RELEVANCE OF THRESHOLD CONCEPTS FOR UNDERSTANDING EVOLUTION

DOCTORAL THESIS

SUBMITTED WITH RESPECT TO THE REQUIREMENTS
FOR THE DOCTORAL DEGREE

Dr. rer. nat

TO THE FACULTY OF MATHEMATICS AND NATURAL SCIENCES
OF THE CHRISTIAN-ALBRECHTS UNIVERSITY AT KIEL

SUBMITTED BY
DANIELA FIEDLER

KIEL, APRIL 2018

First reviewer	Prof. Dr. Ute Harms
Second reviewer	Prof. Dr. Ross H. Nehm
Date of the defense	2018, June 28

The studies were supported by the Swedish Research Council [VR 2012:5344, LT].

SUMMARY

Evolutionary theory is the integrative framework of modern biology and learning its essential tenets is widely considered a necessary feature of scientific literacy. However, research indicates that teachers and students still struggle with teaching and learning evolution, respectively, and have various alternative conceptions. Current research also displays learning difficulties with those evolutionary concepts that are strongly related to abstract concepts like randomness and probability, so-called threshold concepts. Until now, valid tools that assess students' understanding of these threshold concepts to examine the relationships to knowledge and to the acceptance of evolution, as well as to investigate the effectiveness of educational strategies to support a conceptual knowledge of threshold concepts are lacking.

Four empirical studies have been conducted as part of this dissertation project. All four studies focus on students' conceptual knowledge of threshold concepts, particularly on the threshold concepts randomness and probability. Study 1 concentrates on the developmental process of two test instruments to measure students' conceptual knowledge of randomness and probability in an evolutionary and mathematical context (RaProEvo and RaProMath, respectively). In Study 2, the RaProEvo test was used to examine the effectiveness of the simulation software EvoSketch for teaching and learning random and probabilistic processes in evolution. Findings indicate that EvoSketch simulations are a useful tool for learning and teaching these concepts, particularly for fostering long-term understanding. Study 3 deals with the question to which extent conceptual knowledge of randomness and probability is related to knowledge and to acceptance of evolution. Results reveal moderate to strong relationships, while conceptual knowledge of randomness and probability also serves as explaining factor for knowledge and acceptance of evolution. In Study 4, the effect of item features for students' use of threshold concepts was investigated. Findings examine that students' use of threshold concepts in their written evolutionary explanations differs among the three investigated contexts, although no consistent pattern was found. Moreover, fine-grained analyses reveal interesting insights into the different expression of threshold concepts according to item features.

Overall, using qualitative and quantitative methods, the presented dissertation provides new insights into the existing body of work on evolution education by developing a more expansive view of understanding (and accepting) evolution that encompasses aspects of threshold concepts.

ZUSAMMENFASSUNG

Die Evolutionstheorie ist das vereinigende, übergreifende Erklärungsprinzip der Lebenswissenschaften. Das Erlernen der wesentlichen Aussagen der Evolutionstheorie ist deshalb ein unabdingbarer Teil naturwissenschaftlicher Bildung. Dennoch zeigen empirische Untersuchungen, dass Lehrkräfte und Lernende nicht nur Schwierigkeiten mit dem Lehren bzw. Lernen wesentlicher Evolutionsaspekte haben, sondern auch zahlreiche fachlich inadäquate Vorstellungen besitzen. Neue Forschungsergebnisse machen deutlich, dass insbesondere solche Aspekte der Evolution nicht verstanden werden, die mit abstrakten Konzepten wie Zufall oder Wahrscheinlichkeit verknüpft sind, sogenannten Schwellenkonzepten. Bisher gibt es jedoch keine zuverlässigen Testinstrumente, die das Wissen über diese Schwellenkonzepte erfassen. Mit diesen könnten aber nicht nur Zusammenhänge zwischen dem Evolutionswissen und der Akzeptanz der Evolution untersucht werden, sondern auch die Wirksamkeit von Instruktionsmaßnahmen zum Aufbau des konzeptuellen Wissens über diese Schwellenkonzepte.

Die vorliegende Dissertation umfasst vier empirische Studien. Alle vier Studien beschäftigen sich mit dem konzeptuellen Wissen über Schwellenkonzepte, insbesondere über die Konzepte Zufall und Wahrscheinlichkeit. Studie 1 beschreibt den Entwicklungsprozess zweier Testinstrumente zur Messbarmachung des konzeptuellen Wissens über Zufall und Wahrscheinlichkeit im evolutionären und mathematischen Kontext (RaProEvo bzw. RaProMath). In Studie 2 wurde der RaProEvo-Test eingesetzt, um die Effektivität einer Simulationssoftware (EvoSketch) zum Lehren und Lernen zufälliger und probabilistischer Prozesse im evolutionären Kontext zu untersuchen. Die Ergebnisse deuten darauf hin, dass das Lernen mit EvoSketch-Simulationen zu einem langfristigen Wissensaufbau beitragen kann. In Studie 3 wurde der Frage nachgegangen, inwiefern das konzeptuelle Wissen über Zufall und Wahrscheinlichkeit mit dem Evolutionswissen und der Akzeptanz der Evolution zusammenhängt. Die Ergebnisse zeigen nicht nur mittlere bis stark positive Zusammenhänge, sondern auch, dass das konzeptuelle Wissen über Zufall und Wahrscheinlichkeit ein bedeutender Erklärungsfaktor für das Evolutionswissen und die Akzeptanz der Evolution ist. In Studie 4 wurde untersucht, inwiefern Aufgabenmerkmale einen Einfluss auf die Verwendung von Schwellenkonzepten in geschriebenen Antworten haben. Die Ergebnisse weisen darauf hin, dass die Verwendung der Schwellenkonzepte in den Antworten unterschiedlich ausfällt, aber kein konsistentes Muster zu erkennen ist. Detaillierte Analysen zeigten jedoch interessante Ausprägungen bezüglich der Verwendung von Schwellenkonzepten in den einzelnen Aufgaben.

Insgesamt liefert die vorliegende Arbeit durch die Verwendung quantitativer und qualitativer Methoden neue Einblicke bezüglich der Bedeutung der Schwellenkonzepte Zufall und Wahrscheinlichkeit für das Lehren und Lernen der Evolution.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

The overarching goal of science instruction is that citizens develop an understanding of principles and concepts, so that they can apply this knowledge to resolve an extended range of problems in a variety of everyday situations such as interpreting media-described findings (e.g., National Research Council [NRC], 2012). Particular evolution education has a unique educational character since evolution education is designated to foster accurate mental models of the mechanisms of evolutionary theory, the overarching framework of the life sciences, and to introduce an appreciation of the centrality of this framework for a scientific understanding of the living world (American Association for the Advancement of Science, 2006). However, several empirical studies indicate that students across educational programs (e.g., Graf & Soran, 2011; Johannsen & Krüger, 2005; Nehm & Reilly, 2007; Yates & Marek, 2015) and teachers (e.g., Nadelson & Sinatra, 2009; Trani, 2004) still struggle with learning and teaching the tenants of evolutionary theory and have various alternative conceptions. There is a growing body of research that examines (a) specific learning and teaching difficulties (e.g., Nehm, Rector, & Ha, 2010; Sanders & Ngxola, 2009; Trani, 2004), and (b) educational support to successfully promote evolutionary knowledge acquisition (e.g., Basey et al., 2014; Eterovic & Santos, 2013; Neubrand, Borzikowsky, & Harms, 2016).

Currently, the notion of threshold concepts has become a novel focus in evolution education (Tibell & Harms, 2017). Research reveals that underlying abstract concepts such as randomness, probability, temporal scale, and spatial scale hinder the successful learning of evolutionary concepts (Cheek, 2013; Mead & Scott, 2010; Tibell & Harms, 2017). At this moment, the goal is to derive empirical evidence for the relevance of threshold concepts for fruitful teaching and learning of evolutionary concepts. However, until now, valid tools that assess students' understanding of threshold concepts to examine the relationships regarding knowledge and acceptance of evolution or to investigate the effectiveness of educational strategies to support a conceptual knowledge of threshold concepts are lacking.

The presented dissertation addresses these open research fields and provides new insights into the existing body of work on evolution education. The following chapter provides the theoretical and empirical background for the research conducted in this dissertation. The first section (Chapter 2.1) examines the relevance of evolution as a unifying theme and cognitive framework in biology education by focusing on principles and key concepts, students' alternative conceptions, and their acceptance of evolution. Following this (Chapter 2.2), the term threshold concept is clarified. Based on their relevance for evolution education, the two threshold concepts randomness and probability are explained in detail. Afterwards, the theoretical background for learning with visualizations is provided (Chapter 2.3) and will end with the rationale for using visualizations as a tool to foster students' conceptual knowledge (Chapter 2.4). The overarching aim of this dissertation with a short overview of the conducted studies is addressed in the next chapter (Chapter 3). The following chapters will subsequently present the research article (Chapter 4) and manuscripts (Chapter 5–7). Afterwards, a summary of the findings of each study is provided (Chapter 8). At last, the final chapter (Chapter 9) delivers an overall discussion and the general limitations of this dissertation. To sum up, implications for classroom practice and future research are given.

2 THEORETICAL BACKGROUND AND CURRENT STATE OF RESEARCH

2.1 Evolution and Evolutionary Theory

The term *evolution* refers to the process of changes over time in populations or taxa of organisms. Thus, the *theory of evolution* provides the explanations for similarities among organisms, biological diversity, and many features and processes of our world. To date, scientists propose natural selection, gene flow (migration), and genetic drift as primary processes through which evolution can take place (e.g., Heams, 2014; Lynch, 2007). Therefore, learning the essential tenets of these processes is widely considered a fundamental feature of scientific literacy in order to critically address numerous issues associated with students' environment and everyday life (Nationale Akademie der Wissenschaften Leopoldina [Leopoldina], 2017; NRC, 2012). In this dissertation, the predominant focus will be on the understanding of evolution through the process of natural selection, which is still challenging for students. Moreover, various alternative conceptions exist that make learning difficult (see Section 2.1.3).

However, before describing the principles and key concepts of evolution, it is fruitful to know how learning processes in individuals occur. The roots of theories explaining learning processes can be traced back to early Gestalt psychologists (J. R. Anderson, 2007). Nevertheless, research of Piaget (1929, 1930, 1974) and Kuhn (1962) led to a theory of learning processes in science classrooms. Based on the earlier theories and findings, Posner and colleagues (1982) described the conceptual change theory. This framework provides a more detailed basis for explaining learning processes, particularly taking into account that learners bring their own conceptions into science classrooms (Posner, Strike, Hewson, & Gertzog, 1982). This will be explained in detail in the following section.

2.1.1 Conceptual Change

The conceptual change theory considers learning as a rational activity and as a kind of inquiry to make judgments based on available evidence. Thus, learning is not considered as a simple "acquisition of a set of correct responses, a verbal repertoire or a set of behaviors" (Posner et al., 1982, p. 212), but rather as dealing with new phenomena in an existing (prior) conception or the actual replacement and reorganization of central concepts with new ones (Posner et al., 1982). Still, (fundamental) conceptual change mainly refers to the accommodation of new scientific concepts. Posner and colleagues (1982) proposed that learners have to be disappointed or unsatisfied with their existing conceptions before a conceptual change can occur. This means that a learner's everyday explanation is not sufficient to explain a new scientific phenomenon; hence another (accurate) concept is needed. The implementation of a new scientifically accurate concept must be initially plausible, understandable, and fruitful to the learner for a conceptual change to occur.

Although conceptual change is described as “a rational change in a person’s conceptual system” (p. 223), it is not supposed to be abrupt and rather involves going back and forth as well as many false starts and mistakes (Posner et al., 1982). An expansion of Posner’s conceptual change theory was the inclusion of Toulmin’s (1972) idea of conceptual ecology. In this context, a conceptual ecology includes fundamental conceptions that serve as controlling and modifying forces for the process of conceptual change to occur (Strike & Posner, 1992). Still, initial description of conceptual ecology includes anomalies, analogies, metaphors, epistemological commitments, metaphorical beliefs, and knowledge outside the field (Strike & Posner, 1992). But, conceptual change is not purely logically driven. In fact, a learner’s personal need to understand (new) natural phenomenon serves as a significant driving force for the process of conceptual change (Dole & Niederhauser, 1990). Thus, Demastes and colleagues (1995) applied the idea of conceptual ecology to the specific science content of evolutionary theory and proposed the following components as relevant for driving conceptual change within this context:

- prior conceptions (both scientific and alternative; see Section 2.1.2 and Section 2.1.3, but also Section 2.2),
- scientific orientation (the degree to which learners organize their life around scientific activities),
- view of nature of science,
- view of the biological world (in competitive, causal or aesthetics terms),
- religious orientation (the degree to which learners organize their life around religious activities), and
- acceptance of evolution (see Section 2.1.4).

More recently, the idea of conceptual ecology has also been applied to acceptance, with several authors refining the framework by including additional relevant components such as epistemological beliefs, motivations and emotions, reasoning level, and thinking disposition (Athanasiou & Papadopoulou, 2012; Deniz, Donnelly, & Yilmaz, 2008; Sinatra, Brem, & Evans, 2008; Vosniadou, 2007). In summary, conceptual ecology covers a variety of cognitive, affective, and contextual factors influencing the learning (and accepting) of evolutionary principles and key concepts.

2.1.2 Evolutionary Principles and Key Concepts

Understanding evolution through the process of natural selection includes the knowledge about various core and key concepts described in the literature (e.g., D. L. Anderson, Fisher, & Norman, 2002; Gregory, 2009; Mayr, 2001; Nehm & Schonfeld, 2007). Recently, Tibell and Harms (2017) used the term key concepts to refer to nine evolutionary content-oriented concepts (e.g., the origin of variation, reproduction or change within populations) that are organized into the three main evolutionary principles *variation*, *heredity*, and *selection* (cf. Godfrey-Smith, 2007). These principles and their related key concepts are shortly explained in the following sections.

Variation. Evolutionary change can only occur if there is a genetic variation among individuals that “may be manifested as morphological, physiological, or behavioral (phenotypic) differences” (Tibell & Harms, 2017, p. 956). The ultimate sources of genetic variations (*origin of variation*) are random mutations and the reshuffling of existing variation (i.e., genetic recombination, horizontal gene transfer, or recombination of genes during the process of sexual reproduction). Therefore, an *individual’s variation* is based on the set of genes (genotype) that build the foundation for an individual’s structure and behavior (phenotype), and thus, will impact organisms’ survival and reproduction in a particular environment (*differential fitness*) (Gregory, 2009; Mayr, 2001; Tibell & Harms, 2017).

Heredity. Individuals of a population *reproduce* and pass their heritable traits from parent to offspring (*inherited variation*; Tibell & Harms, 2017). Traits that confer advantages over others are more frequently passed on to the next generation and may accumulate through time (Mayr, 2001; Tibell & Harms, 2017). Without heredity, the process of natural selection cannot occur.

Selection. Although populations can exponentially increase in number in a most favorable environment, a large number of produced offspring does not survive to reproduce on their own (*limited survival*; Gregory, 2009). Thus, the effectiveness of existing genetic variation (and through this their phenotype) is determined by biotic and abiotic factors (*selection pressures*), resulting in different potentials for individuals to survive and reproduce (Tibell & Harms, 2017). Hence, the likelihood of survival and reproduction is higher for individuals with advantageous traits in a given environment than for individuals with disadvantageous traits in the same surrounding. Although this process often involves many generations, it will finally lead to *changes in populations* (e.g., favorable traits become more dominant). Over time, isolated populations of the same ancestral population may diverge sufficiently under different selection pressures and become different species (*speciation*; Mayr, 2001; Tibell & Harms, 2017).

2.1.3 Students’ Conceptions about Evolution

It is well known that students come into science classes with deep-rooted ideas for natural phenomena, which were developed as a result of their everyday life experiences to understand, explain, and predict the world (Coley & Tanner, 2012; Sinatra et al., 2008). Although these explanations work well in students’ everyday life, they differ from those explanations accepted by scientists and teachers and are therefore regarded as inaccurate. Accordingly, they are referred to as *misconceptions*, *alternative conceptions, naïve ideas*, *preconceptions,* and by other descriptors (e.g., Graf & Hamdorf, 2011; Leonard, Kalinowski, & Andrews, 2014; Maskiewicz & Lineback, 2013). Even though some differences regarding the terms’ characteristic (e.g., *misconceptions* implies some negative judgment) or their conceptual source (e.g., *preconceptions* are often used to describe students' pre-instructional conceptions) exist, they still share the definition of inaccurate conceptions or ideas based upon the context in which they are used. In this dissertation, I will use the term *alternative conceptions* to refer to scientifically inaccurate conceptions concerning biological evolution.

Alternative conceptions and scientifically accepted conceptions can coexist in a learner's mind (Fosnot & Perry, 2005; Palmer, 1999; Shtulman & Valcarcel, 2012). Thus, learners may maintain both the new (scientifically accepted) and the old (alternative) information in coexisting but different contexts or they construct a new conceptual framework that incorporates the new and the old knowledge.

Decades of empirical research reveal that students of all ages still enter school or university biology courses holding fundamental alternative conceptions of biological evolution (e.g., Baalmann, Frerichs, Weitzel, Gropengießer, & Kattmann, 2004; Beggrow & Nehm, 2012; Bishop & Anderson, 1990; Brumby, 1979; Evans, 2000; Ferrari & Chi, 1998; Graf & Soran, 2011; Heddy & Sinatra, 2013; Kampourakis & Zogza, 2007, 2009; Nehm & Reilly, 2007; Nehm & Ridgway, 2011; Palmer, 1999; Rector, Nehm, & Pearl, 2013; Shtulman, 2006; Spindler & Doherty, 2009; To, Tenenbaum, & Hogh, 2017; Weitzel & Gropengiesser, 2009; Yates & Marek, 2014; Zabel & Gropengiesser, 2011). Some of these alternative conceptions are proposed to be originated from a learner's informal, intuitive way of thinking, also referred to as *cognitive construal*, *cognitive constraint* or *cognitive bias* (Coley & Tanner, 2012, 2015; Evans, 2001; Evans, Rosengren, Lane, & Price, 2012).

In his research article, Gregory (2009) collected and explained the eight most common alternative conceptions that educators from schools and universities will face in biology courses. In the following sections, I will explain four of these alternative conceptions that are often measured in research studies.

Teleology (Need). Human minds are biased towards causal explanations to make sense of many aspects of the world around us (i.e., the need to answer the question why; Kahneman, 2012). Both children and adults tend to explain an event by referring to the consequence of this event (Kelemen, 1999, 2012; Southerland, Abrams, Cummins, & Anzelmo, 2001). In other words, students' faulty reasoning is based on the assumption of a supposed goal, purpose, or function.

Anthropomorphism (or intentionality). When humans need to explain unfamiliar biological species or even processes, they tend to talk about these species and processes by describing an analogy to humans (Coley & Tanner, 2012). In the context of evolution, students ascribe a human-like conscious intent either to the objects of natural selection (i.e., individual organisms evolve in response to a changed environment) or to the process itself (i.e., natural selection or nature itself acts as conscious agent; Kampourakis & Zogza, 2008; Sinatra et al., 2008). Although anthropomorphism is closely related to teleology, this alternative conception stands on its own.

Essentialism. The last intuitive way of thinking that leads humans to faulty reasoning is the tendency to believe that things or organisms belong to categories. This is another way of human thinking to explain and predict an otherwise incomprehensible complex world (Coley & Tanner, 2012; Sinatra et al., 2008): To think that organisms belong to a category means, that members of such a category are conceived to share underlying properties or an essence, while variation among individuals is falsely recognized as anomalous and mostly unimportant deviations (Coley & Tanner, 2012; Gregory, 2009).

Lamarckian (use and disuse). The “Lamarckian conception” is named after its developer Jean-Baptiste Lamarck and is based on his explanation for how evolution occurs. It refers to students’ tendency to describe evolutionary processes as changes of individual organs or traits due to their use or disuse (Gregory, 2009). This alternative conception is not based on a particular intuitive way of thinking. Nevertheless, in some studies, the Lamarckian conception is mistakenly mixed with teleology and need attributions (for discussion on this topic, see Kampourakis & Zogza, 2007).

