution by natural selection (Batzli et al. 2016; Fiedler et al. 2019; Smith et al. 1985; Walck-Shannon et al. 2019).

Second, another explanation for the persistence of mixed models could be the nonrecognition of compatibility (Aptyka et al. 2022; Evans 2001; Opfer et al. 2012) between threshold and key concepts and the contradictions with misconceptions.

Third, it can be surmised from the results of the network analyses that while threshold concepts are not directly correlated with the use of other concepts, they may encourage students to change their perspectives and broaden their understanding of previously inaccessible concepts (cf. Mead and Scott 2010; Meyer and Land 2003; Tibell and Harms 2017). This change of perspective may be indicated by students' restructuring of their knowledge and the way they connect concepts in their reasoning about evolution. If this explanatory approach holds value, it could also account for the structural differences in the network of groups B and M compared to group C.

Overall, a possible future approach to rendering the black box of gradual conceptual change explainable is to adopt theoretical approaches of the threshold concepts (Meyer and Land 2006a) and combine them with research perspectives similar to those of Walck-Shannon et al. (2019) and Settlage (2007). Researchers could focus on analyzing the change of entire networks over a prolonged period with multiple measurement points rather than examining the preference for individual concepts at only a few points.

5.4 | All Intervention Groups Coherently Use the Most Central Concepts of Each Concept Type

To discuss the second RQ further, we use the results of the centrality analyses of concepts in the networks. Our analyses revealed that of the three concept types, the key concepts of *selection factors or limited resources (K4)* and *changes in the distribution of variation (K6)*, misconceptions of *teleology/need (M1)* and *anthropomorphic conception (M3)*, and threshold of *probability (T2)* were most central and influential across all intervention groups. These findings suggest that regardless of the interventional treatment, five concepts appear to be used consistently across all groups. At this point, it is essential to emphasize that concepts may not be activated, used, and connected exclusively as discrete knowledge pieces or elements (cf. diSessa and Sherin 1998; Özdemir and Clark 2007). The observed cohesive pattern implies the potential existence of at least one phase in the process of conceptual change in which students cognitively establish a theory-like pattern of concepts that appears relatively stable across groups (e.g., Carey 1999; Gopnik and Wellman 1994; Ionnides and Vosniadou 2002; Wellman and Gelman 1992). Hence, we can derive the theoretical implication of the network-related findings that the two perspectives on the conceptual change process (i.e., knowledge-in-theory and knowledge-in-pieces) may lie on a shared continuum. Given this assumption, refining and approximating different perspectives on the theory is essential to enhance its robustness.

Beyond the insights gained from the identified patterns of concepts, the centrality indices of the network analyses indicate that the statistically most central key concepts (i.e., *selection factors or limited resources [K4]* and *changes in the distribution of variation [K6]*) identified in our study differed from the established core concepts of variation, heritability, and individual fitness (represented by approximate concepts of presence of variation [E2], heritability of variation [E7], and survival potential [E3] in our study; Federer et al. 2015; Nehm et al. 2012; Nehm and Ha 2011; Nehm et al. 2010; Nehm and Reilly 2007). As researchers propose that students' conceptions of evolution must encompass at least the three core concepts to facilitate scientific reasoning about natural selection in various contexts, it is imperative to elucidate the disparity between the use of central concepts and core concepts. Furthermore, an investigation is warranted to explore how students should be supported by instruction to ensure that the core concepts represent the central and most influential concepts in conceptual networks.

We advise that future research draws value from our novel findings on the centrality of concepts and centralization of networks and adapt instructions more precisely to the specific needs of students, as this could increase the accessibility of knowledge. According to research in cognitive network science, using a particular concept can activate other connected concepts in the semantic memory, a process known as the spreading activation (Anderson 1983; Collins and Loftus 1975). While we cannot determine the direction of spreading activation patterns from our analyses, the statistical results and theoretical underpinnings allow us to propose assumptions that should be critically tested. In promoting core concepts, we suggest that the concepts of *causes of variation* (*K1*; group B) and *survival potential (K3)*, *changes in the distribution of variation* (*K6*), *differential reproductive success (K8)*, and *spatial scales* (*T4*; group C) could activate the concept of *presence of variation (K2)*. Concepts of *spatial scales* (*T4*; group B), *causes of variation* (*K1*; group M), *changes in the distribution of variation (K6)*, and *differential reproductive success* (*K8*; group C) could be proximal to teaching the concept of *heritability of variation (K7)*. Additionally, addressing the concepts of *probability* (*T2*; all intervention groups) and *different survival of individuals* (*K5*; groups M, C) could be beneficial for learning the concept of *survival potential (K3)*.

Furthermore, our results on misconceptions support the existing research on their relatively persistent nature (e.g., Aptyka et al. 2022; Beniermann 2019; Kelemen 2011; Kuschmierz, Meneganzin et al. 2020). The prominence of the misconceptions of *teleology/need (M1)* and *anthropomorphic conception (M3)* in our centrality analyses highlight their influence on the use of other concepts in students' conceptual networks. Additionally, this finding emphasizes the need to address the reduction of these misconceptions more stringently. Consistent with previous research, our results suggest that these misconceptions (i.e., teleological and anthropomorphic concepts) are prevalent in secondary school (Aptyka et al. 2022; Beniermann 2019; Gregory 2009; Kelemen 2011; Smith 2010) and may persist until university education (Aptyka and Großschedl 2022; Bishop and Anderson 1990; Großschedl et al. 2014; Großschedl et al. 2018; Nehm and Ha 2011; Nehm and Reilly 2007; Nehm and Schonfeld 2008; Opfer et al. 2012; Rachmatullah et al. 2018; Rector et al. 2013).

To leverage our insights, we additionally posit assumptions about misconceptions, researchers should scrutinize in future studies. Our analyses reveal that a combination of a higher frequency of the concept of *causes of variation (K1)* and a lower frequency of *anthropomorphic conceptions (M3)* is associated with a lower frequency of the misconception of *teleology/need (M1)* in group B. This finding indicates that promoting the concept of *causes of variation (K1)* and reducing *anthropomorphic conceptions (M3)* could also decrease the use of the misconception of *teleology/need (M1)*. Furthermore, in group M, the misconceptions of *teleology/need (M1)* and *event versus process (M6)*, as well as *anthropomorphic conception (M3)* and *typological conception (M4)*, exhibit positive correlations. The correlations suggest that addressing one of the paired concepts could reduce the use of the other. Similarly, in group C, the concept pairs of *teleology/need (M1)* and *anthropomorphic conception (M3)*, *anthropomorphic conception (M3)* and *typological conception (M4)*, and *anthropomorphic conception (M3)* and *event versus process (M6)* could imply that decreasing one of the concepts curtails the use of the other.