2.1.4 Acceptance of Evolution

Although evolution is the unifying theme of modern biology, many people, including university students and biology teachers, not only lack an understanding of the processes of evolution, but also refuse to accept evolutionary theory as the best scientific explanation for similarities among organisms, biological diversity, and various features and processes of our world (Berkman & Plutzer, 2011; Miller, Scott, & Okamoto, 2006). Individuals’ insufficient knowledge and alternative conceptions (see Section 2.1.2), beliefs (e.g., religiosity), or personality (e.g., thinking disposition) are only some factors to explain this phenomenon (see also Section 2.1.1).

However, a growing carefulness emerged in science education for drawing distinctions between a learner’s *beliefs* in a specific construct and his or her *acceptance* of this construct (e.g., Cohen, 1995; Deniz et al., 2008; Nadelson & Southerland, 2012; Sinatra, Southerland, McConaughy, & Demastes, 2003; Smith & Siegel, 2004, 2016). Particularly, Smith and Siegel (2016) argue that belief is a mental state of having some opinion or faith regarding a construct. Thus, belief cannot be produced at will, carries no conceptual implication about reasoning, and is a matter of degree. In contrast, acceptance is a mental act based on the plausibility of a construct, their validity examination, and the richness of empirical support. Therefore, acceptance is regarded to be more under the voluntary control of an individual, and thus, is rather a decision than a matter of degree. For the sake of clarity, in this dissertation the term acceptance will be used to define learners’ acceptance of evolutionary theory as the best valid scientific explanation based on available evidence.

To date, there is a growing body of theoretical and empirical research on the relationship between understanding and accepting evolution, but a consistent description remains elusive. Some researchers revealed a positive relationship (e.g., Akyol, Tekkaya, Sungur, & Traynor, 2012; Deniz et al., 2008; Dunk, Petto, Wiles, & Campbell, 2017; Gibson & Hoefnagels, 2015; Großschedl, Konnemann, & Basel, 2014; Nadelson & Sinatra, 2009; Peker, Comert, & Kence, 2010), while others documented little to no relationship between acceptance and understanding of evolution (e.g., Demastes et al., 1995; Lord & Marino, 1993; Sinatra et al., 2003).

A simplified theoretical model of Smith and Siegel (2016) described the relationships among knowledge, understanding, beliefs, and acceptance. Based on this model, knowledge of evolutionary principles and key concepts promote or should lead to an understanding of evolution. In turn, understanding promotes acceptance and belief but may lead only to one or the other (if at all). Lastly, belief may promote acceptance (and vice versa), assuming that acceptance (or belief) has not occurred yet. Still, this model is similar to the first description of conceptual change (see Section 2.1.1), and thus, is mostly based on cognitive factors. But in fact, there are other factors that are suggested to influence acceptance (e.g., *feeling of certainty*; Ha, Haury, & Nehm, 2012). In addition, the understanding of underlying threshold concepts could also be another relevant but neglected factor for evolution education. Indeed, the understanding of threshold concepts is regarded as a necessary factor for understanding evolution (Tibell & Harms, 2017; see also Section 2.2), and therefore this may also influence the acceptance of evolution.

2.1.5 Evolution in Curricula

Educational standards are written learning goals of students' competencies and skills regarding specific school subjects like biology. New frameworks of science standards have introduced the notion of disciplinary core ideas (German *Basiskonzepte*) that weave across learning contexts to support a continual integration of knowledge and abilities over multiple years (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005; NRC, 2012). In the life sciences, evolution is the most important organizing theme since "all organisms are related by evolution and […] evolutionary processes have led to the tremendous diversity of the biosphere" (NCR, 2013, p. 139).

In the German biology standards for middle schools (KMK, 2005), aspects of evolution are mentioned throughout the three core ideas: *systems* (e.g., living systems are characterized by genetic and environmental variation and the opportunity for individual and evolutionary development), *structure and function* (e.g., adaptation of organisms to their environment is the result of evolutionary development of structure and function), and particularly *development* (e.g., mutation and selection are causes of intraspecific and phylogenetic development). The standards for high-school graduation in biology (KMK, 2004) list three out of eight core ideas with an explicit connection to the aspects of evolution (i.e., *reproduction*, *variability, and adaptation*; and *history and relatedness*). National standards in other countries similarly consider aspects of evolution as core ideas. For instance, the Next Generation Science Standards of the United States of America (NGSS, 2013) lists *biological evolution: unity and diversity* as a core idea of life sciences, while other aspects of evolution are implemented in the core idea *heredity: inheritance and variation of traits*. The NGSS (2013) illustrates a framework to implement these core ideas into states' curricula from elementary to high school.

However, educational standards are written goals that reflect what students should know and should be able to do, but they do not dictate how specific topics like evolutionary theory have to be implemented into states' curricula nor the manner or methods used to teach these topics (NGSS, 2013). Therefore, states' curricula could be quite diverse (for an overview of standards' implementation in middle schools in the federal states of Germany or the United States, see Fenner, 2013 and Vazquez, 2017, respectively). In Germany, only the federal states of Lower Saxony (Niedersächsisches Kultusministerium, 2009, 2015, 2017) and Schleswig-Holstein (Ministerium für Schule und Berufsbildung des Landes Schleswig-Holstein, 2016) implemented evolutionary theory in their state curriculum as an integrative framework of biology to date.

2.2 Threshold Concepts

More than a decade ago, Meyer and Land (2003) used the term threshold concepts in their report to the project *Enhance Teaching and Learning Environments in Undergraduate Courses (ETL),* which was undertaken in the United Kingdom. Threshold concepts are defined as portals "opening up a new and previously inaccessible way of thinking about something" (Meyer & Land, 2003, p. 1). In this report, an understanding of specific threshold concepts is proposed to transform students' perspectives and lead them to see things through a different lens. Further, threshold concepts are distinguished from key or core concepts as they are more than simple conceptual building blocks towards an understanding within a discipline (Meyer & Land, 2003, 2006). Although the exact nature of threshold concepts is still under review, Meyer and Land (2003, 2006) identify five initial characteristics:

- *Transformative* – once a threshold concept is understood, this will change the way in which students perceive and practice aspects of their discipline or subject.
- *Irreversible* – once particular threshold concepts are mastered, they are unlikely to be forgotten or to become unlearned.
- *Integrative* – threshold concepts can expose previously hidden connections (e.g., between isolated concepts or pieces of knowledge).
- *(Disciplinary) bounded* – threshold concepts have the potential to help identifying boundaries or frontiers of a subject, discipline, or academic territory.
- *Troublesome* – mastering threshold concepts means to deal with conceptually tricky, counter-intuitive, or alien knowledge. In this sense, threshold concepts are also connected to deeply rooted and difficult-to-change alternative conceptions.

Through the years, three more characteristics were added to address the critiques that threshold concepts are difficult to differentiate from other educational concepts (Land, 2011; Taylor, 2006):

- *Reconstitutive* – understanding threshold concepts may require a reconfiguration of learners' prior knowledge and a 'letting go' of alternative conceptions.
- *Discursive* – a shift in perspectives or crossing thresholds results in the use of enhanced or extended language.
- *Liminality* – the understanding of threshold concepts is not simple and often involves chaotic progress back, forth, and across conceptual terrains.

Despite the fact that the aspects mentioned above can help in finding, identifying, and understanding threshold concepts, there is still an ongoing debate about which characteristics are relevant and how many aspects a concept needs to possess to be regarded as threshold concept (Barradell, 2013). Even though some threshold concepts are likely defined by many but not necessarily all of the above-mentioned characteristics, it is said that they have to be – at least – *transformative* and involve crossing through *liminal space* (Meyer, Land, & Baillie, 2010; Taylor, 2006).

Over the last years, threshold concepts have been examined in diverse disciplines, including economics (e.g., Davies & Mangan, 2007; Karunaratne, Breyer, & Wood, 2016), philosophy (e.g., Booth, 2006), biosciences (e.g., Batzli, Knight, Hartley, Maskiewicz, & Desy, 2016; Taylor, 2006), chemistry (e.g., Park & Light, 2009), physics (e.g., Ferreira, Lemmer, & Gunstone, 2017), and computer science (e.g., Zander et al., 2008). The next section focuses on particularly threshold concepts that emerged in biology with a closer look at the topic of evolution.

2.2.1 Threshold Concepts in Biology (and Evolution)

Although threshold concepts often arise from troublesome or difficult content knowledge, not all content and concept areas with difficulties for teachers and learners are necessarily also threshold concepts (Ross, Taylor, Hughes, Kofod, et al., 2010). Moreover, it is suggested that identifying threshold concepts requires a look beyond the particular content to determine the concepts that operate in a broad integrating way (Perkins, 2006; Ross, Taylor, Hughes, Kofod, et al., 2010). Still, a problem for identifying threshold concepts is that they are rarely being made explicit (Davies, 2006). At least, Ross and colleagues (2010) identified a list of potential threshold concepts in biology including "energy, transformations, variation, probability and randomness, proportional reasoning (surface area to volume ration), predictive reasoning (hypothesis and null hypothesis testing), thinking at the subcellular level and integrating these observations with the macroscopic, temporal and spatial scales, and equilibrium" (p. 169).

Even though evolution is widely considered as troublesome to learn and teach, evolution itself is not suggested to be a threshold concept, but instead consists of a complex network of interconnected threshold concepts such as temporal scale, spatial scale, probability, and randomness (Ross, Taylor, Hughes, Whitaker, et al., 2010; Tibell & Harms, 2017). Additionally, variability, inheritance, and reproductive success are also proposed to be threshold concepts (Ross, Taylor, Hughes, Whitaker, et al., 2010). But Tibell and Harms (2017) argue that these concepts can be explained by underlying probabilistic or random processes, and thus should not be regarded as threshold concepts.

Moreover, Tibell and Harms (2017) recently developed a two-dimensional framework connecting principles and key concepts of evolution (Section 2.1.2) with the abovementioned general abstract concepts (randomness, probability, spatial scale, and temporal scale). They proposed that a complete understanding of evolution through natural selection requires the development of knowledge concerning both principles of evolution and general abstract concepts, and furthermore, the ability to freely navigate through this two-dimensional framework. In addition, they stated that an understanding of threshold concepts - as often chaotic and gradual - indicates similarities to the framework of conceptual change (see Section 2.1.1), and thus integrates threshold concepts into the conceptual change framework. Hence, a conceptual knowledge of specific threshold concepts is a prerequisite for changing alternative conceptions to scientifically sophisticated ones when thinking about evolution through natural selection (Tibell & Harms, 2017).

Although Tibell and Harms (2017) propose four threshold concepts to be relevant for evolution through natural selection, the focus of this dissertation will be mainly on the threshold concepts of randomness and probability, which are explained in the next section.

2.2.2 Randomness and Probability

It is stated that students need firm grounding in key statistical concepts such as sampling, probability, distribution, randomness, and uncertainty (Garfield, 2003). A clear understanding of randomness and probability is particularly important for understanding both evolution and molecular/cellular biology (Kærn, Elston, Blake, & Collins, 2005; Lenormand, Roze, & Rousset, 2009; Mead & Scott, 2010). But randomness is often difficult to understand because of its counter-intuitive description and the different meanings in diverse contexts.

In ordinary or everyday language, the term *random* is generally used to explain that a corresponding phenomenon is without order, predictability or pattern (Bennett, 1998; Wagner, 2012). Hence, random processes are purposeless and directionless (Mead & Scott, 2010). Moreover, an event is also referred to as random, if it occurs very rarely or if the occurrence is experienced as rather unusual (Büchter, Hußmann, Leuders, & Prediger, 2005). For instance, meeting a friend on an airplane to an island that you have not seen in ten years is expected to be random. Humans cannot *see* a causal explanation to such a phenomenon, but human minds are constructed to make sense of the world around us (Coley & Tanner, 2012). Moreover, this common perception of randomness or 'chance occurrences' does not change with increasing age (Falk & Konold, 1997; Kattmann, 2015), which hinders an understanding of the concept of randomness (and the closely related concept of stochasticity) in scientific disciplines including mathematics (Kaplan, Rogness, & Fisher, 2014) and biology (Mead & Scott, 2010).

Nevertheless, scientists and mathematicians use the term random to suggest an unpredictability (also given our current state of knowledge), but they do not mean to refer to purposelessness (Buiatti & Longo, 2013; Mead & Scott, 2010). Characteristics of random phenomena can be described as (1) in a given situation there is more than one possible outcome, and (2) the actual outcome that will occur is unpredictable (Batanero, Green, & Serrano, 1998; Kuzmak & Gelman, 1986). Therefore, mathematical models are applied to such situations to describe and understand it (Buiatti & Longo, 2013).

The evolutionary notion of *randomness* is often quite specific referring to events (e.g., mutations, genetic drift) that are independent of an organism's need and of the directionality provided by natural selection in the process of adaptation (Eble, 1999; Mead & Scott, 2010; Millstein, 2000). Random processes occur at every organizational level of the biological world (i.e., from the genetic level to the scale of populations including clade diversification and extinction; Cai, Friedman, & Xie, 2006; Lenormand et al., 2009; Raup, Gould, Schopf, & Simberloff, 1973). While random processes on individual level often refer to the process of genetic drift, population scales are often explained by random environmental changes such as frost, fire, or volcano eruption (Lenormand et al., 2009). Still, one of the most relevant factors for evolution through natural selection to occur are random genetic mutations. Mutations are called to be random, because of two different assumptions. At first, randomness is not meant in the way that all kinds of mutations are equally likely. In fact, mutations occur with statistical probability for each gene (mutation rates; Kattmann, 2015), and transition mutations are typically twice as frequent as transversion mutations (Li, 1997). However, it cannot be predicted precisely *where* and *when* a mutation will appear at a particular nucleotide site and for which generation – at least not with our current knowledge (Heams, 2014; Kattmann, 2015; Mead & Scott, 2010). This is why the occurrence of a mutation is best described as random. Further, mutations are also called random because they occur independently from their phenotypic effects. Hence, they are not directed to individuals' adaptation, and they do not occur more frequently when they are advantageous (Lenski & Mittler, 1993; Sniegowski, Gerrish, Johnson, & Shaver, 2000; Wagner, 2012). Thus, to summarize, randomness refers to both the process and the outcome of single events.

In contrast, *probability* is the likelihood of a particular outcome in the long run (over multiple events). It is assigned a numerical value between zero and one (Feller, 1968). The closer a probability value is to one, the more likely the outcome is. Students in evolution education have to deal with probabilities and probabilistic equations such as Punnett-Square diagrams or Hardy-Weinberg-Equilibrium. Thus, students are required to understand probabilities and that independent probabilities should be combined by multiplication (the AND-rule in mathematics; Masel, 2012). Besides these clear statistical models, natural selection itself can also be described as a probabilistic process because the likelihood of an organism to survive and to produce more offspring depends on the hereditary traits in relation to the surrounding environment (Buiatti & Longo, 2013; Mayr, 2001). Hence, the process of selection can be defined as the probabilities of individuals with differing traits in a given population surviving and reproducing in a specific environment (Tibell & Harms, 2017). Nevertheless, evolution through natural selection depends on random genetic mutations leading to heritable variation on which the probabilistic process of selection can act upon (Andrews et al., 2012; Mix & Masel, 2014).

Still, research indicates that students tend to struggle with both probability and the notion of randomness in the evolutionary context (Brumby, 1979; Deadman & Kelly, 1978; Garvin-Doxas & Klymkowsky, 2008; Robson & Burns, 2011). Moreover, biology students not only struggle to grasp the importance and role of randomness and probability in evolutionary theory (Gregory, 2009), but often have a weak understanding of mathematics (Hester, Buxner, Elfring, & Nagy, 2014; Jungck, 1997). This clearly hinders the teaching and learning of evolution because mathematical descriptions of randomness and probability are key elements of the explanations of random and probabilistic evolutionary (and other) biological processes (Buiatti & Longo, 2013; Wagner, 2012).

Particularly the process of randomness and the connection to the process of natural selection is often counter-intuitive for students (Tibell & Harms, 2017). Moreover, the processes of genetic random mutations that occur on scale levels are often not visible to the naked human eye. To make these processes visible through – for example – visualizations may help students in understanding the importance of random and probabilistic processes in the context of evolution.

2.3 Visualizations in Science Education

“Visualizations are an essential element of teaching, understanding, and creating scientific ideas” (Tversky, 2005, p. 40). They belong to a large class of cognitive tools crafted by people from all cultures and all eras to remember, reason, discover, and communicate (Tversky, 2005). Even though there are some ambiguities with the term *visualizations* and its relationships to other concepts like representations or models, the term is mainly used to describe two types of visualizations: *internal* and *external* visualizations (Gilbert, 2005; Gobert, 2005; Rapp & Kurby, 2008).

Internal visualizations refer to the act of forming an internal mental picture or construct of the external world, also identified as mental models (Craik & Lockhart, 1972; Gobert, 2005). These mental models are only available to the individual learner and cannot be captured directly by others. Further, they tend to be more like a piecemeal and incomplete version of (scientific) concepts, processes or structures (Franco & Colinvaux, 2000; Rapp & Kurby, 2008). In contrast, *external visualizations* describe the act of making something visible to the human eye. This includes visual materials such as pictures, diagrams, models, or simulations that present data or scientific concepts and processes the data in a novel way (Gobert, 2005; Rapp & Kurby, 2008; Schnotz, 2002).

These external visualizations are physically available to others and allow the interaction with the domain-specific content.

However, there are some ambiguities concerning the term visualization and their references to verbal and nonverbal entities (Mayer, 2011; Ryoo & Linn, 2012; Tibell & Rundgren, 2010). In this dissertation, the term visualizations will be used as an explanation for external visual displays. Further, whenever I refer to internal visualizations, the term *internal* is explicitly added.

2.3.1 Learning with Visualizations

Learning with text-picture-combinations is assumed to be more efficient than learning with texts alone (e.g., Adadan, Irving, & Trundle, 2009; Butcher, 2014; Mayer, 2014). Moreover, pictures, whether static or dynamic, seem to facilitate content storage and retrieval under certain conditions (Large, 1996). Over the last decades, a substantial body of theories explained how visual (static or dynamic) materials can enhance learning. The first theory that emerged to describe effective learning with visualizations is the *dual coding theory* (Clark & Paivio, 1991; Paivio, 1986). According to this, verbal and visual (nonverbal/pictorial) information are processed and stored in separated, but connected cognitive systems. Words or sentences are usually handled in the verbal system, while visualizations are encoded in both the nonverbal and verbal system. Furthermore, information that is encoded in both systems will be remembered more easily than information encoded in only one of the systems (Paivio, 1986).

This theory of dual coding was then used to develop the *cognitive theory of multimedia learning* (Mayer, 1997, 2014), and the more detailed *integrative model of text and picture comprehension* (Schnotz, 2014; Schnotz & Bannert, 2003). One of the central assumptions of both models is that information of texts and visualizations are encoded depending on their verbal and nonverbal modality resulting in parallel constructions of verbal or nonverbal mental models in two different cognitive systems inside the working memory. In a final step, these verbal and nonverbal mental models are mapped onto each other to build connections. Further, learners' prior knowledge influences this selection and organization processes of verbal and nonverbal information.

Another relevant theory for learning with visualizations, although not restricted to this topic, is the *cognitive load theory* (Chandler & Sweller, 1991; Paas & Sweller, 2014; Sweller, 1988). The underlying assumption is that the cognitive capacity of the human working memory is limited. Therefore, learning is likely to be hindered when learning tasks require too much capacity. The cognitive load theory distinguishes three types of cognitive load: (a) intrinsic cognitive load that is related to the experienced difficulty of the material in connection with a learner's prior knowledge (cannot be changed by instructional treatments); (b) extraneous cognitive load that is stimulated by the unnecessary processing of the presented instructional material (can be influenced by instructional design); and (c) germane cognitive load that refers to the mental effort of the learning process itself, hence the meaning making of the presented problem and the construction of schemas. Based on this, instructional materials should be aligned with learners' prior knowledge (intrinsic load), avoid unnecessary and confusing information (extraneous load), and stimulate mental processes to build up a conceptually rich and profound knowledge (germane load; de Jong, 2010).

2.3.2 Dynamic Visualizations

Scientific concepts, processes or structures are often difficult to perceive in the everyday world due to their spatial and temporal scales (Gobert, 2000; Rapp, 2005). Especially processes and structures that exist *beyond* the human visible scale (i.e., not directly seen through the human eye) can only be understood on imaginary levels (Lakoff, 1987; Lakoff & Johnson, 1980). For instance, changes in genes (i.e., mutations) are not visible to the human eye, even though technical opportunities like DNA sequencing techniques can make these changes visible. Nevertheless, such experiments are often time-consuming and rather expensive for ordinary school science classes (Euler, Schüttler, & Hausamann, 2015; Scharfenberg, 2005). Furthermore, most of the schools cannot provide this kind of lab work due to the lack of technical equipment.

Dynamic visualizations can overcome these artificial (school) limitations as well as the limitations of natural systems by making unobservable scientific phenomena visible (Ainsworth & VanLabeke, 2004; H.-Y. Chang & Linn, 2013). The advantageous feature of dynamic visualizations is the temporal structure (Ploetzner & Lowe, 2012; Ploetzner, Lowe, & Schlag, 2013). Learning with static visualizations often means that learners have to mentally infer processes that change over time, while dynamic visualizations can in fact show these changes continuously (Ainsworth & VanLabeke, 2004; Hegarty, 1992; Tversky, Morrison, & Betrancourt, 2002). Although the literature reveals mixed results for the effectiveness of static over dynamic visualizations (Hegarty, 2004; Lowe, 2003; Moreno & Valdez, 2005), there are growing hints regarding the advantages of dynamic visualizations (Höffler & Leutner, 2007; Pfeiffer, Scheiter, & Gemballa, 2012; Ryoo & Linn, 2012; Yang, Andre, Greenbowe, & Tibell, 2003).

Moreover, computer simulations may serve as a practical visualization tool to foster students' understanding of (dynamic) scientific concepts like evolution. According to Merriam-Webster's Collegiate Dictionary (2018), a simulation is "the imitative representation of the functioning of one system or process by means of the functioning of another" or an "examination of a problem often not subject to direct experimentation by means of a simulating device". De Jong and van Joolingen (1998) simplified this definition in their research article and described a computer simulation as "a program that contains a model of a system (natural or artificial; e.g., equipment) or a process" (p. 180).

Simulations have the advantage to be interactive and to allow learners to develop their knowledge through the interaction with the simulated (realistic-like) environment (Ainsworth & VanLabeke, 2004; van Berkum & de Jong, 1991). Learners can explore hypothetical situations of events while changing parameters and time-scales, and through this, observe the effects of changes (Ploetzner & Lowe, 2004; van Berkum & de Jong, 1991; Yaman, Nerdel, & Bayrhuber, 2008). Especially simulations of invisible phenomena indicated substantial learning effects (Marbach-Ad, Rotbain, & Stavy, 2008; Trey & Khan, 2008; Yarden & Yarden, 2010). Simulations are rarely sufficient to improve science learning by themselves, but succeed when combined with practical activities and instructional support (H.-Y. Chang & Linn, 2013; Eckhardt, Urhahne, Conrad, & Harms, 2013; Ryoo & Linn, 2012; Yaman et al., 2008).

2.3.3 Instructional Support for Learning with Simulations

Research indicates that problems in simulation-based science learning may be generated due to (1) the learner's prior domain-specific background knowledge (e.g., Ploetzner & Lowe, 2012), (2) design problems of the visualization so that learners are less likely to know what to attend (Rapp, 2005) or (3) learners' intrinsic problems with discovery learning itself, hence generating new hypotheses, designing experiments or interpreting data (de Jong & van Joolingen, 1998). Therefore, it is supposed that learners need instructional support to overcome these learning difficulties and to improve simulation-based learning outcomes (de Jong & van Joolingen, 1998; Kombartzky, Ploetzner, Schlag, & Metz, 2010; Urhahne & Harms, 2006). Zhang and colleagues (2004) distinguished between interpretative, experimental, and reflective support based on the time and methods to help learners in their simulation-based discovery process. In the following sections, I will explain these three types of instructional support.

Interpretative support. Interpretative support is mainly provided before the interaction to scaffold learners' awareness about the meaningfulness of the discovery process and to activate their prior knowledge, and to generate appropriate hypotheses. Learners tend to have a lack of knowledge regarding the hypothesis structure or they are unable to adapt hypotheses based on gathered data (de Jong & van Joolingen, 1998). However, to generate appropriate hypotheses and to construct a coherent understanding of the context, learners need to activate their prior knowledge (de Jong & van Joolingen, 1998). An effective way of interpretative support is to provide access to domain-specific background information (de Jong & van Joolingen, 1998; Reid, Zhang, & Chen, 2003). Still, studies indicate that the timing of providing this information is a critical aspect (Lazonder, Hagemans, & de Jong, 2010). Rather than being presented only beforehand, information should better be accessible through the entire discovery process. Another way to provide effective interpretative support is the use of worked examples. In simulation-based learning, worked examples have shown positive effects on learning outcomes and learners' situational interest (Spanjers, Wouters, van Gog, & van Merriënboer, 2011; Yaman et al., 2008). A typical worked example consists of a problem followed by the worked-out solution that is normally presented to the learner in a step-by-step format (Renkl, 2005). They support learners' knowledge acquisition as well as problem-solving competencies. Additionally, helping learners in regulating or structuring their learning process before they start with the simulation is also regarded as interpretative support. This can be done by providing concrete assignments, exercises or questions that guide learners to conduct their experiment (de Jong et al., 1999; Vreman-de Olde & de Jong, 2006).

Experimental support. Experimental support is allocated with a simulation during the interaction, and scaffolds learners' process of scientific discovery by designing verifiable experiments and drawing valuable conclusions. Learners often possess inefficient experimentation behaviors: they manipulate variables that had nothing to do with their tested hypotheses or vary too many variables at the same time (de Jong & van Joolingen, 1998). Further, students show a tendency to search for evidence that confirms their current hypothesis instead of stating an alternative hypothesis. Effective experimental supports are gradual and cumulative introductions to handle the simulation, explanations of essential parameters located in the simulation, or requesting learners to predict, describe, and interpret the outcome of the simulated experiment (Urhahne & Harms, 2006; Wang, Wu, & Hsu, 2017). Particularly learners with a low ability and inefficient discovery learning strategies are supported by structural guidance or experimental prompts (H.-Y. Chang, 2017; Veenman & Elshout, 1995). This kind of experimental support can be implemented dynamically into the simulation based on a learner's actual experimental behavior. Meaning, hints are given whenever a learner displays inadequate behavior.

Reflective support. Reflective support is provided after the interaction with a simulation and encourages learners to reflect on each experiment as well as overall in order to design valuable conclusions and generalizations. Studies indicate that learners often tend to misinterpret data to fit their conclusions with their current hypothesis (Chinn & Brewer, 1998; Klahr, Fay, & Dunbar, 1993). Adequate support is given through reflective assignment tools or the opportunity to discuss students' own results or results generated by other students in the class (de Jong & van Joolingen, 1998). Prompting learners to reflect upon and justify their experimental activities can raise learners' self-awareness and might also contribute to higher knowledge acquisition (Eckhardt et al., 2013; White & Frederiksen, 1998; Zhang, Chen, Sun, & Reid, 2004). Further, prompting students to criticize someone else's experiments (also possible as a fictional student's experiment) can foster them to recognize confounds in the experiments designed by others and to use this knowledge to conduct their own valid experiments (H.-Y. Chang & Linn, 2013).

To sum up, there are a bunch of possibilities to provide students with instructional support with dynamic visualizations. Moreover, three central assumptions can be summarized concerning learning with visualizations:

- Multimedia learning is often more effective than learning from texts alone.
- Dynamic visualizations like computer simulations have the advantage of making processes tangible that are not visible to the human eye.
- Simulation-based learning is successful when combined with practical activities in the form of instructional support to overcome learners' limitations (e.g., problems with discovery learning).

In the next section, the three major topics evolution, threshold concepts, and visualizations will be combined and explained how dynamic visualizations may foster a better understanding of both threshold concepts and evolution.

2.4 Learning Threshold Concepts through Dynamic Visualizations

Understanding evolutionary principles and key concepts is fundamental for science literacy, while an understanding of underlying threshold concepts is regarded to be essential for a better understanding of evolutionary processes. In fact, Tibell and Harms (2017) stated that most of the problems in learning evolutionary principles and key concepts might be due to a lack of understanding the underlying abstract concepts (threshold concepts) such as randomness and probability.

Nevertheless, these concepts might be tangible through appropriate visualization. Thus, using (dynamic) visualizations can be a useful tool for learning threshold concepts. In fact, there are several reasons that visualizations, and particularly computer simulations, could be a helpful tool for learners to foster their understanding of random and probabilistic processes in evolutionary contexts. For instance, random genetic mutations are essential sources of variation on which the process of natural selection can act upon (e.g., Heams, 2014; Mayr, 2001). However, mutations are not visible to the naked human eye, although they can be visualized technologically (e.g., using DNA sequencing techniques). The consequent lack of possibilities for students to observe these phenomena in everyday situations may result in a misunderstanding of the importance and the nature of random processes in evolution (Garvin-Doxas & Klymkowsky, 2008; Mead & Scott, 2010; Tibell & Harms, 2017). Thus, simulations may overcome these limitations and provide students with tangible visualizations.

Furthermore, if students understand the nature of random genetic mutations, they might realize how the process of natural selection acts upon these random processes. For instance, a mutation that turns out to be beneficial in a particular environment affects survival and reproduction probabilities. Thus, individuals with an advantageous trait are more likely to survive and reproduce, and cumulative change in a population over generations may occur (advantageous traits become a majority within the population). Hence, evolution occurs through random processes (the occurrence of a beneficial mutation) and probabilistic processes (selection of more individuals with advantageous traits). Simulations that explicitly visualize the threshold concept of randomness, and build connections to the probabilistic processes of natural selection may help students to overcome their problems with these concepts.

However, although biology is one of the major subjects with research on dynamic visualizations, the topic of evolution is rarely mentioned (Lee & Tsai, 2013; Rutten, van Joolingen, & van der Veen, 2012). To date, the available computer simulations are quite complex, often designed for undergraduates, or focus on only specific processes of evolution (e.g., Price, Pope, Abraham, Maruca, & Meir, 2016; Soderberg & Price, 2003; Speth, Long, Pennock, & Ebert-May, 2009). Moreover, although the number of available online educational videos increases, they often lack explanations regarding underlying threshold concepts (Bohlin, Göransson, Höst, & Tibell, 2017). Additionally, threshold concepts are often communicated orally, but not visually. Thus, there is a lack of appropriate dynamic visualizations to foster the understanding of threshold concepts in evolution.

3 AIM AND OVERVIEW OF THE CONDUCTED STUDIES

Although evolutionary theory is the integrative framework of biology and learning the essential tenets is widely considered as a fundamental feature of scientific literacy, vast bodies of empirical research have provided insights into teachers' and students' struggle with teaching and learning evolution (and underlying threshold concepts), and the various alternative conceptions that are held. Still, there is a lack of appropriate test instruments to actually measure students' understanding of (specific) threshold concepts. Thus, in a first step, test instruments have to be developed to measure students' conceptual knowledge of threshold concepts. In a next step, the available evidence should be used to develop effective teaching strategies and learning resources to support an understanding of threshold concepts and evolutionary theory.

This dissertation aims to expand the existing body of work on evolution education by (a) developing a test instrument that measures students' conceptual knowledge of randomness and probability, (b) exploring the relationships between knowledge and acceptance of evolution, and (c) using computer simulations focusing on threshold concepts in evolution as learning resources to support students' conceptual knowledge of randomness and probability. The following sections briefly present the rationale of each of the four conducted studies (Chapters 4–7). In addition, Table 3.1 provides an overview of the main information presented in the research article (Chapter 4) and manuscripts (Chapter 5–7).

3.1 Study 1 (Chapter 4): University Students' Conceptual Knowledge of Randomness and Probability in the Contexts of Evolution and Mathematics

Several studies report evolution educational problems associated with underlying abstract threshold concepts such as randomness and probability. Another problem is the lack of appropriate instruments for assessing students' conceptual knowledge of these threshold concepts. Despite the wide variety of instruments measuring students' knowledge of evolution, tools that assess students' understanding of randomness and probability – especially in the context of evolution - are lacking. Furthermore, there is no empirical evidence about students' conceptual structures regarding randomness and probability in biological contexts, and their connections (if any) to conceptual structures in mathematic contexts.

This cross-sectional study addresses this need by focusing on the developmental process of two instruments: The "Randomness and Probability test in the context of Evolution" (RaProEvo), and the "Randomness and Probability test in the context of

Mathematics" (RaProMath). The newly developed instruments were administered to 140 German university students to provide first insights into the empirical structure of students' conceptual knowledge of randomness and probability and the relationship to their evolutionary knowledge.

3.2 Study 2 (Chapter 5): EvoSketch: Simple Simulations for Learning Random and Probabilistic Processes in Evolution, and Effects of Instructional Support on Learners' Conceptual Knowledge

Former results (see study 1) reveal that students' conceptual knowledge of randomness and probability is connected to their knowledge of evolutionary theory. Thus, fostering students' conceptual knowledge of threshold concepts might also increase evolutionary knowledge. The literature indicates that visualizations can help in capturing invisible abstract concepts. Particularly simulations combined with instructional support have shown to be most effective for learning.

Therefore, the presented study tests the effectiveness of the EvoSketch simulations for teaching and learning the roles of randomness and probability in an evolutionary context (i.e., mutation and selection, respectively). A further aim is to identify the optimal kind of additional instructional support (if any) to use. Altogether, 267 German secondary school students participated in this experimental repeated measures design study with four intervention groups.

3.3 Study 3 (Chapter 6): Is Statistical Reasoning Relevant for Evolution Education?

Although a rich body of research explored the relationship between knowledge and acceptance of evolution, this work has not explored the potential contributions to students' conceptual knowledge of randomness and probability yet. Indeed, it is likely that this knowledge impacts not only evolutionary knowledge, given the intimate connections between these domains, but also students' acceptance of evolutionary theory.

Thus, a total of 538 American undergraduate students participated in this explorative, cross-sectional study to examine the relationships among students' conceptual knowledge of randomness and probability, their evolutionary knowledge, and the acceptance of evolution. In addition, the empirical structure of students' conceptual knowledge of randomness and probability in evolution and mathematics (measured by the RaProEvo and RaProMath) is reanalyzed with this international cohort.

3.4 Study 4 (Chapter 7): Item Context Affects the Use of Threshold Concepts in Student Explanation of Evolution by Natural Selection

Empirical studies have already indicated that item contexts affect learners' use of evolutionary principles and key concepts. Moreover, experts and novices seem to be attracted by different features of evolutionary problems. Nevertheless, the use of threshold concepts in students' explanations of evolutionary problems was not yet observed.

Thus, the aim of this study is to describe how students apply key concepts and threshold concepts in their written explanations on evolutionary processes, and to characterize the relation between item features and the expression of threshold concepts. A total of 247 university students from Sweden and Germany participated in this study and were asked to provide written answers to three open response items focusing on evolutionary processes (i.e., antibiotic resistant bacteria, fast running cheetahs, and blind cave salamanders).

Table 3.1 Overview of the main information of the conducted studies

Study 1 (Chapter 4): Randomness and Probability Knowledge	
Publication	Fiedler, D., Tröbst, S., & Harms, U. (2017). University students' conceptual knowledge of randomness and probability in the contexts of evolution and mathematics. *CBE-Life Sciences Education, 16,* ar38. doi:10.1187/cbe.16-07-0230
Context	Evolution; Mathematics; Randomness; Probability
Level	University
Aim	1. Provide first insights of the empirical structure of students' conceptual knowledge of randomness and probability 2. Examine the relationship between conceptual knowledge of randomness and probability and evolutionary knowledge
Design	Cross-sectional online survey study
Sample	*N* = 140 German university students
Instruments	− Randomness and Probability test in the context of Evolution (RaProEvo) − Randomness and Probability test in the context of Mathematics (RaProMath) − Open Response Instrument (ORI; Nehm & Reilly, 2007) − Knowledge Processing subscale of the Berlin Evaluation Instrument for Self-Evaluated Student Competencies (Braun, Gusy, Leidner, & Hannover, 2008)
Study 2 (Chapter 5): Learning Randomness and Probability	
Manuscript	Fiedler, D., Tröbst, S., Großschedl, J., & Harms, U. (submitted, 02/2018). EvoSketch: Simple simulations for learning random and probabilistic processes in evolution, and effects of instructional support on learners' conceptual knowledge. *Journal of Research in Science Teaching.*
Context	Evolution; Randomness; Probability; Simulations
Level	Secondary school
Aim	1. Explore the effectiveness of the EvoSketch simulations for teaching and learning random and probabilistic processes in evolutionary context. 2. Identify the optimal kind of additional instructional support (if any) to use.
Design	Experimental repeated measures design study with four intervention groups
Sample	*N* = 267 tenth grade school students from comprehensive schools in Germany
Instruments	− Randomness and Probability test in the context of Evolution (RaProEvo) − Conceptual Inventory of Natural Selection (CINS; D. L. Anderson et al., 2002) − General Biological Content Knowledge test (GBCK; Neubrand, 2017; Neubrand et al., 2016) − General language proficiency (C-Test; Wockenfuß & Raatz, 2006) − Perceived Cognitive Load (PCL; Urhahne, 2002) − Self-reported test-taking effort (OECD, 2010)

(Continued)

Table 3.1 Continued

Study 3 (Chapter 6): Statistical Reasoning in Evolution Education	
Manuscript	Fiedler, D., Sbeglia, G., Nehm, R. H., & Harms, U. (in preparation). Is statistical reasoning relevant for evolution education? Intended for *Science Education*.
Context	Evolution; Mathematics; Randomness; Probability
Level	University
Aim	1. Investigate the relationships among knowledge of randomness and probability, understanding, and acceptance of evolution. 2. Examine the degree to which knowledge of randomness and probability might serve as predictors for understanding and acceptance.
Design	Cross-sectional online survey study
Sample	*N* = 583 American undergraduate students from an introductory biology course
Instruments	– Randomness and Probability test in the context of Evolution (RaProEvo) – Randomness and Probability test in the context of Mathematics (RaProMath) – Conceptual Assessment of Natural Selection (CANS; Kalinowski, Leonard, & Taper, 2016) – Inventory of Student Evolution Acceptance (I-SEA; Nadelson & Southerland, 2012)
Study 4 (Chapter 7): Use of Threshold Concepts	
Manuscript	Göransson, A., Orraryd, D., Fiedler, D., & Tibell, L. A. E. (in preparation). Item context affects the use of threshold concepts in student explanation of evolution by natural selection. Intended for the *International Journal of Science Education.*
Context	Evolution; Randomness; Probability; Temporal scale; Spatial scale
Level	University
Aim	1. Explore how students apply key concepts and threshold concepts in their written explanation of evolutionary processes. 2. Investigate the relation between item context and the expression of threshold concepts.
Design	– Cross-sectional online survey study – Qualitative analyses of three open response items regarding evolutionary processes (bacteria, cheetah, salamander)
Sample	*N* = 247 university students from Sweden (*n* = 38) and Germany (*n* = 209)
Instruments	– Open Response Instrument (ORI; Nehm & Reilly, 2007)

4 STUDY 1: UNIVERSITY STUDENTS' CONCEPTUAL KNOWLEDGE OF RANDOMNESS AND PROBABILITY IN THE CONTEXTS OF EVOLUTION AND MATHEMATICS[1]

Abstract

Students of all ages face severe conceptual difficulties regarding key aspects of evolution: the central, unifying and overarching theme in biology. Aspects that are strongly related to abstract "threshold" concepts like randomness and probability appear to pose particular difficulties. A further problem is the lack of an appropriate instrument for assessing students' conceptual knowledge of randomness and probability in the context of evolution. To address this problem we have developed two instruments, called "Randomness and Probability test in the context of Evolution" (RaProEvo), and "Randomness and Probability test in the context of Mathematics" (RaProMath), which include both multiple-choice and free-response items. The instruments were administered to 140 university students in Germany, then the Rasch partial credit model was applied to assess them. The results indicate that the instruments generate reliable and valid inferences about students' conceptual knowledge of randomness and probability in the two contexts (which are separable competencies). Furthermore, RaProEvo detected significant differences in knowledge of randomness and probability, as well as evolutionary theory, between biology majors and preservice biology teachers.