Additionally, our analysis of group C suggests that students did not recognize the contradictions between specific pairs of concepts. For example, despite representing theoretically opposite components, the concepts of *teleology/need (M1)* and *differential reproductive success (K8)*, *anthropomorphic conception (M3)* and *survival potential (K3)*, and *anthropomorphic conception (M3)* and *spatial scales (T4)* correlate positively. This finding supports prior studies emphasizing the need to clarify misconceptions (e.g., Andrews et al. 2011; Gartmeier et al. 2008; Limón 2001; Nelson 2008; Rea-Ramírez and Clement 1998). Furthermore, we hypothesize that promoting a cluster of related concepts rather than distinct concepts may be more effective in guiding students' concept development in a preferred direction. Combining specific concepts, such as those listed above, could be a promising approach.

5.5 | Limitations

The limitations of this study must be considered alongside its strengths. First, as we aimed to conduct initial exploratory analyses, we used a relatively small sample. The sample size may have impacted the accuracy and limited the statistical power of some analyses (Cohen 1988). Nevertheless, the study serves its purpose. We aimed to adopt an exploratory approach to our data analyses that allowed us to identify important trends and provide novel starting points for future hypothesis-testing studies and fundamental theoretical research.

Second, the allocated time for the study was relatively brief, comprising four school hours (over 3 days), with 90 min dedicated to the intervention. We opted for a short time frame for this study since teachers frequently report having limited time to teach biology (cf. Kilic 2012). We aimed to avoid overburdening them. Additionally, the shorter time frame has the advantage of reducing uncontrollable confounding variables (e.g., external input on the topic of this study, participant absences, and canceled classes) while still achieving initial intervention effects (cf. Aptyka et al. 2022). However, as discussed above, we cannot exclude the possibility that the short time frame may have restricted the overall learning effects in the form of an increased use of key and threshold concepts as well as a decreased use of misconceptions. Extending the intervention would likely have led to higher learning outcomes (Nehm et al. 2022; Paçacı 2022).

Third, we chose rather transferential and traditional teaching approaches in our intervention than one that aligns more closely with other approaches, such as the one on Socioscientific Issues (SSI) in conceptual change (Högström et al. 2024). We carefully selected the contexts for our study (e.g., the random and probabilistic impacts of climate change; dice games as analogies for randomness; exploring analogies of plants in nature; evolutionary changes in plants and animals) to align with standard school curricula (Ministry for Schools and Further Education of the State North Rhine-westphalia 2022). Simultaneously, we aimed to use methods that minimized the effect of confounding variables (e.g., effects of methods such as group discussions) within the groups to primarily investigate the impact of addressing threshold concepts in evolution education. However, we acknowledge that this approach may represent a limitation in our study. Therefore, SSI could serve as a lens for exploring alternative instructional approaches that align more closely with the transformative goals associated with threshold concepts and offer valuable insights into fostering students' conceptual change.

Fourth, although the measure GAENE 2.1 (Smith et al. 2016) has passed several validation steps, Sbeglia and Nehm (2018) identified psychometric shortcomings regarding individual items' fits and rating scale functioning. This information could be crucial for studies aiming to focus more heavily on the results of the acceptance of evolution. However, because our goal was to use the level of acceptance of evolution only to transparently describe the sample, test the baseline equivalence, and provide a basis for comparison with other studies, the advantages outweigh the potential disadvantages.

A fifth limitation is the restricted comparability of our results with similar studies due to the novelty of the theoretical and methodological approaches we used. Few studies have explored the use of threshold concepts in instruction on evolution and the concept network methodology in evolution education. Additionally, we used the refined coding manual of Göransson et al. (2020), which is relatively new, and the coded concepts may differ marginally in depth and meaning from similar concepts coded with a different coding manual. Despite this, we demonstrated the interrater reliability and convergent validity of the refined coding manual (Göransson et al. 2020), supporting its use in our study rather than other scoring guides such as the one for the ORI and EGALT (Nehm et al. 2010).

6 | Conclusion

In conclusion, this study marks a significant step toward understanding effective evolution education and students' concept networks concerning evolutionary processes. The findings demonstrate that the theoretical idea of threshold concepts is supported so far that it may potentially enhance evolution education and unlock innovative perspectives concerning key and threshold concepts when taught explicitly and (inter-) disciplinarily. The findings suggest that threshold concepts can promote a more incredible wealth of knowledge and facilitate restructuring students' understanding by fostering more interconnected concept networks. However, our results do not support the theoretical idea that threshold concepts are necessary for discarding misconceptions.

Our findings of the concept network analysis can guide educators in enriching traditional instruction on evolution and highlight the potential of analyzing conceptions of evolution as concept networks. Treating conceptions of evolution as concept networks and analyzing them based on the network theory goes beyond previous work and provides a more nuanced understanding. To meaningfully link with our results, we strongly advocate future research selecting a similar perspective and focusing on complementary analyses. Future research projects can thus expand these results, for example, by investigating differences in concept networks between novices and experts or in different contexts of evolution (e.g., the polarity of trait change in animals vs. plants). Longitudinal studies can also examine changes in networks or network clusters over time. The theoretical approach of network analyses can provide a more nuanced understanding of beneficial networks or context-related structures and holds great potential for future research in evolution education and related scientific disciplines.

Ultimately, it should be emphasized that the implementation of the study and the analysis of the data may vary significantly depending on the theoretical framework underpinning the instructional approaches (e.g., Socio-Scientific Issues [SSI] within the context of conceptual change) and the interpretative lens applied to the data. Given the intricate interplay of instructional effects on students' learning processes and cognitive structures, this study does not purport to resolve the longstanding black box of over four decades of conceptual change research or to provide definitive conclusions free from contention. Nevertheless, it succeeds in offering a pioneering impetus to catalyze new interdisciplinary discussions grounded in diverse theoretical perspectives.