Keywords:

Evolution, Randomness, Probability, Threshold concepts, Conceptual knowledge

[1] This is the peer-reviewed version of the following article: Fiedler, D., Tröbst, S., & Harms, U. (2017). University students' conceptual knowledge of randomness and probability in the contexts of evolution and mathematics. *CBE-Life Sciences Education, 16,* ar38. doi:10.1187/cbe.16-07-0230, which has been published in final form at https://doi.org/10.1187/cbe.16-07-0230. This article may be used for non-commercial purposes in accordance with the American Society for Cell Biology (ASCB) Terms and Conditions for Use of Self-Achieved Versions.

4.1 Introduction

Evolution through natural selection is a central, unifying and overarching theme in biology. Evolutionary theory is the integrative framework of modern biology and provides explanations for similarities among organisms, biological diversity, and many features and processes of our world. For example, the evolution of oxygenic photosynthesis massively affected geochemistry, and the evolution of organisms with calcareous shells led to the formation of limestone (e.g., Castanier, Le Métayer-Levrel, & Perthuisot, 1999; Kopp, Kirschvink, Hilburn, & Nash, 2005). It is also applied in numerous other fields, both biological (e.g., agriculture and medicine) and non-biological (e.g., economics and computer science). Therefore, the essential tenets of evolutionary theory have long been regarded as key parts of the foundations of science education (e.g., Beardsley, 2004; Bishop & Anderson, 1990; Nehm & Reilly, 2007; Pugh, Linnenbrink-Garcia, Koskey, Stewart, & Manzey, 2010; Speth et al., 2014). Accordingly, the American Association for the Advancement of Science (AAAS, 2006), the Next Generation Science Standards (NGSS 2013), the National Education Standards of Germany (Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany [KMK], 2005a) as well as official documents of many other countries, all describe evolution as an organizing principle for biological science and include the topic as a learning goal.

Although evolutionary processes may occur in numerous kinds of systems, unless specified otherwise evolution generally refers to changes over time (also referred as between generations) in populations or taxa of organisms due to the generation of variation and natural selection (Gregory, 2009). There is a massive empirical body of work on evolution, myriads of processes involved have been elucidated (e.g. genetic drift, genetic linkage, endosymbiosis, adaptive radiation and speciation), and extensive terminology has been developed (e.g., Rector, Nehm, & Pearl, 2013; Reinagel & Speth, 2016). However, biologists generally agree that three principles are necessary and sufficient for explaining evolutionary change by means of natural selection: (1) the generation of variation, (2) heritability of variation, and (3) differential reproductive success of individuals with differing heritable traits (Endler, 1986; Gregory, 2009). This framework is deceptively simple, because myriads of interactions are involved in phenomena such as adaptive radiation (the diversification of taxa leading to the filling of vacant ecological niches; Schluter, 2000). Furthermore, key processes such as speciation may occur gradually over long times and numerous generations or in a single generation, if a massive chromosomal change or polyploidization is involved. Similarly, some important processes involve atomic-level phenomena while others involve large-scale spatio-temporal variations in environmental variables and populations' genetic structures. Moreover, natural selection acts on phenotypes (organisms' observable traits), but adaptive changes are mediated by genetic changes that generally either enhance organisms' reproductive success (thereby allowing the alleles they carry to spread in their respective populations) or enable colonization of new niches (Schluter, 2000). Hence, evolutionary change is far from simple, and it is still poorly understood by students throughout the educational hierarchy (Nehm & Reilly, 2007; Shtulman, 2006; Spindler & Doherty, 2009), science teachers (Nehm, Kim, & Sheppard, 2009; Osif, 1997), and the general public (Evans et al., 2010). This poor understanding has been attributed to diverse cognitive, epistemological, religious, and emotional factors (for an overview see Rosengren, Brem, Evans, & Sinatra, 2012).

Tibell and Harms (2017) concluded that complete understanding of evolutionary theory might require the understanding of more general abstract concepts like randomness, probability, and different scales in space and time. These general abstract concepts coincide with a set of recently proposed ‘threshold concepts’ in genetics and evolution (Ross et al., 2010; Taylor, 2006). According to emerging theory initiated by Meyer and Land (2006), such concepts are portals that provide access to new ways of thinking; acquisition of understanding of these concepts is said to alter students’ perspectives and lead them to see things through a different lens. Threshold concepts are distinguished from ‘key’ or ‘core’ concepts as they are more than mere building blocks towards understanding within a discipline and are tentatively proposed to have five characteristics: *transformative* (occasioning a shift in perception and practice), probably *irreversible* (unlikely to be forgotten or unlearned), *integrative* (surfacing patterns and connections), often disciplinarily *bounded*, and *troublesome* (Meyer & Land, 2006). Threshold concepts in diverse disciplines have been examined, including: economics (Davies & Mangan, 2007), chemistry (Park & Light, 2009), biology (Taylor & Cope, 2007), biochemistry (Loertscher, Green, Lewis, Lin, & Minderhout, 2014), and computer science (Zander et al., 2008).

In the context of evolution, Tibell and Harms (2017) developed a two-dimensional framework connecting principles and key concepts of evolutionary theory with the abovementioned general abstract concepts like randomness and probability. They propose that complete understanding of evolutionary theory requires the development of knowledge concerning not only the principles of evolution but also general abstract concepts like randomness and probability, and the ability to freely navigate through this two-dimensional framework.

4.1.1 Randomness, Stochasticity and Probability

Random and probabilistic processes are key elements of evolutionary theory, and several studies report educational problems associated with the underlying abstract concepts (Robson & Burns, 2011; Ross et al., 2010). When considering random processes in evolution, students are reportedly challenged by both the terminology (Mead & Scott, 2010) and conceptual complexity (Garvin-Doxas & Klymkowsky, 2008).

The term *random,* as used in everyday life and scientific contexts (e.g. mathematics and biology), is connected to various conceptions and interpretations. In everyday life, an event is often called random if it is very rare, strange or unusual, and hence unpredictable or uncertain (Bennett, 1998). This common perception of randomness or ‘chance occurrences’ does not change with increasing age (Falk &

Konold, 1997; Kattmann, 2015), which hinders understanding of the concept of randomness (and the closely related concept stochasticity) in scientific disciplines including mathematics (Kaplan, Rogness, & Fisher, 2014) and biology (Mead & Scott, 2010). Random and stochastic are widely treated as synonymous terms, and definitions vary, but most mathematical texts and dictionaries note a distinction. Here, the term random is used when referring to phenomena (such as rolling dice) “where the outcome is probabilistic rather than deterministic in nature; that is, where there is uncertainty as to the result." (Smith, 2012, p. 1). In accordance with *oxforddictionaries.com*, stochastic is used to describe processes for which outcomes have “a random probability distribution or pattern that may be analyzed statistically but may not be predicted precisely”. It should be noted that random and stochastic can often be used in the same contexts, because a process may be random in the sense that it is influenced by random variables and stochastic in the sense that it has probabilistic outcomes. More formally, “a stochastic process is a family of random variables $X\theta$, indexed by a parameter θ, where θ belongs to some index set Θ” (Breuer, 2006, p. 1).

Randomness and stochasticity are fundamental elements of biological theories related to phenomena at all scales and levels, including the evolutionary gene-, individual-, population- and environment-level processes involved in both the generation of variation and natural selection (Heams, 2014; Tibell & Harms, 2017). For example, the individual-level processes of mutation and recombination are regarded as random. Mutations may occur (at low frequencies) either in coding regions (thereby potentially affecting the structure and function of encoded proteins) or non-coding regions (thereby potentially affecting expression patterns). Hence, mutations of either kind may profoundly change organisms’ phenotypes. Clearly, the reactions involved must follow physico-chemical laws, but they are regarded as random because the individual-level outcomes are far beyond our ability to model predictively at this level (Heams, 2014), although we can determine population-level (stochastic) frequencies of mutations at given sites or sequences of DNA. Further, at population or environmental levels random processes may involve, for example, the death of single organisms through causes that cannot be directly linked to selective (dis)advantages (Tibell & Harms, 2017), so even organisms close to an adaptive peak may die as juveniles. Thus, randomness and stochasticity are major elements of biological processes generally, and evolution specifically. However, a desire to ascribe causes to all events appears to be an intrinsic element of human nature (Falk, 1991), which may lead to a denial of chance in general, and explain why students have difficulties perceiving evolutionary events as aimless random occurrences (Kattmann, 2015). Furthermore, students tend to perceive biological processes as efficient, and random processes as inefficient (Garvin-Doxas & Klymkowsky, 2008).

To summarize, randomness and stochasticity (as defined here) are closely related, but randomness refers to processes or variables that are uncertain rather determinate, while stochasticity refers to probabilities of outcomes of processes in or affecting populations. Probability is the likelihood of a particular outcome and is assigned a numerical value between zero and one (Feller, 1968). The closer a probability value is to one, the more likely the outcome. Crucially, an outcome that is extremely rare at individual level, such as a given beneficial mutation, is extremely likely to occur at least once in a population that is sufficiently large or over a sufficiently long timeframe (in terms of number of generations). In the context of evolution, probability plays a role in all three of the principles mentioned above, but particularly selection and inheritance (Tibell & Harms, 2017). For example, fertilization in sexual reproduction involves probabilistic events like the choice of mate. The best-adapted individuals are most likely to survive to reproductive maturity, mate and thus to reproduce. Hence, the frequencies of organisms with given traits in a given environment depend on many random events, and the process of selection can also be defined as the probabilities of individuals with differing traits in a given population surviving and reproducing in a specific environment. Although reproduction depends upon survival and many other different factors (as mentioned above), it is still reproduction, herby including fitness, that is relevant evolutionarily. However, it should also be remembered that selection acts on random processes involved in generation of variation (Mayr, 2001), but the importance of these processes seems to be a learning obstacle for students (Garvin-Doxas & Klymkowsky, 2008; Lynch, 2007).

Moreover, biology students not only struggle to grasp the importance and roles of randomness, probability and stochasticity in evolutionary theory (Gregory, 2009), but also often have a weak understanding of mathematics (Hester, Buxner, Elfring, & Nagy, 2014; Jungck, 1997). This clearly hinders the teaching and learning of evolution as mathematical descriptions of randomness and probability are key elements of the explanations of random and stochastic evolutionary (and other) biological processes (Buiatti & Longo, 2013; Wagner, 2012). To date, there is no empirical evidence about students' conceptual structures regarding randomness and probability in biological contexts, and their connections (if any) to conceptual structures in mathematics contexts. However, some studies indicate that mathematical modeling can generally lead to improvements in problem-solving and qualitative conceptual knowledge, i.e. students' ability to predict likely outcomes of processes (Chiel, McManus, & Shaw, 2010; Schuchardt & Schunn, 2016). Thus, there is a need to explore the possible connections between understanding of evolutionary theory and conceptual knowledge of randomness and probability in both evolutionary and mathematical contexts.

4.1.2 Development of Content-Related Knowledge in Higher Education

In Germany, higher education in biology is divided into two stages, generally consisting of a 3–4 years course leading to a Bachelor's degree followed by a 1–2 years course leading to a Master's degree (KMK, 2010). Bachelor's courses are intended to equip students with a broad qualification by providing academic subject-specific foundations, methodological skills and competences related to the professional field, while Master's courses provide further subject and academic specialization (KMK, 2010).

Higher education to become a teacher includes at least two subjects and students can take – depending on the Land (federal state) or higher education institution – either a basic foundation course (concluding with the first state exam) or a graded course (with Bachelor's and Master's degrees) (KMK, 2010). In all programs, subject areas, subject didactics and educational science components are coupled and supplemented with practical components in the form of school internships. The relative amounts of time allocated to subject areas and educational science depend on the Land and type of school the students aspire to teach in. Typically, the contents of preservice biology teachers' education in their chosen subjects (e.g., biology) account for 30–40 % of the total in Bachelor's and basic foundation courses, and 20–25 % in Master's courses (VBIO, 2006).

At the beginning of Bachelor's programs most universities offer a compulsory module on the topic of *general biology*. This module should enable students to gain sound knowledge about the structure and function of cells, acquire insights into the diversity and evolution of plants and animals, and learn the basic techniques of biological investigations. Students subsequently take various compulsory or elective modules, depending on the university and whether they are biology majors or preservice teachers, like *genetics*, *ecology*, *evolution*, *cell biology*, and/or *molecular biology* (VBIO, 2006). Regarding *evolution* (or evolutionary theory), both sets of students are normally exposed to the topics of mechanisms of evolution, micro and macro evolution, evolutionary theories, and abiotic and biotic factors (see Supplemental Material 4.7.1). Nevertheless, there is a substantial difference in development of biological knowledge between biology majors and preservice biology teachers. Although some seminars are attended by both, preservice biology teachers have fewer opportunities to learn the subject. Therefore, preservice biology teachers may tend to have less deep and detailed knowledge about specific biological processes. As evolution is described as an organizing principle for biological science and an explicitly stated learning goal in diverse standards (e.g. AAAS, 2006; KMK, 2005a; NGSS, 2013) both biology majors and preservice teacher students should ideally have a shared core of knowledge regarding evolutionary changes through natural selection. Further, this general knowledge is important, because evolutionary theory is the integrative framework of modern biology and its essential tenets are key parts of the foundations of, and for, science education.

4.1.3 Research Objective

Diverse instruments have been developed for measuring evolutionary knowledge (e.g., D. L. Anderson, Fisher, & Norman, 2002; Nadelson & Southerland, 2009; Nehm, Beggrow, Opfer, & Ha, 2012; Price et al., 2014). However, we are not aware of any tool for measuring understanding of randomness and probability, although they play major roles in evolutionary processes (Tibell & Harms, 2017). Thus, a robust test instrument for measuring understanding of these two abstract concepts, and their roles in evolution, is required to advance evolution education research and assess both biological courses and students. In efforts to meet this need we have developed an instrument called the "Randomness and Probability test in the context of Evolution" (RaProEvo) and a sister instrument called the "Randomness and Probability test in the context of Mathematics" (RaProMath), to explore the empirical structure of biology students' conceptual knowledge of randomness and probability, and the relationship of this knowledge to their conceptual knowledge of evolutionary theory. During development of these instruments we applied previous findings on students' common difficulties when trying to learn evolutionary concepts (e.g., Gregory, 2009; Mead & Scott, 2010). Here, we describe their development, provide indications of their valid measures (expert ratings and criterion-related valid measures), and present results of field tests of the instruments on biology majors and preservice biology teachers.

4.2 Methods

4.2.1 Participants

During the 2015–2016 academic year we recruited 140 biology students (26.4% male) — 72 biology majors (30.6% male) and 68 preservice biology teachers (22.1% male) — enrolled at 23 German universities to complete an online survey. The participants' average age was 22.9 years (*SD* = 3.7); 22.2 years (*SD* = 2.9) for biology majors and 23.7 years (*SD* = 4.3) for preservice biology teachers. On average, they had attended 5.3 semesters (*SD* = 3.6) in tertiary education, with a mean of 4.7 semesters (*SD* = 3.9) for biology majors and 5.8 semesters (*SD* = 3.2) for preservice biology teachers. A total of 79 students (56.4% of all participants; 41 biology majors, 38 preservice biology teachers) had taken compulsory modules on evolution or evolutionary biology and had been introduced to the topic of evolution (e.g., mechanisms of evolution, micro and macro evolution, evolutionary theories, and abiotic and biotic factors). Further, 48 of these students (34.3% of all participants) had also taken compulsory modules in genetics, ecology and cell or molecular biology, while 10 students (7.1% of all participants) had only taken the evolutionary module. Students were also asked to provide Likert-type responses ranging from 1 (*Not at all*) to 4 (*Intensively*) to the items regarding their learning opportunities in the contexts of evolution, genetics, and ecology. Their self-reported statements indicate that considerable attention was paid to evolution (*M* = 9.51, *SD* = 1.8) genetics (*M* = 8.43, *SD* = 2.68) and ecology (*M* = 8.67, *SD* = 2.27) during their higher education.

4.2.2 Procedure

Participants responded to a basic demographic questionnaire (including items probing their academic self-concept) and completed tests on conceptual knowledge of randomness and probability in both evolutionary and mathematical contexts. The structure of the online survey was the same for all participants and had no time limit. On average, the students took 58 min 56 s (*SD* = 15 min 14 s; range 20 min 4 s to 94 min) to complete the survey. All respondents were given the opportunity to participate in a lottery for 10 vouchers, each worth 50 Euros (approximately US$ 54 at the time of the survey).

4.2.3 Measures

Randomness and Probability Knowledge Test

Development. The first step in developing or considering an instrument to measure students' conceptual knowledge of randomness and probability in the context of evolution is to clarify the types of such knowledge they should develop during their education. To do so we first designated two focal topics (contexts): evolution and mathematics. For the evolution context, we identified the following five aspects in which randomness and probability play important roles that biology graduates and teachers should understand: (1) o*rigin of variation*, (2) *accidental death (single events* such as death of one individual rather than another that is not linked to differences in their adaptation to their environment, e.g. an individual could be struck by lightning while less well adapted individuals escape injury and produce more offspring*)*, (3) *random phenomena*, (4) *process of natural selection*, and (5) *probability of events*. For the mathematics context we selected the following five topics: (1) s*ingle events*, (2) *random phenomena*, (3) *probability as ratio*, (4) *sample reasoning*, and (5) *probability of events*. To explore knowledge of these topics (explained in Table 4.1), we reviewed previously published instruments for testing evolutionary knowledge (e.g., D. L. Anderson et al., 2002; Bowling et al., 2008; Fenner, 2013; Robson & Burns, 2011) and knowledge of randomness and/or probability in various fields (e.g., Eichler & Vogel, 2012; Falk & Konold, 1997; Garfield, 2003; Green, 1982). Items deemed suitable were included in a pool of questions ($N = 65$ items; Table 4.2). Most items were translated from English into German and almost all were modified more than once to fit the specific purpose of the instrument. Additionally, a number of questions were created by three researchers of the *EvoVis* project group (*EvoVis: Challenging Threshold Concepts in Life Science - enhancing understanding of evolution by visualization*). Distractors for these items were mainly based on students' alternative conceptions reported in previous studies (e.g., Gregory, 2009). A coding scheme was provided for each item.

Table 4.1 Explanation of randomness and probability topics in evolution and mathematics contexts and corresponding questions in the test instruments

Topic	**Learning objective – Students should be able to:**	**Question No.**
Evolution		
Origin of variation	Explain the cause of genetic variability (e.g., mutation, recombination), their impact on survival, and their importance for evolutionary processes.	E01, E02, E03, E07, E11, E12, E17
Accidental death (single event)	Evaluate sudden death of single individuals in a population are not per see due to natural selection, and thus, a random process.	E04, E09
Random phenomena	Identify and explain common processes in evolution that are called to be random (e.g., mutations).	E05, E13, E14, E15
Process of natural selection	Determine that natural selection acts on phenotypes of populations with different organisms produce different numbers of offspring. Over time, this can result in specialization for particular ecological niches.	E06, E10, E16, E18
Probability of events	Apply mathematical modeling to biological processes and thus, to argue with.	E08, E19
Mathematics		
Single event	Determine the definitions of random processes [(i) unpredictability of single outcomes, but (ii) predictable in long terms], and thus, to argue with.	M02, M03, M06, M10, M14, M17, M25, M26, M29, M33
Random phenomena	Interpret results as outcomes of random phenomena.	M23, M27
Probability as ratio	Distinguish between equally likely and non-equally likely experiments, and thus can predict the probability of simple experiments.	M01, M04, M08, M12, M15, M16, M18, M20, M21, M22, M28
Probability of events	Applying appropriate methods to predict the probability of multi-stage experiments (e.g., probability tree diagram or combinatorics).	M05, M07, M09, M11, M13, M24, M37
Sample reasoning	Explain how samples are linked to populations and which conclusions can be made from samples to populations.	M19, M30, M31

Table 4.2 Sources of the final RaProEvo and RaProMath test items

Context	Topic	Item	Source of the idea/item (Item code)
Evolution	Origin of variation	E01	Fenner, 2013 (item #24 pretest)
		E02	Fenner, 2013 (item #26 pretest)
		E03	Robson & Burns, 2011 (item #5 pretest)
		E07	Campbell & Reece, 2011 (item #3, chapter 23)
		E11	Author[1]
		E12	Bowling *et al.*, 2008 (item #9)
		E15	Author
		E17	Campbell, Reece, & Markl, 2006 (item #8)
	Accidental death (single event)	E04	Author
		E09	Author
	Random phenomena	E05	Campbell *et al.*, 2006 (item #16)
		E13	Author
		E14	Klymkowsky, Underwood, & Garvin-Doxas, 2010 (item #4)
	Process of natural selection	E06	Author
		E10	Fenner, 2013 (item #20 pretest)
		E16	Author
		E20	Author
	Probability of events	E08a	Author
		E08b	Author
		E19a	Green, 1982 (item #7)
		E19b	Author
Mathematics	Single event	M02	Green, 1982 (item #8)
		M03	Green, 1982 (item #1)
		M06	Green, 1982 (item #21a)
		M10	Green, 1982 (item #21d)
		M14	Jones, Langrall, Thornton, & Mogill, 1997 (item #CP1)
		M17	Author
		M25	Eichler & Vogel, 2009
		M26	Author
		M29	Author
		M33	Green, 1982 (item #25)
	Random phenomena	M23	Author
		M27	Falk & Konold, 1997
	Probability as ratio	M01	Garfield, 2003 (item #8)
		M04	Green, 1982 (item #3)
		M08	Green, 1982 (item #2)
		M12	Jones *et al.*, 1997 (item #CP2)
		M15	Green, 1982 (item #17)
		M16	Author
		M18	Green, 1982 (item #6d)
		M20	Herget, Kösters, & Merziger, 2009 (item #1a, test part 3)
		M21	Weber & Mathea, 2008 (item #5, test form 2)
		M22	Jones *et al.*, 1997 (item #CP2)
		M28	Herget *et al.*, 2009 (item #1b, test part 3)
	Probability of events	M05	Author
		M07	Garfield, 2003 (item #18)
		M09	Green, 1982 (item #22)
		M11	Garfield, 2003 (item #13)
		M13	Garfield, 2003 (item #19)
		M24	Garfield, 2003 (item #9)
		M37	Weber & Mathea, 2008 (item #6, test form 1)
	Sample reasoning	M19	Garfield, 2003 (item #14)
		M30	Green, 1982 (item #21d)
		M31	Green, 1982 (item #23)

[1]Developed by authors together with other members of the *EvoVis* project

Two preliminary versions of tests were developed to capture biology students' conceptual knowledge of randomness and probability in the contexts of evolution and mathematics, designated RaProEvo and RaProMath, respectively. The RaProEvo test included a mixture of dichotomously scored (0 = no credit, 1 = full credit) and partial credit (0 = no credit, 1 = partial credit, 2 = full credit) items, while items of the RaProMath test were all dichotomously scored (0 = no credit, 1 = full credit). In order to assess interrater reliability of the open-ended items, two raters independently coded the responses using scoring rubrics. Cohen's kappa interrater reliability statistics (Cohen, 1960) for these RaProEvo and RaProMath versions were .93 and .91, respectively. Discrepancies were resolved via deliberation between the raters. Items with a negative or low discrimination index ($r_{it} < .10$) were excluded from further analysis ($n = 3$).

Faculty review. We examined content valid measures of the developed test instruments by soliciting faculty input to help validate the items. For this purpose, we administered an online version of RaProEvo to evolutionary biology faculty members (hereafter: biology experts) and an online version of RaProMath to faculty members with expertise in stochastics and/or probability (hereafter: mathematics experts) of different institutions. Biology experts were asked to select the correct response for each item, and were asked if the item (1) tests the intended learning objective (Table 4.1) and (2) is scientifically accurate. A summary of their alignment is presented in Table 3. Experts could also add comments regarding each item and provide feedback. Mathematics experts were asked to follow the same procedure but evaluate the mathematical accuracy of the items (Table 4.3). A total of 13 biology experts (10 faculty members and three PhD students) and 10 mathematics experts (eight faculty members and two PhD students) provided feedback on the instruments. In all cases, items with an agreement <80% had been flagged as potentially problematic, and thus were deleted or critically revised. The experts' suggestions on the intended learning objective were primarily to reword questions to increase precision and eliminate possible ambiguities, but two RaProEvo and six RaProMath items were scientifically or mathematically incorrect, and thus deleted. At the end of this process, we were left with a 21-item RaProEvo test (16 multiple-choice, three free-response, and two matching items; see Supplemental Material 4.7.2) and a 33-item RaProMath test (30 multiple-choice, and three free-response items; see Supplemental Material 4.7.3).

Table 4.3 Summary of RaProEvo and RaProMath faculty review

	Items with given faculty agreement		
	>90%	>80%	<80%
RaProEvo			
The item tests the intended learning objective	18	4	1
The information given in the item is scientifically accurate	15	5	3
RaProMath			
The item tests the intended learning objective	32	0	7
The information given in the item is mathematically accurate	32	0	7

Test of Evolutionary Knowledge

Students' conceptual knowledge of evolutionary theory was assessed using the Open Response Instrument (ORI) published by Nehm and Reilly (2007). The instrument was designed to determine how successfully biology majors can answer questions about natural selection at different levels of complexity and to identify both student knowledge and alternative conceptions. We used the following three, of five, items from this instrument:

- Explain why some bacteria have evolved resistance to antibiotics (that is, the antibiotics no longer kill the bacteria).
- Cheetahs (large African cats) can run faster than 60 miles (97 km) per hour when chasing prey. How would a biologist explain how the ability to run fast evolved in cheetahs, assuming their ancestors could run at only 20 miles (32 km) per hour?
- Cave salamanders (amphibious animals) are blind (they have eyes that are not functional). How would a biologist explain how blind cave salamanders evolved from ancestors that could see?

To score students' evolutionary explanations, we established and refined two scoring rubrics in a pilot study with a set of 39 biology students. The first scoring rubric "key concepts" covered eight key concepts: (1) origin of variation (e.g., mutation and recombination), (2) individual variation, (3) differential survival potential linked to specific traits, (4) inheritance of traits, (5) reproductive success, (6) selection pressure including limitations of resources, (7) limited survival, and (8) changes in populations or distributions of individuals with certain traits (explained in Table 4.4). Two raters independently coded their responses in these terms to compute interrater reliability, and Cohen's kappa interrater reliability was found to be .76. In cases of disagreement, all coding discrepancies were resolved via deliberation. This scoring rubric was used to quantify the presence or absence of the eight key concepts in each of the students' responses. The mean numbers of key concepts each student referred to in responses to all three items (hereafter: key concept score) and in responses to each of the three items (hereafter: key concept diversity, KCD) were found to be 8.01 (*SD* = 4.89, range 0 to 19, out of a maximum possible score of 24) and 4.61 (*SD* = 2.42, range 0 to 8, out of a maximum possible score of 8), respectively. The second scoring rubric, "alternative conceptions concerning natural selection" (hereafter: alternative conceptions), was developed using seven common, well-known alternative conceptions that have been extensively documented in research literature (Bishop & Anderson, 1990; Gregory, 2009; Nehm et al., 2012; Nehm & Reilly, 2007): (1) need, (2) use and disuse, (3) anthropomorphism, (4) essentialism, (5) soft inheritance, (6) events vs. processes, and (7) source vs. sorting of variation (explained in Table 4.4). Two raters independently coded their responses in these terms to compute interrater reliability, and Cohen's kappa interrater reliability was found to be .73. In cases of disagreement, all coding discrepancies were resolved via deliberation. This scoring rubric was used to quantify the presence or absence of the seven common alternative conceptions in each of the students' responses. The mean numbers of alternative concepts each student referred to in responses to all three items (hereafter: alternative concept score) and in responses to each of the three items (hereafter: alternative concept diversity, ACD) were found to be 0.55 (*SD* = .71, range 0 to 3, out of a maximum possible score of 21) and .35 (*SD* = .56, range 0 to 3, out of a maximum possible score of 7), respectively.

To quantify students' evolutionary knowledge in terms of key concept and alternative conception measures more fully, we used the Natural Selection Performance Quotient (NSPQ) of Nehm and Reilly (2007). The NSPQ is derived by multiplying KCD/(KCD+ACD) and KCD/maximum possible key concept score, and expresses the product on a scale of 0–100. The "first term expresses the proportion of students' answers that were correct, and the second expresses how the correct proportion compared to the most complete possible answer" (Nehm & Reilly, 2007, p. 266). Further, the NSPQ distinguishes between students who have significant knowledge of natural selection, but conceptual problems, and those with no alternative conceptions but differing levels of knowledge (Nehm & Schonfeld, 2007). The mean NSPQ of our sample was .55 (*SD* = .31).

High School Grade Point Average (GPA)

The high school grade point average (GPA) is one of the most important criteria for selecting candidates for higher education in Germany (Heine *et al.*, 2006) and is widely used as a proxy for cognitive ability (J. R. Anderson & Lebière, 1998). Thus, we used self-reported GPA to assess the convergent valid measures of the RaProEvo and RaProMath tests. GPA was captured by a single item, with scores ranging from 1 (*good performance*) to 4 (*poor performance*). The results indicate that our students' GPA, and hence cognitive ability, covered a sufficiently wide range for robustly testing our instruments (*M* = 2.17, *SD* = 0.53, Min = 1.00, Max = 3.50).

Table 4.4 Explanations of key concepts and alternative conceptions

Topic	The response refers to the following aspects:
Key concepts	
Origin of variation	Changes are caused by mutation or recombination.
Individual variation	Differences in the traits of individuals are addressed (e.g. the fastest).
Differential survival potential	Individuals have different survival potentials due to specific traits (e.g. higher survival potential, evolutionary advantage).
Inheritance of traits	Traits are passed on from the individual to their offspring (or next generation).
Reproductive success	Some individuals have higher reproductive success than others.
Selection pressure	Designation of selection factors, selection pressure or limited resources (e.g. light, prey).
Limited survival	Imagine that some individuals will survive, while others die.
Changes in populations	[Beneficial] traits are getting more frequent.
Alternative conceptions	
Need	Individuals develop the new trait or behavior because they *need* it to survive (or the trait disappears because they do not need it)
Use and disuse	New trait or physical changes result from the use or non-use and are passed on directly to the offspring.
Anthropomorphism	The individual *knows* about the benefit / non-use of a characteristic and therefore it appears or disappears.
	Natural selection (nature) is understood as a sorting-out force.
Essentialism	The individuals of a population change at the same time and develop the new feature.
Soft inheritance	Characteristics learned by an individual during the lifetime are passed on to the offspring.
Events vs. processes	Natural selection is an event with start and end (and is not understood as continuous).
Source vs. sorting of variation	Mutations appear because of a changed environment and are therefore advantageous.

Students' Academic Self-Concept

To investigate the criterion-related valid measures of the RaProEvo and RaProMath tests we assessed students' academic self-concept. This is reportedly a highly important and influential predictor of cognitive and behavioral outcomes such as performance and self-worth, and it also seems to be strongly related to academic achievement (Marsh & Martin, 2011). Further, Paulick, Großschedl, Harms, and Möller (2016) showed that preservice biology teachers' academic self-concept is positively related to their biological knowledge. Hence, RaProEvo and RaProMath score should be positively correlated with academic self-concept in evolutionary theory and stochastics, respectively.

To assess our participants' academic self-concept we used the "Knowledge Processing" subscale of the Berlin Evaluation Instrument for Self-Evaluated Student Competencies (BEvaKomp; Braun, Gusy, Leidner, & Hannover, 2008). This instrument operationalizes knowledge processing as students' self-reported competency (based on self-knowledge and evaluation of value or worth of one's own capabilities) regarding a specific subject. We adapted the (five) selected BEvaKomp items to the topics of evolutionary theory and stochastics, then asked our students to provide Likert-type responses ranging from 1 (*Does not apply at all*) to 4 (*Fully applies*) to the items regarding both contexts (see Supplemental Material 4.7.4, for items). The results indicate that our students had medium self-reported competency in evolutionary theory (M = 3.04, SD = 0.72, Min = 1.00, Max = 4.00; Cronbach's alpha = .93) and somewhat lower self-reported competency in stochastics (M = 2.42, SD = 0.82, Min = 1.00, Max = 4.00; Cronbach's alpha = .94). All rating levels were chosen by at least five students.

4.2.4 Statistical Analysis

Test Instrument Dimensionality

In order to tackle whether students' conceptual knowledge of randomness and probability in the context of evolution and mathematics follow a single dimension or are better modeled as two separate dimensions, we first conducted a principle component analysis on Rasch scores in IBM SPSS Statistics (Version 23). It has been suggested to assume unidimensionality when the first component explains at least 20% of the total variance (Reckase, 1979). Further, a single dimension is supported with one large eigenvalue and a large ratio of the first and second eigenvalue (Hutten, 1980; Lord, 1980).

Rasch analysis was applied in ACER ConQuest® (Version 1; Wu, Adams, Wilson, & Haldane, 2007) to analyze the psychometric distinction of students' conceptual knowledge of randomness and probability in the contexts of evolution and mathematics. Since the two tests were designed to capture students' conceptual knowledge of randomness and probability in two contexts, a two-dimensional model was fitted to the data, based on the assumption that students have separable competencies for evolution and mathematics, which can be captured as the latent traits "competency in RaProEvo" (measured by the 21 evolutionary items) and "competency in RaProMath" (measured by the 33 mathematical items), respectively. This model was compared to a one-dimensional model presuming a single competency, i.e. that items represent one latent trait ("competency in Randomness and Probability", measured by 21 evolutionary combined with 33 mathematically items).

To determine which model provides the best fit to the acquired data, we calculated final deviance values, which are negatively correlated with models' fits (and thus indicate degrees of support for underlying assumptions). To test whether the two-dimensional model fits the data significantly better than the one-dimensional model, we applied a χ^2 test (Bentler, 1990). In addition, we applied two information-based criteria, Akaike's (1981) Information Criterion (AIC) and Bayes' Information Criterion (BIC), to compare the two models. These criteria do not enable tests of the significance of differences between models, but generally the values are negatively correlated to the strength of models' fits to the data (Wilson, De Boeck, & Carstensen, 2008).

Test Instrument Evaluation by Rasch Modeling

Assuming that evolution and mathematics competencies differ, the reliable measures and internal structure of the RaProEvo and RaProMath instruments were evaluated by analyzing the participants' responses using the Rasch partial credit model (PCM) and Wright Maps. The PCM is rooted in Item Response Theory and provides a means for dealing with ordinal data (Bond & Fox, 2015; Wright & Mok, 2000), by converting them into interval measures, thus allowing the calculation of parametric descriptive and inferential statistics (Bond & Fox, 2015; Smith Jr, 2000; Wright & Mok, 2000). The discrepancy between a considered PCM and the data is expressed by so-called fit statistics (Bond & Fox, 2015). Since person and item measures are used for further analyses, only items fitting the model should be included, otherwise values of these measures could be skewed and lead to wrong conclusions in further analyses. To calculate fit statistics for the RaProEvo and RaProMath instruments we used ACER ConQuest® item response modeling software (Version 1; Wu et al., 2007). ConQuest provides *outfit* and *infit mean square statistics* (hereafter *outfit* and *infit*, respectively) to measure discrepancies between observed and expected responses. The *infit* statistic is mainly used for assessing item quality as it is highly sensitive to variation in discrepancies between models and response patterns, while *outfit* is more sensitive to outliers (Bond & Fox, 2015). Furthermore, aberrant *infit* statistics usually raise more concern than aberrant *outfit* statistics (Bond & Fox, 2015). Therefore, we used the Weighted Mean Square (WMNSQ): a residual-based fit index with an expected value of 1 (if the underlying assumptions are not violated), ranging from 0 to infinity. We deemed WMNSQ values acceptable if they were within the range 0.5 to 1.5 (Wright & Linacre, 1994) and had *t*-values that did not significantly deviate from 1.0 (being within the range -2.0 to 2.0).

To test whether the developed test instruments fit the Rasch model, model fit indices regarding the items and participants' abilities ('person ability') were calculated. Person ability and item difficulty were estimated using Masters' (1982) partial credit model as RaProEvo includes a mixture of dichotomously scored and partial credit items. The partial credit model allows analysis of items scored in more than two ordered categories, with different measurement scales for different items, and estimates a distinct threshold parameter for each item (Wright & Mok, 2000). Four reliability indices — person reliability, person separation, item reliability, and item separation — were calculated (Bond & Fox, 2015). For further analysis, person parameters were estimated by calculating weighted maximum likelihood estimation (WLE) values.

Valid Measures Check

Spearman's rho correlation coefficients were used to assess criterion-related (convergent/discriminate) valid measures of the applied instruments and the relationship between students' knowledge of evolutionary theory and their conceptual understanding of randomness and probability. The instruments' convergent valid measures was assessed by testing the association between the participants' person ability scores and GPA (assumed to be negatively correlated), while their discriminant valid measures was assessed by testing the association between their person ability scores and academic self-concepts (assumed to be stronger for corresponding than for non-corresponding self-concepts).

Furthermore, we applied one way analyses of covariance (ANCOVA) to explore differences between biology majors' and preservice biology teachers' knowledge in terms of: (a) person RaProEvo ability, (b) person RaProMath ability, and (c) students' evolutionary knowledge (KCD, ACD, and NSPQ). In all cases participants' GPA was a covariate.

4.3 Results

4.3.1 Test Instrument Dimensionality

Rasch scores principle component analysis was conducted to tackle the issue of dimensionality. The first component obtained in this analysis explained 11.25% of the total variance. In order, the eigenvalues of the first five components were 6.08, 3.71, 2.88, 2.39 and 2.13. Correspondingly, the ratio of the first and second eigenvalues was 1.64, indicating the lack of a dominant single dimension.

To determine whether students' conceptual knowledge of randomness and probability in the context of evolution is psychometrically distinct from their mathematical knowledge of randomness and probability, we compared two-dimensional and one-dimensional partial credit models fitted to data obtained from coding 140 biology students' responses to the two instruments. Rasch analysis results and AIC values indicate that the two-dimensional model provides a better fit to the data, although values of the other information-based criterion applied (BIC) indicates that the one-dimensional model provides a better fit (Table 4.5). Nevertheless, results of a χ^2 test show that the two-dimensional model significantly outperformed the one-dimensional model: χ^2 $(2, N = 140) = 6.23$, $p = .044$. Thus, students' conceptual knowledge of randomness and probability in evolutionary and mathematical contexts appear to be empirically separable competencies. Accordingly, the Spearman's correlation coefficients between their knowledge in the two contexts were $r_{\text{latent}} = .86$ and $r_{\text{manifest}} = .59$ ($p < .001$), indicating that the two competencies are closely related but distinct.

Table 4.5 Final deviance and information criteria for comparing the two- and one-dimensional models of students' conceptual knowledge of randomness and probability (*N* of items = 54)

	Context of conceptual knowledge	One-dimensional model	Two-dimensional model
Allocation to dimension	Evolution	A	A
	Mathematics	A	B
Deviance		6178.15	6171.92
(no. of free parameters)		(57)	(59)
AIC		6292.15	6289.92
BIC		6459.83	6463.48

Note. A = indicator(s) of dimension 1; B = indicator of dimension 2.

4.3.2 Test Instrument Analysis

As the two-dimensional model represents students' conceptual knowledge of randomness and probability slightly better than the one-dimensional model, the results regarding the reliable measures and internal structure of RaProEvo (*N* = 140, 21 items) and RaProMath (*N* = 140, 33 items) are presented separately (see Supplemental Material 4.7.5 and 4.7.6, for item parameter estimates).

RaProEvo. The Wright map acquired from analysis of the RaProEvo test results (Figure 4.1) was used to analyze the internal structure of the instrument (Boone & Rogan, 2005). In such a map, the distributions of persons and items of the instrument (or more strictly person ability and item difficulty estimates) are plotted along the same dimension (conventionally to the left and right, respectively) and can be directly compared. Items of equivalent difficulty are located at the same position on the scale (e.g., rpE14 and nsE18; Figure 4.1), and persons at the same position or height on the scale as a particular item have a 50% chance of answering that item correctly, while those located above and below an item respectively have a higher and lower than 50% chance of answering it correctly. The RaProEvo Wright map suggests that a typical respondent would answer most questions correctly, as 38.1% and 61.9% of the items were respectively above and below the position of the mean person (dotted line). Nevertheless, fits for items forming the test for conceptual knowledge of randomness and probability in evolution were acceptable, with WMNSQ values ranging from 0.81 to 1.07 and *t*-values from -1.9 to 0.7. For the 21 items of the RaProEvo test, an item separation reliability of .98, a WLE person separation reliability of .58, and a Cronbach's alpha (internal consistency) value of .66 were computed. Mean person ability (person parameters) was found to be 0.01 (*SD* = 0.08) and the mean score (item parameters) was 17.27 points (*SD* = 2.93, range 6 to 22 points, maximum possible score = 23 points).