Author Contributions

J.G., D.F., and H.A. played a central role in shaping the research objectives and the research design. H.A. mainly contributioned to the literature review and was responsible for data collection. H.A. led the manuscript writing, conducted the statistical analyses, and created the figures. All authors were involved in critically reviewing and providing feedback on the final manuscript, which they all read and approved.

Acknowledgments

We thank the secondary school teachers and school students for their valuable help in successfully implementing this study. The author(s) received no financial support from third-party funders for the research, authorship, and/or publication of this article. Open Access funding enabled and organized by Projekt DEAL.

Ethics Statement

This study followed educational legislation, ethical guidelines, and research participant protection protocols. Using a unique identifier enabled the anonymous pooling of materials, and no compensation or incentives were offered to the subjects.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data set used to draw the inferences in this study is based on secondary school students and is inaccessible to the public. Nevertheless, interested parties may obtain access to the data set through a reasonable request to the corresponding author.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. 2014. *Standards for Educational and Psychological Testing*. American Educational Research Association.

Amin, T. G., C. L. Smith, and M. Wiser. 2014. "Student Conceptions and Conceptual Change: Three Overlapping Phases of Research." In *Handbook of Research on Science Education (Vol. 2)*, edited by N. Lederman and S. Abell, 71–95. Routledge. https://doi.org/10.4324/9780203097267-12.

Anderson, J. R. 1983. "A Spreading Activation Theory of Memory." *Journal of Verbal Learning and Verbal Behavior* 22, no. 3: 261–295. https://doi.org/10.1016/S0022-5371(83)90201-3.

Andrews, T. M., M. J. Leonard, C. A. Colgrove, and S. T. Kalinowski. 2011. "Active Learning Not Associated With Student Learning in a Random Sample of College Biology Courses." *CBE—Life Sciences Education* 10, no. 4: 394–405. https://doi.org/10.1187/cbe.11-07-0061.

Aptyka, H., D. Fiedler, and J. Großschedl. 2022. "Effects of Situated Learning and Clarification of Misconceptions on Contextual Reasoning About Natural Selection." *Evolution: Education and Outreach* 15, no. 5: 1–15. https://doi.org/10.1186/s12052-022-00163-5.

Aptyka, H., and J. Großschedl. 2019. "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.

Aptyka, H., and J. Großschedl. 2022. "Analyzing Pre-Service Biology Teachers' Intention to Teach Evolution Using the Theory of Planned Behavior." *Evolution* 15, no. 16: 16. https://doi.org/10.1186/s12052-022-00175-1.

Barthlott, W., M. D. Rafiqpoor, and W. R. Erdelen. 2016. "Bionics and Biodiversity—Bio-Inspired Technical Innovation for a Sustainable Future." In *Biomimetic Research for Architecture and Building Construction: Biologically-Inspired Systems (Vol. 8)*, edited by J. Knippers, K. G. Nickel, and T. Speck, 11–55. Springer. https://doi.org/10.1007/978-3-319-46374-2_3.

Batzli, J. M., J. K. Knight, L. M. Hartley, A. C. Maskiewicz, and E. A. Desy. 2016. "Crossing the Threshold: Bringing Biological Variation to the Foreground." *CBE—Life Sciences Education* 15, no. 4: es9. https://doi.org/10.1187/cbe.15-10-0221.

Becker, N., E. Dittmer, and J. Weber. 2019. *Natura Biologie für Gymnasien—Oberstufe [Natura Biology for grammar schools—upper school]*. Westermann Verlag.

Beniermann, A. 2019. *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]*. Springer Spektrum. https://doi.org/10.1007/978-3-658-24105-6.

Beniermann, A., P. Kuschmierz, A. Pinxten, et al. 2021. 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. https://zenodo.org/record/4554742.

Bennett, D. 2011. "Defining Randomness." In *Philosophy of Statistics*, edited by P. S. Bandyopadhyay and M. R. Forster, 7, 633–639. Elsevier. https://doi.org/10.1016/B978-0-444-51862-0.50020-4.

Bishop, B. A., and C. W. Anderson. 1990. "Student Conceptions of Natural Selection and Its Role in Evolution." *Journal of Research in Science Teaching* 27, no. 5: 415–427. https://doi.org/10.1002/tea.3660270503.

Bojanowski, M. 2023. "Intergraph: Coercion Routines for Network Data Objects." https://mbojan.github.io/intergraph/.

Bonacich, P. 1972. "Technique for Analyzing Overlapping Memberships." In *Sociological Methodology*, edited by H. L. Costner, 4, 176–185. Jossey-Bass. https://doi.org/10.2307/270732.

BonJour, L. 1985. *The Structure of Empirical Knowledge*. Harvard University Press.

Boshuizen, H. P. A., H. Gruber, and J. Strasser. 2020. "Knowledge Restructuring Through Case Processing: The Key to Generalise Expertise Development Theory Across Domains?" *Educational Research Review* 29: 100310. https://doi.org/10.1016/j.edurev.2020.100310.

Bressan, P. 2002. "The Connection Between Random Sequences, Everyday Coincidences, and Belief in the Paranormal." *Applied Cognitive Psychology* 16, no. 1: 17–34. https://doi.org/10.1002/acp.754.

Brock, R. 2017. "Tacit Knowledge in Science Education." In *Science Education: An International Course Companion*, edited by K. S. Taber and B. Akpan, 133–142. Sense Publishers. https://doi.org/10.1007/978-94-6300-749-8_10.

Buiatti, M., and G. Longo. 2013. "Randomness and Multilevel Interactions in Biology." *Theory in Biosciences* 132, no. 3: 139–158. https://doi.org/10.1007/s12064-013-0179-2.

Burger, J., A. M. Isvoranu, G. Lunansky, et al. 2023. "Reporting Standards for Psychological Network Analyses in Cross-Sectional Data." *Psychological Methods* 28: 806–824. https://doi.org/10.1037/met0000471.

Butts, C. T. 2024. "sna: Tools for Social Network Analysis." https://CRAN.R-project.org/package=sna.

Carey, S. 1999. "Sources of Conceptual Change." In *Conceptual Development: Piaget's Legacy*, edited by K. N. E. K. Scholnick and P. Miller, 293–326. Lawrence Erlbaum Associates.

Carroll, S. P., P. S. Jørgensen, M. T. Kinnison, et al. 2014. "Applying Evolutionary Biology to Address Global Challenges." *Science (New York, N.Y.)* 346, no. 1245993: 1245993. https://doi.org/10.1126/science.1245993.