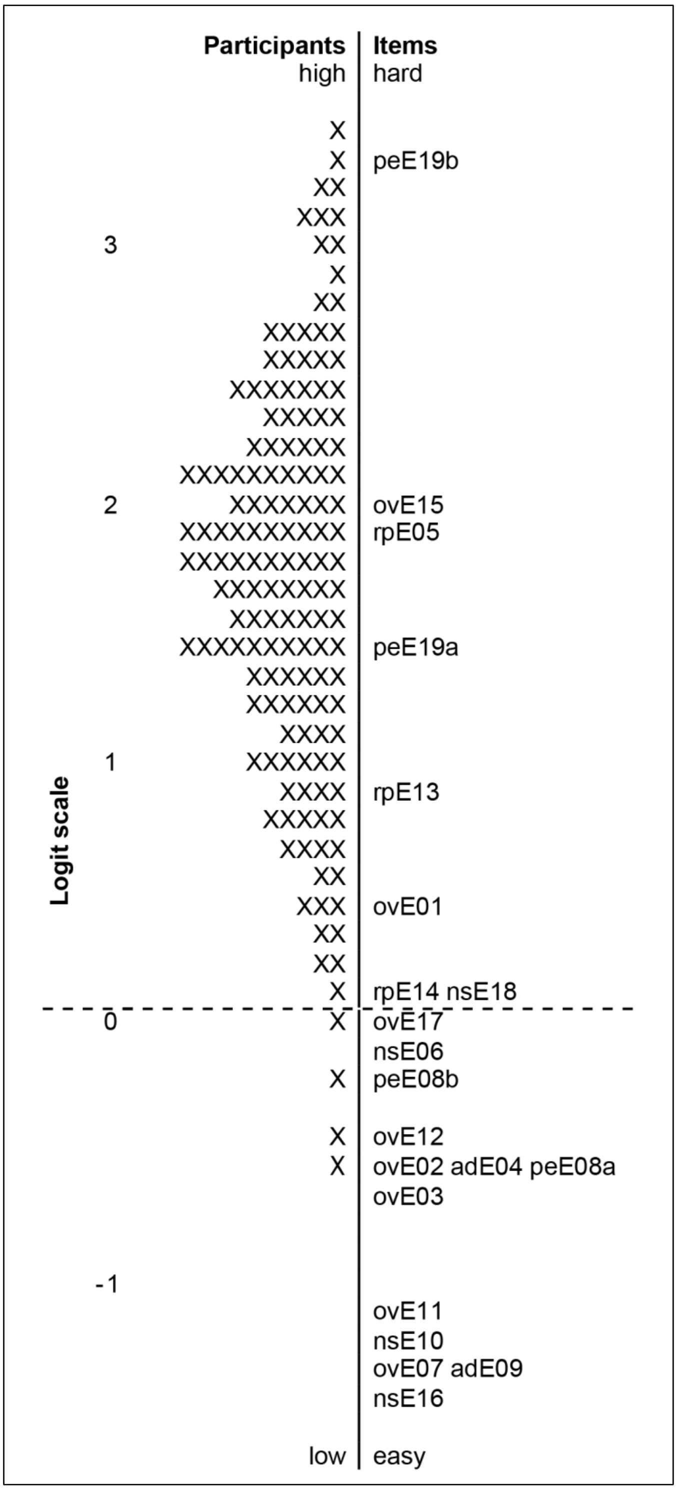


Figure 4.1 Wright map of responses to items of the RaProEvo test (N = 140; 21 items). Abilities of persons who took the test are displayed on the left and difficulty of the (coded) items on the right. Each X indicates 0.9 individuals in the sample. The first two letters stand for: *ov* origin of variation, *ad* accidental death (single event), *rp* random phenomena, *ns* process of natural selection, and *pe* probability of events. *E* represents the content of evolutionary theory, while *01* to *19* indicates the item number in the RaProEvo test and the last letter stands for *a* item 1 and *b* item 2 within a similar item task.

RaProMath. The Wright map acquired from analysis of the RaProMath test results (Figure 4.2) was also used to assess the internal structure of the instrument (Boone & Rogan, 2005). Like the RaProEvo map, it suggests that a typical respondent would answer most questions correctly, as 42.4% and 57.6% of the items were respectively above and below the position of the mean person (dotted line). Like those in the RaProEvo test, the items forming the test for conceptual knowledge of randomness and probability in mathematics had acceptable fit, with WMNSQ values ranging from 0.80 to 1.12 and *t*-values from -1.8 to 1.8. For the 33 items of the RaProMath test, an item separation reliability of .99, a WLE person separation reliability of .68, and a Cronbach's alpha (internal consistency) value of .69 were computed. Mean person ability (person parameters) was found to be 0.02 (SD = 0.90) and the mean score (item parameters) was 24.01 points (SD = 3.66, range 10 to 31 points, maximum possible score = 33 points).

Additionally, the joint Wright map generated from the two-dimensional model (Figure 4.3), which enables comparison of patterns of knowledge of randomness and probability in evolution and mathematics contexts, shows that the two instruments detected similar spread in our students' abilities.

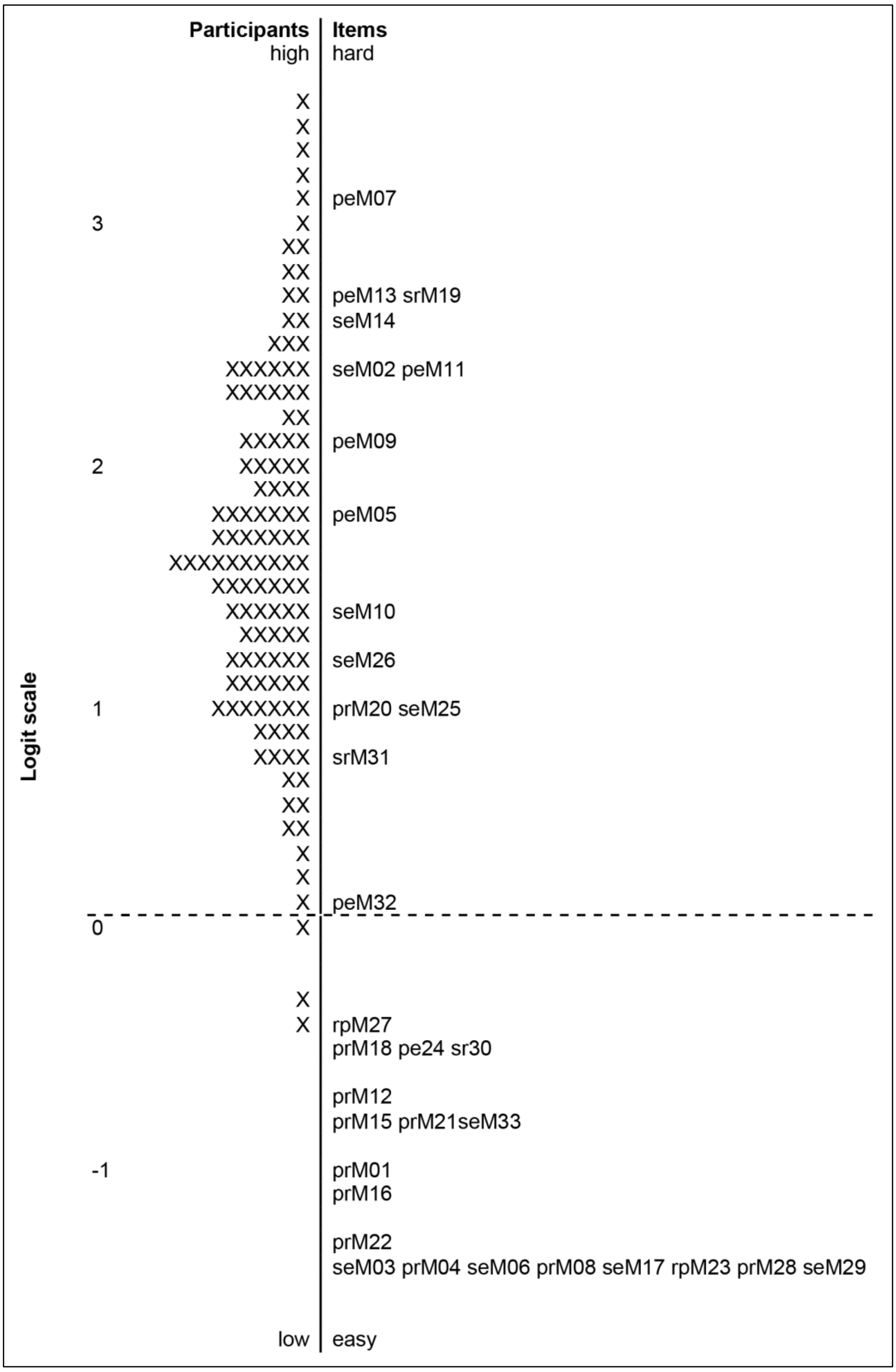


Figure 4.2 Wright map of responses to items of the RaProMath test (N = 140; 33 items). Abilities of persons who took the test are displayed on the left and difficulty of the (coded) items on the right. Each X indicates 1.1 individuals in the sample. The first two letters stand for: *se* single event, *rp* random phenomena, *pr* probability as ratio, *pe* probability of events, and *sr* sample reasoning. *M* represents the content of mathematics, while *01* to *33* indicates the item number in the RaProMath test.

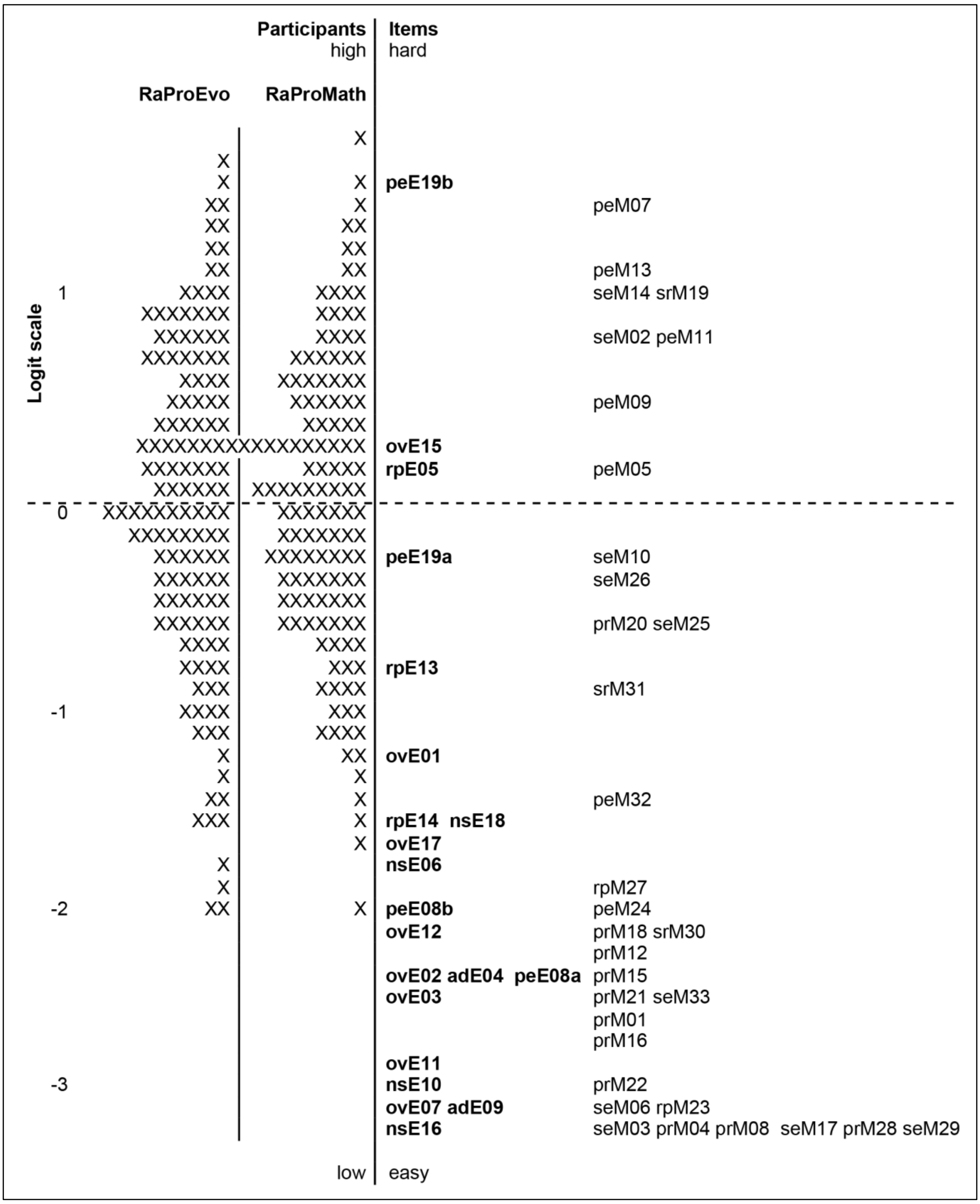


Figure 4.3. Wright map of responses to items linked to the two dimensions of the RaProEvo test (*bold*; *N* = 140; 21 items) and RaProMath test (*N* = 140; 33 items). Abilities of persons who took the test are displayed on the left and the difficulty of the (coded) items on the right. Each X indicates 1.0 individuals in the sample. The first two letters stand for: *ov* origin of variation, *ad* accidental death (single event), *rp* random phenomena, *se* single event, *ns* process of natural selection, *pe* probability of events, *pr* probability as ratio, and *sr* sample reasoning. *E01* to *E19* indicates the item number in the RaProEvo test, while *M01* to *M33* represents the item number in the RaProMath test. The last letter stands for *a* item 1 and *b* item 2 within a similar item task.

4.3.3 Valid Measure Check

To test the instruments' valid measures, we first analyzed the relationships between the participants' GPA and person ability in the two knowledge dimensions of randomness and probability in evolutionary and mathematical contexts to assess their convergent valid measures. The results confirmed our hypotheses that GPA values would be negatively correlated with both RaProEvo and RaProMath person abilities ($r_s = -.25$, $p = .004$ and $r_s = -.33$, $p < .001$, respectively; $n = 129$ in both cases).

Next, we analyzed the relationship between the two dimensions of person ability (knowledge of randomness and probability in evolutionary and mathematical contexts) and the participants' academic self-concepts to assess the tests' discriminant valid measures. The results confirmed our hypothesis that participants' academic self-concepts in the contexts of evolutionary theory and mathematics would be more strongly connected to their RaProEvo and RaProMath composite scores, respectively (Table 4.6).

The results also showed that KCD in students' responses was significantly positively related to their person abilities as measured by both RaProEvo ($r_s = .45$) and RaProMath ($r_s = .35$), while ACD was significantly negatively related to these abilities ($r_s = -.32$ and -.17, respectively). Furthermore, the NSPQ was significantly positively related to their abilities measured by RaProEvo ($r_s = .47$) and RaProMath ($r_s = .36$). These findings ($p < .001$, $N = 140$, in all cases) confirm the hypothesis that their conceptual knowledge of evolutionary theory would be positively correlated with their conceptual knowledge of randomness and probability.

Table 4.6 Spearman's rank correlation coefficients (rho) between students' academic self-concepts of evolutionary theory and stochastics, and their knowledge of randomness and probability

	Academic self-concept	
	Evolutionary theory	Stochastics
RaProEvo	.40 **	.13
RaProMath	.19 *	.23 **

* $p < .05$, ** $p < .01$

Note. RaProEvo = person parameters in conceptual knowledge of roles of randomness and probability in evolution; *RaProMath* = person parameters in conceptual knowledge of randomness and probability in mathematics.

4.3.4 Biology Majors vs. Preservice Biology Teachers

To assess effects of study program on the participants' performance we applied one-way analyses of covariance (ANCOVA) to compare RaProEvo- and RaProMath-measured abilities of biology majors and preservice biology teachers, their KCD scores, ACD scores, and NSPQ while controlling for cognitive ability as indicated by GPA.

We detected a significant effect of study program on RaProEvo scores: biology majors obtained significantly higher RaProEvo person ability scores (adj $M = 0.33$, *SEM* = 0.11, *n* = 69) than preservice biology teachers (adj *M* = -0.33, *SEM* = 0.12, *n* = 60): $F(1, 126) = 15.97$, $p < .001$. In contrast, the study program had no significant effect on RaProMath scores: $F(1, 126) = 1.54$, $p = .217$. Regarding KCD and ACD scores, biology majors (adj *M* = 5.02, *SEM* = 0.29, *n* = 69) used significantly more key concepts in their answers than preservice biology teachers (adj *M* = 4.00, *SEM* = 0.31, *n* = 60), $F(1, 126) = 5.78$, $p = .018$, while study program has no significant effect on numbers of alternative conceptions identified in their responses: $F(1, 126) = 1.40$, $p = .239$. Nevertheless, biology majors obtained significantly higher NSPQs (adj *M* = .60, *SEM* = .04, *n* = 69) than preservice biology teachers (adj *M* = .46, *SEM* = .04, *n* = 60): $F(1, 126) = 6.27$, $p < .001$.

4.4 Discussion

We have attempted to address the need for instruments capable of measuring understanding of two important abstract concepts underlying the biological concepts in evolutionary theory (randomness and probability) and advance evolutionary education research. Using the presented instruments we explored the psychometric distinction of biology students' conceptual knowledge of randomness and probability in the context of both evolution (RaProEvo) and mathematics (RaProMath). We then assessed the reliable and valid measures of the RaProEvo and RaProMath instruments. Finally, we investigated the relationships of RaProEvo and RaProMath scores with evolutionary knowledge (KCD, ACD, and NSPQ) and the difference in this knowledge between biology majors and pre-service biology teachers.

Several of the empirical findings are of potential interest, particularly given the importance of understanding randomness and probability, both in science generally, as highlighted in national and international education standards (NGSS, 2013; KMK, 2005a, 2005b), and specifically in teaching and learning evolution (Mead & Scott, 2010; Tibell & Harms, 2017). First, the percentage of the total variance explained by the first components of a Rasch scores principle components analysis as well as the ratio of the first and second eigenvalues of this principle components analysis reveals a lack of unidimensionality. Second, Rasch analysis also indicated that a two-dimensional model fits the participants' responses slightly but significantly better than a one-dimensional model, supporting the assumption that RaProEvo and RaProMath measure separate competencies. We obtained promising indications of the instruments' reliable measures, albeit preliminary due to the small sample size, and their valid measures was confirmed by experts and criterion-related indications. Furthermore, biology students' RaProEvo scores, KCD and NSPQ (but not ACD) were all higher than those of the pre-service teachers, indicating that they had more evolutionary knowledge. In contrast, RaProMath scores did not differ between biology students and pre-service teachers.

4.4.1 Randomness and Probability Knowledge

There was a good fit between the dataset and Rasch model, indicating that the tests had strong internal valid measures. Detailed analysis indicated that the RaProEvo instrument's difficulty was not optimal for our sample of biology students: numerous items clustered at the low end of the scale, and there was a lack of sufficiently difficult items to distinguish high performers. Nevertheless, the Wright maps obtained from our analysis of responses to items of the tests provided indications of informative patterns regarding students' thinking (which require further verification), as outlined below.

The RaProEvo Wright map indicates that most students could satisfactorily answer items regarding the *process of natural selection* (Figure 4.1), which mainly concerned broad, probabilistic aspects of the process, rather than specific contributory processes or key associated concepts. Illustrative phenomena used in these questions might be mostly familiar, such as changes in color of foxes' fur in adaptive responses to environmental changes, a frequently used example of natural selection-mediated change that many students may learn from textbooks. In contrast, only high-performing students correctly answered questions with complex probabilistic backgrounds (psE19; *probability of events*).