Clark, D. B. 2006. "Longitudinal Conceptual Change in Students' Understanding of Thermal Equilibrium: An Examination of the Process of Conceptual Restructuring." *Cognition and Instruction* 24: 467–563.

Cohen, J. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20, no. 1: 37–46. https://doi.org/10.1177/001316446002000104.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed. Laurence Erlbaum and Associates).

Collins, A. M., and E. F. Loftus. 1975. "A Spreading-Activation Theory of Semantic Processing." *Psychological Review* 82, no. 6: 407–428. https://doi.org/10.1037/0033-295X.82.6.407.

Colton, J., G. Sbeglia, S. Finch, and R. H. Nehm. 2018. "A quasi-Experimental Study of Short-and Long-Term Learning of Evolution in Misconception-Focused Classes." Paper Presented at the American Educational Research Association International conference (AERA), New York, NY.

Common Core State Standards Initiative. 2010. "Common Core State Standards for Mathematics." https://www.nctm.org/uploadedFiles/Standards_and_Positions/Common_Core_State_Standards/Math_Standards.pdf.

Csárdi, G., and T. Nepusz. 2006. "The igraph Software Package for Complex Network Research." *Interjournal Complex Systems* 1695, no. 5: 1–9.

Csárdi, G., T. Nepusz, V. Traag, et al. 2025. "igraph: Network Analysis and Visualization in R." In (Version R Package Version 2.1.4) https://CRAN.R-project.org/package=igraph.

Dablander, F., and M. Hinne. 2019. "Node Centrality Measures Are a Poor Substitute for Causal Inference." *Scientific Reports* 9: 6846. https://doi.org/10.1038/s41598-019-43033-9.

Deniz, H., L. A. Donnelly, and I. Yilmaz. 2008. "Exploring the Factors Related to Acceptance of Evolutionary Theory Among Turkish Preservice Biology Teachers: Toward a More Informative Conceptual Ecology for Biological Evolution." *Journal of Research in Science Teaching* 45, no. 4: 420–443. https://doi.org/10.1002/tea.20223.

diSessa, A. A. 1993. "Toward an Epistemology of Physics." *Cognition and instruction* 10, no. 2–3: 105–225. https://doi.org/10.1080/07370008.1985.9649008.

diSessa, A. A., N. M. Gillespie, and J. B. Esterly. 2004. "Coherence Versus Fragmentation in the Development of the Concept of Force." *Cognitive Science* 28, no. 6: 843–900. https://doi.org/10.1016/j.cogsci.2004.05.003.

diSessa, A. A., and B. L. Sherin. 1998. "What Changes in Conceptual Change?" *International Journal of Science Education* 20, no. 10: 1155–1191. https://doi.org/10.1080/0950069980201002.

Duit, R., and D. F. Treagust. 2003. "Conceptual Change: A Powerful Framework for Improving Science Teaching and Learning." *International Journal of Science Education* 25, no. 6: 671–688. https://doi.org/10.1080/09500690305016.

Entwistle, N. 2008. "Threshold Concepts and Transformative Ways of Thinking Within Research Into Higher Education." In *Threshold Concepts Within the Disciplines*, edited by R. Land, J. H. F. Meyer, and J. Smith, 21–35. Brill. https://doi.org/10.1163/9789460911477_003.

Epskamp, S., D. Borsboom, and E. I. Fried. 2018. "Estimating Psychological Networks and Their Accuracy: A Tutorial Paper." *Behavior Research Methods* 50, no. 1: 195–212. https://doi.org/10.3758/s13428-017-0862-1.

Esders, S., W. Kleesattel, A. Gnoyke-Sitterz, et al. 2015. *Biologie Oberstufe SII [Biology Upper Level SII]* (3rd ed. Cornelsen Verlag).

Eurydice. 2021. "National Specificities of the Education System." https://eacea.ec.europa.eu/national-policies/eurydice/content/germany_en.

Evans, E. M. 2001. "Cognitive and Contextual Factors in the Emergence of Diverse Belief Systems: Creation Versus Evolution." *Cognitive Psychology* 42, no. 3: 217–266. https://doi.org/10.1006/cogp.2001.0749.

Federer, M. R., R. H. Nehm, J. E. Opfer, and D. Pearl. 2015. "Using a Constructed-Response Instrument to Explore the Effects of Item Position and Item Features on the Assessment of Students' Written Scientific Explanations." *Research in Science Education* 45, no. 4: 527–553. https://doi.org/10.1007/s11165-014-9435-9.

Feller, W. 1968. "*An Introduction to Probability Theory and Its Applications (Vol. 1)*." John Wiley & Sons, Inc.

Fiedler, D., A. Moormann, and A. Beniermann. 2023. "Using Different Acceptance Measures: The Interplay of Evolution Acceptance, Evolution Understanding, and Religious Belief Among German Preservice Biology Teachers, Secondary School Students, and Creationists." *Science Education* 108, no. 1: 223–274. https://doi.org/10.1002/sce.21833.

Fiedler, D., G. C. Sbeglia, R. H. Nehm, and U. Harms. 2019. "How Strongly Does Statistical Reasoning Influence Knowledge and Acceptance of Evolution?" *Journal of Research in Science Teaching, John Wiley and Sons Inc* 56, no. 9: 1183–1206. https://doi.org/10.1002/tea.21547.

Fiedler, D., S. Tröbst, J. Großschedl, and U. Harms. 2018. "Evosketch: Simple Simulations for Learning Random and Probabilistic Processes in Evolution, and Effects of Instructional Support on Learners' Conceptual Knowledge." *Evolution: Education and Outreach* 11, no. 15: 1–17. https://doi.org/10.1186/s12052-018-0089-3.

Fiedler, D., S. Tröbst, and U. Harms. 2017. "University Students' Conceptual Knowledge of Randomness and Probability in the Contexts of Evolution and Mathematics." *CBE—Life Sciences Education* 16, no. 2: ar38. https://doi.org/10.1187/cbe.16-07-0230.

Fruchterman, T. M., and E. M. Reingold. 1991. "Graph Drawing by Force-Directed Placement." *Software: Practice and experience* 21, no. 11: 1129–1164. https://doi.org/10.1002/spe.4380211102.

Gartmeier, M., J. Bauer, H. Gruber, and H. Heid. 2008. "Negative Knowledge: Understanding Professional Learning and Expertise." *Vocations and Learning* 1, no. 2: 87–103. https://doi.org/10.1007/s12186-008-9006-1.