Similar patterns were discerned in responses to the RaProMath instrument. Questions concerning *probability as ratio* seemed quite easy for the participants. This may seem unsurprising, as pupils learn to calculate ratios in primary school (KMK, 2005b). However, only high performers correctly answered items concerning *probability of events*, although students should have learned this topic in school too (KMK, 2004, 2015). The finding corroborates indications presented by various authors (e.g., Chi, Feltovich, & Glaser, 1981) that students tend to ignore connections to underlying concepts (e.g., probability), which could allow them to transfer their understanding to other problems. It is a concern as students have to calculate and apply ratios in biology explicitly in topics such as Mendelian inheritance and Hardy-Weinberg equilibrium (e.g., Campbell et al., 2006), and (more often) implicitly in diverse contexts (e.g. the influence of alleles' selective strength on the probability of fixation as a function of the strength of genetic drift), which increases the sophistication of the required conceptualization (Tibell & Harms, 2017).

In the mathematical context, students found some of the *single event* items challenging (some were apparently easy, but responses to more than half of them were distributed across the scale). Even when asked about the (un)predictability of single events students seemed to think about predictability in aggregate terms. Similarly, in the evolutionary context, items regarding *origin of variation*, either generally (e.g., ovE03, ovE17) or specific sources of variation like recombination (e.g., ovE07) and mutation (e.g., ovE12) were also distributed across the entire scale. Finally, *random phenomena* seemed quite challenging for our students in evolutionary contexts. When they had to explain why evolutionary change through natural selection is a nonrandom process, they often forgot that natural selection acts upon randomly generated variation. Indeed, as noted by Mayr (2001): "Without variation, there would be no selection". Even Darwin (1859) suggested that variation is a fundamental requirement for evolutionary change in *On the Origin of Species by Means of Natural Selection* (for more information see Gregory, 2009), although he could not explain where the variation comes from. Nevertheless, only 17% of the participants stated this in their answers.

A particularly important source of new variation in the focal contexts is mutation, which is regarded as a random process, partly because the probability of mutations occurring is not affected by the selective consequences, and partly because their occurrence in a given individual at a given time is far beyond our modelling capacities (Gregory, 2009; Heams, 2014). Nevertheless, several studies have indicated that students tend to struggle with both the importance of random processes such as the origin of variation in evolutionary processes and understanding why mutations are called random (e.g., Garvin-Doxas & Klymkowsky, 2008; Smith, Wood, & Knight, 2008; Speth et al., 2014). Our results corroborate these findings that random processes pose learning difficulties.

4.4.2 Differences between Biology Majors and Preservice Biology Teachers

Most studies of evolutionary knowledge focus on differences between novice and advanced students attending similar study programs (e.g., Frasier & Roderick, 2011; Nehm & Ridgway, 2011). However, possible differences between biology majors and preservice biology teachers are also potentially important, particularly as the latter will form the next generation to teach evolutionary theory. So, it might be acceptable for preservice biology teachers to lack detailed knowledge of specific associated processes, and thus obtain lower scores in tests such as RaProEvo, but they should have similar general understanding (as measured, for example, by KCD and NSPQ) of evolutionary change through natural selection. Alarmingly, we found significant deficits (relative to the biology majors) in both their conceptual knowledge of randomness and probability in evolutionary contexts and their evolutionary knowledge. These findings cannot be explained by differences in cognitive abilities, because we accounted for variations in participants' GPA, and Klusmann (2013) found no differences in cognitive characteristics between students attending teacher and other university education courses. However, we cannot exclude the possibility that these findings are simply a manifestation of differences that existed between the groups before their higher educational training.

Regardless of the reasons for the preservice teachers' lower RaProEvo scores there will clearly be potential problems in teaching evolution if the next generation of teachers has only modest knowledge of (or harbors misconceptions about) it. Thus, when considering strategies to improve biology students' understanding it is important not only to foster development of accurate evolutionary knowledge, but also to ensure that the next generations of teachers develop an adequate knowledge base. As proposed by Tibell and Harms (2017), a step towards appropriate solutions could be to deepen students' knowledge of abstract concepts underlying evolutionary processes.

4.4.3 Limitations and Future Research

Mathematics is a compulsory subject in school and mathematical concepts, particularly randomness and probability, are fundamental elements of descriptions of myriads of biological interactions, relationships, and processes (Chiel et al., 2010; Jungck, 1997). However, most previous studies on evolutionary knowledge have solely considered biological aspects (Tibell & Harms, 2017). A major implication of our study is that conceptual knowledge of randomness and probability is important for biology students' understanding of evolutionary theory. In contrast to other studies on students' misunderstanding of random processes (Garvin-Doxas & Klymkowsky, 2008; Robson & Burns, 2011), we also detected clear differences in students' conceptual knowledge of randomness and probability in evolutionary and mathematical contexts. In developing the instruments we also tried to extend extant research by showing that threshold concepts are important factors for a deeper conceptual knowledge of evolutionary theory, and thus important in students' education.

Nevertheless, instruments such as RaProEvo and RaProMath have intrinsic limitations, partly because they need to be reasonably short and not require much time to complete or mark. Thus, they must include only a few items targeting each concept. Hence, the instruments should be used mainly for formative purposes, i.e. for instructors to identify obstacles their students are currently facing. The instruments were not intended to be summative evaluation tools. The utility of RaProEvo and RaProMath lies in their proposed ability to assess students' general conceptual knowledge about randomness and probability in two contexts (evolution and mathematics), rather than exhaustively assess their knowledge of specific constructs (e.g., genetic drift).

We also note the obvious limitation of the small sample size in our study. We obtained promising preliminary results, but the reliable inferences of the instruments must be confirmed with a larger group of students. Further, the participants were all German students from a single cohort. To assess the generality of the findings and identify causes of possible variations in findings, tests of the instruments internationally and with other cohorts are required.

Finally, we hope that our instruments will facilitate efforts to design further tools to assess students' conceptual knowledge of randomness and probability in the future. In addition, having developed an instrument for measuring conceptual knowledge of randomness and probability in the context of evolution (RaProEvo), instruction about randomness and probability connected to evolutionary concepts warrants attention. Therefore, an objective of ongoing research is to investigate if visualizations and/or instructional support can help students to develop better conceptual knowledge of the roles of randomness and probability in evolution and, hence, better evolutionary knowledge.

4.5 Acknowledgments

This study is part of the Swedish-German cooperation project *“EvoVis: Challenging Threshold Concepts in Life Science – enhancing understanding of evolution by visualization”*, funded by the Swedish Research Council (VR 2012:5344, LT). We thank the whole EvoVis group, Dr Jörg Großschedl and Charlotte Neubrand for helpful discussions that contributed to this study. Special thanks to Dr John Blackwell for the language review and valuable and insightful comments on evolution aspects of the paper. We are also very thankful to the students who participated in this study and the experts for their comments on the instruments.

4.6 References

Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics, 16*, 3–14. doi:10.1016/0304-4076(81)90071-3

American Association for the Advancement of Science. (2006). Evolution on the front line: An abbreviated guide for teaching evolution (Project 2061). http://www.project2061.org/publications/TeachingGuides.htm (accessed 01 March 2017)

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

Anderson, J. R., & Lebière, C. (1998). *The atomic components of thought.* Mahawah, NJ: Lawrence Erlbaum.

Beardsley, P. M. (2004). Middle school student learning in evolution: Are current standards achievable? *The American Biology Teacher, 66*, 604–612. doi:10.1662/0002-7685(2004)066[0604:MSSLIE]2.0.CO;2

Bennett, D. J. (1998). *Randomness.* Cambrigde, United Kingdom: Harvard University Press.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.

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

Bond, T., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). New York, NY: Routledge.

Boone, W., & Rogan, J. (2005). Rigour in quantitative analysis: The promise of Rasch analysis techniques. *African Journal of Research in Mathematics, Science and Technology Education, 9*, 25–38.

Bowling, B. V., Acra, E. E., Wang, L., Myers, M. F., Dean, G. E., Markle, G. C., . . . Huether, C. A. (2008). Development and evaluation of a genetics literacy assessment instrument for undergraduates. *Genetics, 178*, 15–22. doi:10.1534/genetics.107.079533

Braun, E., Gusy, B., Leidner, B., & Hannover, B. (2008). Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). [The Berlin evaluation instrument for selfapproximated, student competence]. *Diagnostica, 54*, 30–42. doi:10.1026/0012-1924.54.1.30

Breuer, L. (2006). *Introduction to stochastic processes.* Retrieved from www.kent.ac.uk/smsas/personal/lb209/files/notes1.pdf (accessed 01 March 2017)

Buiatti, M., & Longo, G. (2013). Randomness and multilevel interactions in biology. *Theory in Biosciences, 132*, 139–158. doi:10.1007/s12064-013-0179-2

Campbell, N. A., & Reece, J. B. (2011). *Campbell Biologie Gymnasiale Oberstufe, Übungsbuch*. Munich, Germany: Pearson Schule.

Campbell, N. A., Reece, J. B., & J, M. (2006). *Biologie* (Vol. 6. aktualisierte Auflage). Munich, Germany: Pearson Studium.

Castanier, S., Le Métayer-Levrel, G., & Perthuisot, J.-P. (1999). Ca-carbonates precipitation and limestone genesis—the microbiogeologist point of view. *Sedimentary Geology, 126*, 9–23. doi:10.1016/S0037-0738(99)00028-7

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152. doi:10.1207/s15516709cog0502_2

Chiel, H. J., McManus, J. M., & Shaw, K. M. (2010). From biology to mathematical models and back: teaching modeling to biology students, and biology to math and engineering students. *CBE-Life Sciences Education, 9*, 248–265. doi:10.1187/cbe.10-03-0022

Cohen, J. (1960). A coefficient of agreement for nominal scale. *Educational and Psychological Measurement, 20*, 37–46.

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life.* London, United Kingdom: John Murray.

Davies, P., & Mangan, J. (2007). Threshold concepts and the integration of understanding in economics. *Studies in Higher Education, 32*, 711–726. doi:10.1080/03075070701685148

Eichler, A., & Vogel, M. (2012). Basic modelling of uncertainty: young students' mental models. *ZDM Mathmatics Education, 44*, 841–854. doi:10.1007/s11858-012-0451-9

Endler, J. A. (1986). *Natural selection in the wild.* Princeton, NJ: Princeton University Press.

Evans, E. M., Spiegel, A. N., Gram, W., Frazier, B. N., Tare, M., Thompson, S., & Diamond, J. (2010). A conceptual guide to natural history museum visitors' understanding of evolution. *Journal of Research in Science Teaching, 47*, 326–353. doi:10.1002/tea.20337

Falk, R. (1991). Randomness—an ill-defined but much needed concept. *Journal of Behavioral Decision Making, 4*, 215–218. doi:10.1002/bdm.3960040307

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review, 104*, 301–318. doi:10.1037/0033-295x.104.2.301

Feller, W. (1968). *An introduction to probability theory and its applications* (Vol. 1). London, United Kingdom: John Wiley & Sons.

Fenner, A. (2013). *Schülervorstellungen zur Evolutionstheorie: Konzeption und Evaluation von Unterricht zur Anpassung durch Selektion [Students conceptions about the theory of evolution: Design and evaluation of a teaching sequence about adaption through selection]* (doctoral dissertation). Justus-Liebig-University, Giessen, Germany. Retrieved from http://geb.uni-giessen.de/geb/volltexte/2013/9250/ (accessed 01 March 2017)

Frasier, T. R., & Roderick, C. (2011). Improving how evolution is taught: facilitating a shift from memorization to evolutionary thinking. *Evolution: Education and Outreach, 4*, 298–307. doi:10.1007/s12052-011-0327-4

Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal, 2*, 22–38.

Garvin-Doxas, K., & Klymkowsky, M. W. (2008). Understanding randomness and its impact on student learning: lessons learned from building the Biology Concept Inventory (BCI). *CBE-Life Sciences Education, 7*, 227–233. doi:10.1187/cbe.07-08-0063

Green, D. (1982). *Probability concepts in 11-16 year old pupils Report of research sponsored by the Social Science Research Council*. Loughborough, United Kingdom: CAMET, University of Technology.

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

Heams, T. (2014). Randomness in biology. *Mathematical Structures in Computer Science, 24*, e240308. doi:10.1017/S096012951200076X

Heine, C., Briedis, K., Didi, H. J., Haase, K., & Trost, G. (2006). Bestandsaufnahme von Auswahl- und Eignungsfeststellungsverfahren beim Hochschulzugang in Deutschland und ausgewählten Ländern. [University entrance selection procedures and aptitude tests in Germany and selected countries. A survey]. Retrieved from http://www.dzhw.eu/pdf/pub_kia/kia200603.pdf (accessed 01 March 2017)

Herget, W., Kösters, E., & Merziger, P. (2009). *Stochastisches Denken – Lehren und Lernen [Stochastic thinking – Teaching and learning]*. Retrieved from http://disk.mathematik.uni-halle.de/lehrerseite/Stochastik_Reader_mit_Aufgaben_und_Loesungen_0909.pdf (accessed 01 March 2017)

Hester, S., Buxner, S., Elfring, L., & Nagy, L. (2014). Integrating quantitative thinking into an introductory biology course improves students' mathematical reasoning in biological contexts. *CBE-Life Sciences Education, 13*, 54–64. doi:10.1187/cbe.13-07-0129

Hutten, L. R. (1980). Some empirical evidence for latent trait model selection. *Paper presented at the AERA*, Boston, MA.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children‘s thinking in probability. *Educational Studies in Mathematics, 32*, 101–125. doi:10.1023/A:1002981520728

Jungck, J. R. (1997). Ten equations that changed biology: mathematics in problem-solving biology curricula. *BioScience, 23*, 11–36.

Kaplan, J. J., Rogness, N. T., & Fisher, D. G. (2014). Explotting lexical ambiguity to help students understand the meaning of random. *Statistics Education Research Journal, 13*, 9–24.

Kattmann, U. (2015). *Schüler besser verstehen: Alltagsvorstellungen im Biologieunterricht [Understanding school students: Everyday images in biology lessons]*. Hallbergmoos, Germany: Aulis.

KMK (Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany) (2004). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss (Jahrgangsstufe 10) (Beschluss der Kultusministerkonferenz vom 4.12.2003) [National educational standards for the subject of mathematics concerning the 10th grade]*. Munich, Germany: Luchterhand.

KMK (Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany) (2005a). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss (Jahrgangsstufe 10) (Beschluss der Kultusministerkonferenz vom 16.12.2004) [National educational standards for the subject of biology concerning the 10th grade]*. Munich, Germany: Luchterhand.

KMK (Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany) (2005b). *Bildungsstandards im Fach Mathematik für den Primarbereich (Jahrgangsstufe 4) (Beschluss der Kultusministerkonferenz vom 15.10.2004) [National educational standards for the subject of mathematics concerning the 4th grade]*. Munich, Germany: Luchterhand.

KMK (Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany) (2010). *Common structural guidelines of the Länder for the accreditation of Bachelor's and Master's study courses.* Retrieved from http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2003/2003_10_10-Accredidation-Bachelors-Masters-study-courses.pdf (accessed 01 March 2017)

KMK (Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany) (2015).

Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife (Jahrgangsstufe 12/13) (Beschluss der Kultusministerkonferenz vom 118.10.2012) [National educational standards for the subject of mathematics concerning the 12th/13th grade]. Colone, Germany: Carl Link.

Klusmann, U. (2013). Individual characteristics of prospective teachers. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers* (pp. 311–319). Heidelberg, Germany: Springer.

Klymkowsky, M. W., Underwood, S. M., & Garvin-Doxas, R. K. (2010). Biological Concepts Instrument (BCI): A diagnostic tool for revealing student thinking. *arXiv preprint* arXiv:1012.4501.

Kopp, R. E., Kirschvink, J. L., Hilburn, I. A., & Nash, C. Z. (2005). The Paleoproterozoic snowball Earth: a climate disaster triggered by the evolution of oxygenic photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 11131–11136. doi:10.1073/pnas.0504878102

Loertscher, J., Green, D., Lewis, J. E., Lin, S., & Minderhout, V. (2014). Identification of threshold concepts for biochemistry. *CBE-Life Sciences Education, 13*, 516–528. doi: 10.1187/cbe.14-04-0066

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* New York, NY: Routledge.

Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences, 104*, 8597–8604. doi:10.1073/pnas.0702207104

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology, 81*, 59–77. doi:10.1348/000709910X503501

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149–174. doi:10.1007/BF02296272

Mayr, E. (2001). *What evolution is.* New York, NY: Basic books.

Mead, L. S., & Scott, E. C. (2010). Problem concepts in evolution part II: Cause and chance. *Evolution: Education and Outreach, 3*, 261–264. doi:10.1007/s12052-010-0231-3

Meyer, J. H., & Land, R. (2006). *Overcoming barriers to student understanding: threshold concepts and troublesome knowledge*. Oxon, NY: Routledge.

Nadelson, L. S., & Southerland, S. A. (2009). Development and preliminary evaluation of the Measure of Understanding of Macroevolution: Introducing the MUM. *The Journal of Experimental Education, 78*, 151–190. doi:10.1080/00220970903292983

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

Nehm, R. H., Kim, S. Y., & Sheppard, K. (2009). Academic preparation in biology and advocacy for teaching evolution: Biology versus non-biology teachers. *Science Education, 93*, 1122–1146. doi:10.1002/sce.20340

Nehm, R. H., & Reilly, L. (2007). Biology majors' knowledge and misconceptions of natural selection. *BioScience, 57*, 263–272. doi:10.1641/b570311

Nehm, R. H., & Ridgway, J. (2011). What do experts and novices "see" in evolutionary problems? *Evolution: Education and Outreach, 4*, 666–679. doi:10.1007/s12052-011-0369-7

Nehm, R. H., & Schonfeld, I. S. (2017). Does increasing biology teacher knowledge of evolution and the nature of science lead to greater preference for the teaching of evolution in schools? *Journal of Science Teacher Education, 18*, 699-723. doi:10.1007/s10972-007-9062-7

NGSS Lead States (2013). *Next Generation Science Standards: For States, By States.* Washington, D.C.: The National Academies Press.

Osif, B. A. (1997). Evolution & religious beliefs: a survey of Pennsylvania high school teachers. *The American Biology Teacher, 59*, 552–556. doi:10.2307/4450382

Park, E. J., & Light, G. (2009). Identifying atomic structure as a threshold concept: Student mental models and troublesomeness. *International Journal of Science Education, 31*, 233–258. doi:10.1080/09500690701675880

Paulick, I., Großschedl, J., Harms, U., & Möller, J. (2016). Preservice teachers' professional knowledge and its relation to academic self-concept. *Journal of Teacher Education, 67,* 1–10. doi:10.1177/0022487116639263

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

Pugh, K. J., Linnenbrink-Garcia, L., Koskey, K. L., Stewart, V. C., & Manzey, C. (2010). Motivation, learning, and transformative experience: A study of deep engagement in science. *Science Education, 94*, 1–28. doi:10.1002/sce.20344

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4*, 207–230. doi:10.3102/10769986004003207

Rector, M. A., Nehm, R. H., & Pearl, D. (2013). Learning the language of evolution: Lexical ambiguity and word meaning in student explanations. *Research in Science Education, 43*, 1107–1133. doi:10.1007/s11165-012-9296-z

Reinagel, A., & Speth, E. B. (2016). Beyond the central dogma: Model-based learning of how genes determine phenotypes. *CBE-Life Sciences Education, 15*, ar4. doi:10.1187/cbe.15-04-0105

Robson, R. L., & Burns, S. (2011). Gain in student understanding of the role of random variation in evolution following teaching intervention based on Luria-Delbruck experiment. *Journal of Microbiology & Biology Education: JMBE, 12*, 3–7. doi:10.1128/jmbe.v12i1.272

Rosengren, K. S., Brem, S. K., Evans, E. M., & Sinatra, G. M. (2012). *Evolution challenges: Integrating research and practice in teaching and learning about evolution*. New York, NY: Oxford University Press.

Ross, P. M., Taylor, C. E., Hughes, C., Whitaker, N., Lutze-Mann, L., Kofod, M., & Tzioumis, V. (2010). Threshold concepts in learning biology and evolution. *Biology International, 47*, 47–52.

Schluter, D. (2000). *The ecology of adaptive radiation.* Oxford, United Kingdom: Oxford University Press.