Garvin-Doxas, K., and M. W. Klymkowsky. 2008. "Understanding Randomness and Its Impact on Student Learning: Lessons Learned From Building the Biology Concept Inventory (BCI)." *CBE—Life Sciences Education* 7, no. 2: 227–233. https://doi.org/10.1187/cbe.07-08-0063.

Gopnik, A., and H. M. Wellman. 1994. "The Theory Theory." In *Mapping the Mind: Domain Specificity in Cognition and Culture*, edited by L. A. H. and S. A. G., 257–293. Cambridge University Press.

Göransson, A., D. Orraryd, D. Fiedler, and L. A. E. Tibell. 2020. "Conceptual Characterization of Threshold Concepts in Student Explanations of Evolution by Natural Selection and Effects of Item Context." *CBE—Life Sciences Education* 19, no. 1: ar1. https://doi.org/10.1187/cbe.19-03-0056.

Green, D. A., J. Loertscher, V. Minderhout, and J. E. Lewis. 2017. "For Want of a Better Word: Unlocking Threshold Concepts In Natural Sciences With a Key From the Humanities?" *Higher Education Research & Development* 36, no. 7: 1401–1417. https://doi.org/10.1080/07294360.2017.1325848.

Gregory, T. R. 2009. "Understanding Natural Selection: Essential Concepts and Common Misconceptions." *Evolution: Education and Outreach* 2, no. 2: 156–175. https://doi.org/10.1007/s12052-009-0128-1.

Großschedl, J., C. Konnemann, and N. Basel. 2014. "Pre-Service Biology Teachers' Acceptance of Evolutionary Theory and Their Preference for Its Teaching." *Evolution: Education and Outreach* 7, no. 1: 1–18. https://doi.org/10.1186/s12052-014-0018-z.

Großschedl, J., F. Seredszus, and U. Harms. 2018. "Angehende Biologielehrkräfte: evolutionsbezogenes Wissen und Akzeptanz der Evolutionstheorie [Pre-service Biology Teachers: Knowledge and Acceptance of Evolutionary Theory]." *Zeitschrift für Didaktik der Naturwissenschaften* 24, no. 1: 51–70. https://doi.org/10.1007/s40573-018-0072-0.

Gutowski, R., H. Aptyka, and J. Großschedl. 2023. "An Exploratory Study on Students' Denominations, Personal Religious Faith, Knowledge About, and Acceptance of Evolution." *Evolution: Education and Outreach* 16, no. 9: 1–17. https://doi.org/10.1186/s12052-023-00187-5.

Harms, U., and M. J. Reiss. 2019. *Evolution Education Re-Considered*. Springer. https://doi.org/10.1007/978-3-030-14698-6.

Harrison, A. G., D. J. Grayson, and D. F. Treagust. 1999. "Investigating a Grade 11 Student's Evolving Conceptions of Heat and Temperature." *Journal of Research in Science Teaching* 36, no. 1: 55–87. https://doi.org/10.1002/(SICI)1098-2736(199901)36:1<55::AID-TEA5>3.0.CO;2-P.

Hartelt, T., and H. Martens. 2024. "Influence of Self-Assessment and Conditional Metaconceptual Knowledge on Students' Self-Regulation of Intuitive and Scientific Conceptions of Evolution." *Journal of Research* in Science Teaching 61, no. 5: 1134–1180. https://doi.org/10.1002/tea.21938.

Hattie, J. 2008. *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement* (1st ed. Routledge). https://doi.org/10.4324/9780203887332.

Hayes, A. F., and J. J. Coutts. 2020. "Use Omega Rather Than Cronbach's Alpha for Estimating Reliability. But." *Communication Methods and Measures* 14, no. 1: 1–24. https://doi.org/10.1080/19312458.2020.1718629.

Högström, P., N. Gericke, J. Wallin, and E. Bergman. 2024. "Teaching Socioscientific Issues: A Systematic Review." *Science & Education*. https://doi.org/10.1007/s11191-024-00542-y.

Hommel, G. 1988. "A Stagewise Rejective Multiple Test Procedure Based on a Modified Bonferroni Test." *Biometrika* 75, no. 2: 383–386. https://doi.org/10.1093/biomet/75.2.383.

Hunter, D. R., M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris. 2008. "ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks." *Journal of Statistical Software* 24, no. 3: 1–29. https://doi.org/10.18637/jss.v024.i03.

Ionnides, C., and S. Vosniadou. 2002. "The Changing Meaning of Force." *Cognitive Science Quarterly* 2, no. 1: 5–61.

Kampourakis, K., and V. Zogza. 2009. "Preliminary Evolutionary Explanations: A Basic Framework for Conceptual Change and Explanatory Coherence in Evolution." *Science & Education* 18, no. 10: 1313–1340. https://doi.org/10.1007/s11191-008-9171-5.

Kelemen, D. 2011. "Teleological Minds: How Natural Intuitions About Agency and Purpose Influence Learning About Evolution." In *Evolution Challenges: Integrating Research and Practice in Teaching and Learning About Evolution*, edited by K. Rosengren, S. K. Brem, E. M. Evans, and G. M. Sinatra, (1st ed, 66–92. Oxford University Press). https://doi.org/10.1093/acprof:oso/9780199730421.003.0004.

Kilic, D. S. 2012. "Biyoloji öğretmen adaylarinin evrim öğretimi niyetleri [Pre-Service Biology Teacher's Intention to Teach Evolution]." *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi* 42: 250–261.

Klepsch, M., F. Schmitz, and T. Seufert. 2017. "Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load." *Frontiers in Psychology* 8, no. NOV: 1997. https://doi.org/10.3389/fpsyg.2017.01997.

Klepsch, M., and T. Seufert. 2020. "Understanding Instructional Design Effects by Differentiated Measurement of Intrinsic, Extraneous, and Germane Cognitive Load." *Instructional Science* 48, no. 1: 45–77. https://doi.org/10.1007/s11251-020-09502-9.

Koponen, I. T., and M. Nousiainen. 2014. "Concept Networks in Learning: Finding Key Concepts in Learners' Representations of the Interlinked Structure of Scientific Knowledge." *Journal of Complex Networks* 2, no. 2: 187–202. https://doi.org/10.1093/comnet/cnu003.

Koponen, I. T., and M. Pehkonen. 2010. "Coherent Knowledge Structures of Physics Represented as Concept Networks in Teacher Education." *Science & Education* 19, no. 3: 259–282. https://doi.org/10.1007/s11191-009-9200-z.

Koschützki, D., K. A. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl, and O. Zlotowski. 2005. "Centrality Indices." In *Network Analysis: Methodological Foundations*, edited by U. Brandes and T. Erlebach, 3418, 16–61. Springer. https://doi.org/10.1007/978-3-540-31955-9_3.

Kuschmierz, P., A. Beniermann, and D. Graf. 2020. "Development and Evaluation of the Knowledge About Evolution 2.0 Instrument (KAEVO 2.0)." *International Journal of Science Education* 42, no. 15: 2601–2629. https://doi.org/10.1080/09500693.2020.1822561.

Kuschmierz, P., A. Meneganzin, R. Pinxten, et al. 2020. "Towards Common Ground in Measuring Acceptance of Evolution and Knowledge About Evolution Across Europe: A Systematic Review of the State of Research." *Evolution: Education and Outreach* 13, no. 1: 1–18. https://doi.org/10.1186/s12052-020-00132-w.

Land, R., J. H. Meyer, and M. T. Flanagan. 2016. *Threshold Concepts in Practice* (1st ed. Sense Publishers). https://doi.org/10.1007/978-94-6300-512-8.

Landis, J. R., and G. G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33, no. 1: 159–174. https://doi.org/10.2307/2529310.

Leppink, J., and A. van den Heuvel. 2015. "The Evolution of Cognitive Load Theory and Its Application to Medical Education." *Perspectives on Medical Education* 4, no. 3: 119–127. https://doi.org/10.1007/s40037-015-0192-x.

Limón, M. 2001. "On the Cognitive Conflict as an Instructional Strategy for Conceptual Change: A Critical Appraisal." *Learning and Instruction* 11, no. 4–5: 357–380. https://doi.org/10.1016/S0959-4752(00)00037-2.

Luke, D. A. 2015. *A user's Guide to Network analysis in R (Vol. 72)*. Springer. https://doi.org/10.1007/978-3-319-23883-8.

Lüpsen, H. 2019. "Multiple Mittelwertvergleiche-parametrisch und nichtparametrisch-sowie α-Adjustierungen mit praktischen Anwendungen mit R und SPSS. Version 2.0 [Multiple Mittelwertvergleiche-parametrisch und nichtparametrisch-sowie α-Adjustierungen mit praktischen Anwendungen mit R und SPSS. Version 2.0]." https://kups.ub.uni-koeln.de/9566/1/mult-comp.pdf.

Manikas, M., I. Leonardos, K. Plakitsi, and K. Sotiropoulos. 2023. "The Evaluation of GAENE 2.1 in Greek Students." *Journal of Mathematics and Science Teacher* 3, no. 1: em031. https://doi.org/10.29333/mathsciteacher/13069.

Maxwell, S. E., H. D. Delaney, and K. Kelley. 2017. *Designing Experiments and Analyzing Data: A Model Comparison Perspective* (3rd ed. Routledge). https://doi.org/10.4324/9781315642956.

Mead, L. S., and E. C. Scott. 2010. "Problem Concepts in Evolution Part II: Cause and Chance." *Evolution: Education and Outreach* 3, no. 2: 261–264. https://doi.org/10.1007/s12052-010-0231-3.

Metzger, K. J., D. Montplaisir, D. Haines, and K. Nickodem. 2018. "Investigating Undergraduate Health Sciences Students' Acceptance of Evolution Using MATE and GAENE." *Evolution: Education and Outreach* 11, no. 1: 1–18. https://doi.org/10.1186/s12052-018-0084-8.

Meyer, J. H., and R. Land. 2003. "Threshold Concepts and Troublesome Knowledge: Linkages to Ways of Thinking and Practising Within the Disciplines." In *Improving Student Learning—Theory and Practice Ten Years on*, edited by C. Rust, 412–424. Oxford Brookes University.

Meyer, J. H., and R. Land. 2006a. *Overcoming Barriers to Student Understanding: Threshold Concepts and Troublesome Knowledge* (1st ed. Routledge). https://doi.org/10.4324/9780203966273.

Meyer, J. H., and R. Land. 2006b. "Threshold Concepts and Troublesome Knowledge: Issues of Liminality." In *Overcoming Barriers to Student Understanding*, edited by J. Meyer and R. Land, 19–32. Routledge.

Meyer, J. H., R. Land, and C. Baillie. 2010. *Threshold Concepts and transformational learning*. Brill. https://doi.org/10.1163/9789460912078.

Meyer, J. H. F., and R. Land. 2005. "Threshold Concepts and Troublesome Knowledge (2): Epistemological Considerations and a Conceptual Framework for Teaching and Learning." *Higher Education* 49, no. 3: 373–388. https://doi.org/10.1007/s10734-004-6779-5.

Ministry for Schools and Further Education of the State of North Rhine-Westphalia. 2022. "Kernlehrplan für die Sekundarstufe II Gymnasium/Gesamtschule in Nordrhein-Westfalen. Biologie [Core Curriculum for the Upper Secondary Level Gymnasium/Gesamtschule in North Rhine-Westphalia. Biology]." https://www.schulentwicklung.nrw.de/lehrplaene/lehrplan/325/gost_klp_bi_2022_06_07.pdf.

Moharreri, K., M. Ha, and R. H. Nehm. 2014. "Evograder: An Online Formative Assessment Tool for Automatically Evaluating Written

Evolutionary Explanations." *Evolution: Education and Outreach* 7, no. 1: 15. https://doi.org/10.1186/s12052-014-0015-2.

National Research Council. 2001. *Knowing What Students Know: The Science and Design of Educational Assessment*. National Academies Press.

National Research Council. 2012. *Thinking Evolutionarily: Evolution Education Across the Life Sciences: Summary of a Convocation*. National Academies Press. https://doi.org/10.17226/13403.

National Research Council. 2013. *Next Generation Science Standards: For States, by States*. National Academies Press. https://doi.org/10.17226/18290.

Nehm, R. H. 2018. "Evolution." In *Teaching Biology in Schools*, edited by K. Kampourakis and M. J. Reiss, (1st ed., 164–177. Routledge). https://doi.org/10.4324/9781315110158.

Nehm, R. H., E. P. Beggrow, J. E. Opfer, and M. Ha. 2012. "Reasoning About Natural Selection: Diagnosing Contextual Competency Using the Acorns Instrument." *American Biology Teacher* 74, no. 2: 92–98. https://doi.org/10.1525/abt.2012.74.2.6.

Nehm, R. H., S. J. Finch, and G. C. Sbeglia. 2022. "Is Active Learning Enough? The Contributions of Misconception-Focused Instruction and Active-Learning Dosage on Student Learning of Evolution." *BioScience* 72, no. 11: 1105–1117. https://doi.org/10.1093/biosci/biac073.

Nehm, R. H., and M. Ha. 2011. "Item Feature Effects in Evolution Assessment." *Journal of Research in Science Teaching* 48, no. 3: 237–256. https://doi.org/10.1002/tea.20400.

Nehm, R. H., M. Ha, M. Rector, et al. 2010. "Scoring Guide for the Open Response Instrument (ORI) and Evolutionary Gain and Loss Test (EGALT). (27 Apr 2023)." 1–40. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.471.7384&rep=rep1&type=pdf.

Nehm, R. H., and L. Reilly. 2007. "Biology Majors' Knowledge and Misconceptions of Natural Selection." *BioScience* 57, no. 3: 263–272. https://doi.org/10.1641/B570311.

Nehm, R. H., and I. S. Schonfeld. 2008. "Measuring Knowledge of Natural Selection: A Comparison of the Cins, an Open-Response Instrument, and An Oral Interview." *Journal of Research in Science Teaching* 45, no. 10: 1131–1160. https://doi.org/10.1002/tea.20251.

Nelson, C. E. 2008. "Teaching Evolution (and All of Biology) More Effectively: Strategies for Engagement, Critical Reasoning, and Confronting Misconceptions." *Integrative and Comparative Biology* 48, no. 2: 213–225. https://doi.org/10.1093/icb/icn027.

Neubrand, C. 2017. "Evolution unterrichten: Untersuchung von Aspekten erfolgreichen Lehrens und Lernens mit Fokus auf Beispielaufgaben und vorwissensangepassten Prompts [Teaching Evolution: Exploring Aspects of Successful Teaching and Learning With a Focus on Example Tasks and Pre-Knowledge Prompts] IPN-Leibniz Institute for Science and Mathematics Education]." Kiel, Germany.

Opfer, J. E., R. H. Nehm, and M. Ha. 2012. "Cognitive Foundations for Science Assessment Design: Knowing What Students Know About Evolution." *Journal of Research in Science Teaching* 49, no. 6: 744–777. https://doi.org/10.1002/tea.21028.

Oser, F. K., C. Näpflin, C. Hofer, and P. Aerni. 2012. "Towards a Theory of Negative Knowledge (NK): Almost-Mistakes as Drivers of Episodic Memory Amplification." In *Human Fallibility. Professional and Practice-Based Learning (Vol. 6)*, edited by J. Bauer and C. Harteis, 53–70. Springer. https://doi.org/10.1007/978-90-481-3941-5_4.

Özdemir, G., and D. B. Clark. 2007. "An Overview of Conceptual Change Theories." *Eurasia Journal of Mathematics, Science and Technology Education* 3, no. 4: 351–361. https://doi.org/10.12973/ejmste/75414.

Paas, F., A. Renkl, and J. Sweller. 2003. "Cognitive Load Theory and Instructional Design: Recent Developments." *Educational Psychologist* 38, no. 1: 1–4. https://doi.org/10.1207/S15326985EP3801_1.

Paçacı, Ç. 2022. "Effectiveness of Conceptual Change Strategies in Science Education: A Meta-Analysis Middle East Technical University]." https://hdl.handle.net/11511/98191.

Papo, D., J. M. Buldú, S. Boccaletti, and E. T. Bullmore. 2014. "Complex Network Theory and the Brain." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 369, no. 1653: 20130520. https://doi.org/10.1098/rstb.2013.0520.

Park, H. J. 2007. "Components of Conceptual Ecologies." *Research in Science Education* 37, no. 2: 217–237. https://doi.org/10.1007/s11165-006-9023-8.

Perkins, D. 2008. "Beyond Understanding." In *Threshold Concepts Within the Disciplines*, edited by R. Land, J. H. F. Meyer, and A. Smith, 3–19. Sense Publishers.

Posner, G. J., and K. A. Strike. 1992. *A Revisionist Theory of Conceptual Change*. State University of New York Press.

Posner, G. J., K. A. Strike, P. W. Hewson, and W. A. Gertzog. 1982. "Accommodation of a Scientific Conception: Toward a Theory of Conceptual Change." *Science Education* 66, no. 2: 211–227. https://doi.org/10.1002/sce.3730660207.

R Core Team. 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing https://www.R-project.org/.

Rachmatullah, A., R. H. Nehm, F. Roshayanti, and M. Ha. 2018. "Evolution Education in Indonesia: Pre-Service Biology Teachers' Knowledge, Reasoning Models, and Acceptance of Evolution." In *Evolution Education Around the Globe*, edited by H. Deniz and L. Borgerding, 335–355. Springer International Publishing. https://doi.org/10.1007/978-3-319-90939-4_18.

Ravinder, E. B., and A. Saraswathi. 2020. "Literature Review of Cronbach Alpha Coefficient (Α) and Mcdonald's Omega Coefficient (Ω)." *European Journal of Molecular & Clinical Medicine* 7, no. 6: 2943–2949.

Rea-Ramírez, M. A., and J. J. Clement. 1998. "In Search of Dissonance: The Evolution of Dissonance in Conceptual Change Theory." Proceedings of National Association for Research in Science Teaching (NARST), San Diego, CA.

Rector, M. A., R. H. Nehm, and D. Pearl. 2013. "Learning the Language of Evolution: Lexical Ambiguity and Word Meaning in Student Explanations." *Research in Science Education* 43, no. 3: 1107–1133. https://doi.org/10.1007/s11165-012-9296-z.

Richey, J. E., and T. J. Nokes-Malach. 2015. "Comparing Four Instructional Techniques for Promoting Robust Knowledge." *Educational Psychology Review* 27, no. 1: 181–218. https://doi.org/10.1007/s10648-014-9268-0.

Ross, P. M., C. E. Taylor, C. Hughes, et al. 2010. "Threshold Concepts: Challenging the Way We Think, Teach and Learn in Biology." In *Threshold Concepts and Transformational Learning*, edited by J. H. F. Meyer, R. Land, C. Baillie, 165–177. Sense Publishers. https://doi.org/10.1163/9789460912078_011.

Salwén, H. 2021. "Threshold Concepts, Obstacles or Scientific Dead Ends?" *Teaching in Higher Education* 26, no. 1: 36–49. https://doi.org/10.1080/13562517.2019.1632828.

Sbeglia, G. C., and R. H. Nehm. 2018. "Measuring Evolution Acceptance Using the GAENE: Influences of Gender, Race, Degree-Plan, and Instruction." *Evolution: Education and Outreach* 11, no. 18: 1–19. https://doi.org/10.1186/s12052-018-0091-9.

Schoenemann, B., and E. N. K. Clarkson. 2023. "The Median Eyes of Trilobites." *Scientific Reports* 13, no. 1: 3917. https://doi.org/10.1038/s41598-023-31089-7.

Schöps, S. 2021. Ganz einfache 10-Minuten-Übungen Stochastik: Wahrscheinlichkeitsrechnung leicht gemacht [Very Simple 10-minute Exercises in Stochastics: Probability Made Easy]. Auer.

Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany. 2020. Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife [Educational Standards in Biology for the General University Entrance Qualification]. Carl Link. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020/2020_06_18-BildungsstandardsAHR_Biologie.pdf.

Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany. 2022. "Bildungsstandards für das Fach Mathematik Erster Schulabschluss (ESA) und Mittlerer Schulabschluss (MSA) [Educational Standards for the Subject Mathematics First School Certificate (ESA) and Secondary School Certificate (MSA)]." https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2022/2022_06_23-Bista-ESA-MSA-Mathe.pdf.

Settlage, J. 2007. "Conceptions of Natural Selection: A Snapshot of the Sense-Making Process." *Journal of Research in Science Teaching* 31, no. 5: 449–457. https://doi.org/10.1002/tea.3660310503.

Siew, C. S. Q., D. U. Wulff, N. M. Beckage, and Y. N. Kenett. 2019. "Cognitive Network Science: A Review of Research on Cognition Through the Lens of Network Representations, Processes, and Dynamics." *Complexity* 2019: 1–24. https://doi.org/10.1155/2019/2108423.

Smith, C., S. Carey, and M. Wiser. 1985. "on Differentiation: A Case Study of the Development of the Concepts of Size, Weight, and Density." *Cognition* 21, no. 3: 177–237. https://doi.org/10.1016/0010-0277(85)90025-3.

Smith, M. U. 2010. "Current Status of Research in Teaching and Learning Evolution: II. Pedagogical Issues." *Science & Education* 19, no. 6–8: 539–571. https://doi.org/10.1007/s11191-009-9216-4.

Smith, M. U., S. W. Snyder, and R. S. Devereaux. 2016. "The GAENE—Generalized Acceptance of Evolution Evaluation: Development of a New Measure of Evolution Acceptance." *Journal of Research in Science Teaching* 53, no. 9: 1289–1315. https://doi.org/10.1002/tea.21328.

Solís-Pinilla, J., and C. Merino Rubilar. 2024. "Threshold Concepts in the Teaching of Natural Sciences: A Systematic Review." *Journal of Technology and Science Education* 14, no. 4: 947–964. https://doi.org/10.3926/jotse.2385.

Southerland, S. A., E. Abrams, C. L. Cummins, and J. Anzelmo. 2001. "Understanding Students' Explanations of Biological Phenomena: Conceptual Frameworks or p-Prims?" *Science Education* 85, no. 4: 328–348. https://doi.org/10.1002/sce.1013.

Statnet Development Team. 2003–2024. "statnet: Software Tools for the Statistical Modeling of Network Data." http://statnet.org.

Strike, K. A., and G. J. Posner. 1982. "Conceptual Change and Science Teaching." *European Journal of Science Education* 4, no. 3: 231–240. https://doi.org/10.1080/0140528820040302.

Taylor, C. 2006. "Threshold Concepts in Biology: Do They Fit the Definition?" In *Overcoming Barriers to Student Understanding*, edited by J. Meyer and R. Land, (1st ed., 87–99. Routledge).

Texas Education Agency. 2023. Chapter 112. Texas Essential Knowledge and Skills for Science. Subchapter C. High School. Texas Education Agency. https://tea.texas.gov/sites/default/files/ch112c.pdf.

Thagard, P. 1992. *Conceptual Revolutions*. Princeton University Press. https://www.jstor.org/stable/j.ctv36zq4g.

Tibell, L. A. E., and U. Harms. 2017. "Biological Principles and Threshold Concepts for Understanding Natural Selection: Implications for Developing Visualizations as a Pedagogic Tool." *Science & Education* 26, no. 7–9: 953–973. https://doi.org/10.1007/s11191-017-9935-x.

To, C., H. R. Tenenbaum, and H. Hogh. 2017. "Secondary School Students' Reasoning About Evolution." *Journal of Research in Science Teaching* 54, no. 2: 247–273. https://doi.org/10.1002/tea.21347.

Tulis, M., G. Steuer, and M. Dresel. 2016. "Learning From Errors: A Model of Individual Processes." *Frontline Learning Research* 4, no. 4: 12–26. https://doi.org/10.14786/flr.v4i2.168.

Vosniadou, S., and C. Ioannides. 1998. "From Conceptual Development to Science Education: A Psychological Point of View." *International Journal of Science Education* 20, no. 10: 1213–1230. https://doi.org/10.1080/0950069980201004.

Wagner, A. 2012. "The Role of Randomness in Darwinian Evolution." *Philosophy of Science* 79, no. 1: 95–119. https://doi.org/10.1086/663239.

Walck-Shannon, E., J. Batzli, J. Pultorak, and H. Boehmer. 2019. "Biological Variation as a Threshold Concept: Can We Measure Threshold Crossing?" *CBE—Life Sciences Education* 18, no. 3: ar36. https://doi.org/10.1187/cbe.18-12-0241.

Wellman, H. M., and S. A. Gelman. 1992. "Cognitive Development: Foundational Theories of Core Domains." *Annual Review of Psychology* 43: 337–375. https://doi.org/10.1146/annurev.ps.43.020192.002005.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis* (2nd ed. Springer International Publishing). https://doi.org/10.1007/978-3-319-24277-4.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.