Schuchardt, A. M., & Schunn, C. D. (2016). Modeling scientific processes with mathematics equations enhances student qualitative conceptual understanding and quantitative problem solving. *Science Education, 100*, 290–320. doi:10.1002/sce.21198

Shtulman, A. (2006). Qualitative differences between naïve and scientific theories of evolution. *Cognitive Psychology, 52*, 170–194. doi:10.1016/j.cogpsych.2005.10.001

Smith Jr, E. V. (2000). Metric development and score reporting in Rasch measurement. *Journal of Applied Measurement, 1*, 303–326.

Smith, M. (2012). *Random variables.* Retrieved from http://www.ma.utexas.edu/users/mks/M358KInstr/RandomVariables.pdf (accessed 01 March 2017)

Smith, M. K., Wood, W. B., & Knight, J. K. (2008). The genetics concept assessment: a new concept inventory for gauging student understanding of genetics. *CBE-Life Sciences Education, 7*, 422–430. doi:10.1187/cbe.08-08-0045

Speth, E. B., Shaw, N., Momsen, J., Reinagel, A., Le, P., Taqieddin, R., & Long, T. (2014). Introductory biology students' conceptual models and explanations of the origin of variation. *CBE-Life Sciences Education, 13*, 529–539. doi:10.1187/cbe.14-02-0020

Spindler, L., & Doherty, J. (2009). Assessment of the teaching of evolution by natural selection through a hands-on simulation. *Teaching Issues and Experiments in Ecology, 6*, 1–20.

Taylor, C. (2006). Threshold concepts in biology: do they fit the definition? In J. H. Meyer & R. Land (Eds.), *Overcoming barriers to student understanding: Threshold concepts and troublesome knowledge* (pp. 87–99). Abington, United Kingdom: Routledge.

Taylor, C., & Cope, C. (2007). Are there multiple thresholds in the concept of evolution, and can they be identified using dimensions of variation? *Proceedings of the Symposium Science Teaching and Learning Research*, September 27 & 28, 2007, UniServe Science: Sydney, 101–106.

Tibell, L. A., & Harms, U. (2017). Biological principles and threshold concepts for understanding natural selection. *Science & Education, 26*, 953–973. doi:10.1007/s11191-017-9935-x

VBIO (Verband Deutscher Biologen und biowissenschaftlicher Fachgesellschaften e.V.). (2006). *Studienführer Biologie: Biologie - Biochemie - Biotechnologie – Biomedizin. [Study guide biology: biology – biochemistry – biotechnology - biomedicine]*. Heidelberg, Berlin, Germany: Spektrum Akademischer Verlag.

Wagner, A. (2012). The role of randomness in Darwinian evolution. *Philosophy of Science, 79*, 95–119. doi:10.1086/663239

Weber, F., & Mathea, B. (2008). *SINUS-TRANSFER: Zufällige Erscheinungen - Hinweise zur Umsetzung der Leitidee Daten und Zufall im rheinland-pfälzischen Rahmenlehrplan Mathematik [SINUS-TRANSFER: Random phenomena – Tips for implementation of the central idea data and coincidence of the curriculum mathematics in Rhineland-Palatinate]*. Mainz, Germany: Ministerium für Bildung, Wissenschaft, Jugend und Kultur.

Wilson, M., de Boeck, P., & Carstensen, C. H. (2008). Explanatory item response models: A brief introduction. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of Competencies in Educational Contexts* (pp. 91–120). Göttingen, Germany: Hogefe & Huber.

Wright, B. D., & Linacre, M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*, 370.

Wright, B. D., & Mok, M. (2000). Understanding Rasch measurement: Rasch models overview. *Journal of Applied Measurement, 1*, 83–106.

Wu, M., Adams, R., Wilson, M., & Haldane, S. (2007). *ACER ConQuest 2.0: General item response modelling software [computer program manual]*. Camberwell, Australia: Australian Council for Educational Research.

Zander, C., Boustedt, J., Eckerdal, A., McCartney, R., Moström, J.-E., Ratcliffe, M., & Sanders, K. (2008). Threshold concepts in computer science: a multi-national empirical investigation. In R. Land, J. H. F. Meyer, & J. Smith (Eds.), *Threshold concepts within the disciplines* (pp. 102–118). Rotterdam, The Netherlands: Sense.

4.7 Supplemental Material

4.7.1 Overview of compulsory modules

Table 4.7 Compulsory modules (genetics, evolution, ecology, molecular biology, and cell biology) of six universities that sets of our participants attended, with brief synopses of subject matter

University	Students	Bachelor's program 1st year	2nd year	3rd year
Europa-Universität Flensburg	preservice biology teachers	-------	***Evolution and Functional Morphology (3 or 5. Sem.)*** (1) Evolutionary theories and mechanisms of evolution, (2) phylogeny of life including theories about genesis of life, and (3) biomechanical knowledge about structure and function of selected bodies of vertebrates. *Ecology (4. Sem.)* (1) Basics of ecology	***Evolution and Functional Morphology (3. or 5. Sem.)*** content see left
Technical University of Munich	biology majors	*Genetics (2. Sem.)* (1) Structure of genes and genomes, (2) gene expression (transcription and translation), (3) passing on of genetic information, (4) genetic recombination in eukaryotes and bacteria, (5) recombinant DNA and genetic technology, (6) genomic, (7) mutation and genetic analyses of complex biological processes, and (8) regulation of gene expression. *Cell Biology (2. Sem.)* no content required	*Ecology (4. Sem.)* (1) Abiotic factors, (2) population and sexuality, (3) communication systems, (4) population in time and space, (5) population dynamics, (6) growth models, (7) intra and inter specific concurrence, (8) zoogeography, (9) ecosystems of the world, (10), global change, and (11) invasive species.	***Evolution, Biodiversity and Biogeography (6. Sem.)*** (1) Basics of evolution, (2) population and speciation, (3) evolution of plants, (4) evolution of animals, (5) biodiversity, extinction and climate change, (6) genetic diversity, (7) sexual selection, (8) basics of biogeography, (9) plant geography, (10) symbiosis, and (11) animal geography.
Universität Hamburg	biology majors	*Genetics and Molecular Biology (2. Sem.)* (1) Classical and formal genetics (Mendelian inheritance, population genetics), (2) cytogenetics, (3) human genetics, (4) structure and function of nucleic acid (replication, transcription, translation, mutation, recombination), (5) gene regulation, (6) developmental genetics, and (7) methods of molecular biology and genetic technology.	*Genetics and Molecular Biology (3. Sem.)* content see left *Ecology (4. Sem.)* (1) Function, principles and methods of ecology, (2) recording and investigation of species in their environment, and (3) abiotical factors.	-------
Carl von Ossietzky Universität Oldenburg	biology majors	*Microbiology and Cell Biology (2. Sem.)* (1) Molecules of life, (2) energy and enzymes, (3) central metabolism, (4) breathing, (5) photosynthesis, (6) anaerobic metabolism, (7) chemolithotrophy, (8) prokaryotic and eukaryotic cell structures, (9) microbial diversity, (10) importance of microorganisms for human beings, plants, animals, biotechnology and earth system, (11) signal transmission and communication between cells, (12) meiosis, mitosis, Mendelian inheritance, and chromosomal and molecular basis of inheritance, (13) replication, transcription, translation, (14) genomic organization, and (15) mutation and repair.	*Genetics (3. Sem.)* (1) General and molecular genetics, (2) mechanisms of mutation, recombination, DNA repair, and regulation of transcription, (3) quantitative experiments with prokaryotes and eukaryotes, and (4) human genome project and personalized medicine.	-------

(Continued)

Table 4.7 (Continued)

University	Students	Bachelor's program 1st year	2nd year	3rd year
Carl von Ossietzky Universität Oldenburg	preservice biology teachers	*Microbiology and Cell Biology (2. Sem.)* content see above	-------	-------
Christian-Albrechts-Universität zu Kiel	biology majors	*Basics of Zoology and Cell Biology (1. Sem.)* (1) Blueprint of representatives of the important large animal groups, (2) functional units of animal organism, (3) basic knowledge of construction and function of the animal cell, and (4) evolution of animal body structures.	*Ecology (3. Sem.)* (1) Influence of environmental factors: radiation, temperature, humidity/water availability, (2) energy balance of animals and plants, (3) resistance and acclimatization, (4) host parasite and predator-prey interactions, competition, and gender conflicts, and (5) mechanisms of evolution in populations *Cell Biology Animal (3. Sem.)* (1) Simple cell biology and molecular biology techniques, (2) experimental handling and phenomenological observation of different cell types and invertebrate organisms under different experimental conditions and under adequate control, and (3) technics: light microscopy, fluorescence microscopy, polymerase chain reaction. *Cell Biology Plant (4. Sem.)* (1) Fluorescence- and electron microscopy of plant cell, (2) protein biochemical methods: electrophoresis, density gradient centrifugation, and (3) in situ hybridization *Genetics and Microbiology (4. Sem.)* (1) Classical genetics, (2) cytogenetics, (3) human genetics, (4) molecular genetics (DNA, RNA, genomes, replication, transcription, translation, gene regulation, epigenetics), (5) recombination, (6) mutation, (7) gene technology, (8) development, (9) basics of microbiological methods (microscopy, enrichment, cultivation), (10) morphological and physiological differentiation of microorganisms (Gram-staining, antibiotics), and (11) genetic exchange between microorganisms.	
	preservice biology teachers	*Basics of Zoology and Cell Biology (1. Sem.)* content see above	*Ecology (3. Sem.)* content see above *Cell Biology Animal (4. Sem. or Plants 5. Sem.)* content see above	*Cell Biology Plant (5. Sem. or Animals 4. Sem.)* content see above *Genetics and Microbiology (6. Sem.)* content see above

References

Carl von Ossietzky Universität Oldenburg (2016). *Veranstaltungsverzeichnis - Campusmanagementsystem Stud.IP [online lecture course schedule – campus management system Stud.IP]*. Retrieved from https://elearning.uni-oldenburg.de/plugins.php/veranstaltungsverzeichnis/verzeichnis/module (accessed 01 March 2017)

Christian-Albrechts-Universität zu Kiel (2016). *Studieninformationsblatt Biologie [Study information sheet of biology]*. Retrieved from https://www.studium.uni-kiel.de/de/studienangebot/studienfaecher/biologie-ba (accessed 01 March 2017)

Europa-Universität Flensburg (2015). *Modulkatalog Biologie, B.A., Bildungswissenschaften [Module manual biology, bachelor of arts, educational sciences]*. Retrieved from https://www.uni-flensburg.de/fileadmin/content/portale/studium_und_lehre/dokumente/po-studiengaenge/bachelor-of-arts/bildungswissenschaften/2015/modulkataloge/mk-babw-po2015-biologie.pdf (accessed 01 March 2017)

Technical University of Munich (2013). *Modulhandbuch [Module manual]*. Retrieved from http://sf-biowiss.wzw.tum.de/fileadmin/Dokumente/BIO/ModulhandbuchMSc/Modulhandbuch_Bio_MSc_FPSO2007_13.11.13.pdf (accessed 01 March 2017)

Universität Hamburg (2016). *Modulhandbuch Bachelorstudiengang Biologie [module manual of bachelor's program biology]*. Retrieved from https://www.biologie.uni-hamburg.de/studium/bachelor/bachelor-biologie.html (accessed 01 March 2017)

Universität Hildesheim (2014). *Studien- und Prüfungsornungen [Study and examination regulations]*. Retrieved from https://www.uni-hildesheim.de/studium/studienangebot/studienangebot/ordnungen2faecherbachelor/ (accessed 01 March 2017)

4.7.2 The Randomness and Probability test in the context of Evolution (RaProEvo)

*Note: * = correct answer; contact author for guide of free-response items*

E01 (Source: cf. Fenner, 2013)
You observe the following situation. In the South Pole, a male penguin with normally thick plumage and a female penguin with very thick plumage generate an offspring. This one also has very thick plumage. **How can this be explained with the theory of evolution?**

- □ Because it is very cold at the South Pole, the offspring's body had to receive the thickest plumage that the parental genetic pool had to offer.
- □ *The offspring was lucky. It could also have got less thick plumage.
- □ The plumage grew stronger, because otherwise the offspring would have frozen to death.

E02 (Source: cf. Fenner, 2013)
The litter of a cheetah includes two offspring: one with an advantageous mutation and one without this mutation. **What can you say about survival of the two offspring?**

- □ The offspring with the advantageous mutation will survive.
- □ The offspring without the advantageous mutation will survive.
- □ *Both offspring may either survive or die.

E03 (Source: cf. Robson and Burns, 2011)
Milkweed leaves are toxic to most insects, but a subspecies of beetle has been found that can eat milkweed leaves with no ill effects. **Which of the following do you think best explains the evolution of some beetles' ability to eat milkweed leaves without getting sick, even though eating milkweed leaves kills other, closely-related beetles?**

- □ Eating milkweed makes beetles produce an enzyme that destroys milkweed toxins, so that the more milkweed the beetles eat, the less it bothers them.
- □ *A few beetles just happened to make an enzyme that destroys milkweed toxins. When beetles eat milkweed leaves, only those that happen to make this enzyme are able to survive.
- □ Beetles become immune to milkweed toxin the more milkweed leaves they eat. Then, when the beetles reproduce, they pass their immunity to milkweed toxin on to their offspring.

E04 (Source: *EvoVis* project)
There is a pack of 17 grey and 13 brown wolves. The wolves hunt and run along a canyon. Suddenly, some stones fall and hit one of the wolves, which then falls down the canyon. **Mark the probable color of the wolf that fell:**

- □ Grey.
- □ Brown.
- □ *Grey or brown.

E05 (Source: cf. Campbell et al., 2006)
Explain why the statement "Evolution through natural selection is a *random process.*" is wrong.

E06 (Source: *EvoVis* project)
The human eye is composed, like a camera, of many parts, all of which are needed for the eye to work. **Which of the following statements about the eye do you find most credible?**

- □ If you remove one part of the eye it will stop functioning. Therefore, the eye cannot have evolved through a gradual process. It must have appeared in one step as a functioning unit.
- □ *Since small improvements are often favored by natural selection, the eye probably evolved gradually.
- □ The probability that such a complex organ as the eye can evolve by mere chance is so small there must be some thought behind it.
- □ Since animals need to see in order to find food they have evolved eyes.
- □ The eye has evolved solely by chance.

E07 (Source: cf. Campbell and Reece, 2011)
Every person is genetically unique. **Mark the answer that best explains the most common cause of this uniqueness according to the theory of evolution:**

- □ Random mutations that have occurred in previous generation.
- □ *New combinations of alleles during sexual reproduction.
- □ Genetic drift associated with small populations.
- □ Geographical variability within the population.
- □ Environmental influences.

E08 (Source: *EvoVis* project)
The gender of a child depends on whether the sperm involved in his or her conception carries the father's X- or Y-chromosome. Assume that there are equal proportions of sperm carrying these chromosomes, and by chance just one fuses with the female gamete (ovum, carrying an X chromosome from the mother), resulting in the conception of either a girl (XX) or a boy (XY).

a. Given the information above, mark the statement that appropriately describes the likelihood of conception of a girl or a boy:

- □ A girl is more likely to be conceived than a boy.
- □ A boy is more likely to be conceived than a girl.
- □ *Girls and boys are equally likely to be conceived.

b. Respond to the statement "You can be sure to get at least one girl if you give birth to three children."

E09 (Source: *EvoVis* project)
More sheep than people live in New Zealand. A hundred sheep stand in a pasture with no shelter: 68 with a mutation that is advantageous for their survival and 32 without this mutation. Suddenly, there is a flash of lightning. **Mark which sheep could be hit by the flash:**

- □ A sheep with the advantageous mutation for its survival.
- □ A sheep without the advantageous mutation for its survival.
- □ *A sheep either with or without the advantageous mutation for its survival.

E10 (Source: cf. Fenner, 2013)
In former times, half of the foxes in Northern Europe had white fur, while the other half had brown fur. Today, nearly all foxes have white fur. **How can this change be explained by the theory of evolution?**

- □ *Foxes with lighter fur could hunt prey more easily, produce more offspring, and pass on their genetic basis for fur color to more descendants.
- □ The foxes wanted to improve their adaptation to the surrounding landscape by enhancing their camouflage.
- □ The foxes recognized that they needed white fur for their survival.

E11 (Source: *EvoVis* project)
In an experiment four populations of white lab mice are observed. Sometimes a mutation occurs that changes their white fur to brown. **Mark which of the statements best describes what you can tell about changes in the populations after one generation:**

- □ The mutation will occur in every population, so brown mice will appear in all four populations.
- □ The mutation will occur and brown mice will appear in two of the four populations.
- □ *It is impossible to tell whether or not the mutation will occur in the populations.
- □ No mutation will appear in any of these populations, because they are lab mice and do not mutate.

E12 (Source: cf. Bowling et al., 2008)
Mutations in DNA occur in the genomes of all organisms, including humans. **Why are mutations most important according to the theory of evolution?**

- □ Mutations allow the production of new genes in individuals.
- □ Mutations allow the production of new enzymes in individuals.
- □ Mutations are sources of new cells for individuals.
- □ *Mutations are sources of genetic variation for future generations.
- □ Mutations allow the production of new chromosomes for future generations.

E13 (Source: *EvoVis* project)
In a very dry area a sudden violent storm leads to flooding of the whole area. **Mark the statement that best describes which animals and plants survive this catastrophic event according to the theory of evolution:**

- □ Only individuals with potentially advantageous genetic combinations for the new surroundings can survive the catastrophe, all others die.
- □ *Individuals both with and without potentially advantageous genetic combinations may survive the catastrophe.
- □ All individuals will survive, because the flood is only a brief event so it will not affect their survival.

E14 (Source: cf. Klymkowsky et al., 2010)
Why is a catastrophic global event regarded as a random phenomenon?

- □ Because undesirable genes are removed.
- □ Because new genes originate.
- □ *Because only some species survive the event.
- □ Because there are only brief effects, which disappear over time.

E15 (Source: *EvoVis* project)
Explain the meaning of the statement "Mutations are random".

E16 (Source: *EvoVis* project)
The following information is given:

Species	**Bodyweight**	**Generation time***
Mouse	14 g	2 months
Wolf	40 kg	2 years
Viper	200 g	5 years
Elephant	5000 kg	14 years

* The average time between birth and reproductive maturity in a given population or taxon.

Mark which of the four species can adapt most readily through natural selection to sudden, drastic environmental changes:

- □ *Mouse
- □ Wolf
- □ Viper
- □ Elephant

E17 (Source: cf. Campbell et al., 2006)
A large part of variability in the fur coloring and pattern in every generation of wild mustangs is probably due to…

- □ … mutations that occurred in the previous generation.
- □ *… recombination of alleles.
- □ … genetic drift associated with small populations.
- □ … geographical variability within the population.
- □ … environmental influences.

E18 (Source: *EvoVis* project)
Assume that two identical populations of laboratory mice are placed in two different habitats (G and W). Allele 1 is more advantageous in habitat G than in habitat W, while allele 2 is equally advantageous in both habitats (G and W). **Mark what will happen in your opinion to the frequency of allele 1 <u>after 100 generations</u> in both populations:**

- □ *It will become higher in the population living in habitat G than in the population living in habitat W.
- □ It will become higher in the population living in habitat W than in the population living in habitat G.
- □ It will rise in both the population living in habitat G and the population living in habitat W.

E19 (Source: cf. Green, 1982; *EvoVis* project)
a. Match each of the four notions with one of the five statements (A-E). Note, you can match statements with more than one notion:

Statements

A: Cannot happen.
B: Cannot happen very often.
C: Happens rather often.
D: Happens almost always.
E: Always happens.

Notions

a.	very likely	*D
b.	unlikely	*B
c.	likely	*C or D
d.	not very likely	*B

b. Now, match the three biological examples with one of the five statements (A-E). Again, you can match statements with more than one notion:

Biological examples

The descendants of sexually reproducing organisms are genetically identical to their parents.	*A
A bottleneck reduces the genetic variability of a population.	*C or D
Non-resident species that are introduced into an environment in which the climatic conditions differ from those in their original environment can spread and bring local species close to extinction.	*B

References

Bowling, B. V., Acra, E. E., Wang, L., Myers, M. F., Dean, G. E., Markle, G. C., . . . Huether, C. A. (2008). Development and evaluation of a genetics literacy assessment instrument for undergraduates. *Genetics, 178*, 15–22.

Campbell, N. A., & Reece, J. B. (2011). *Campbell Biologie - Gymnasiale Oberstufe, Übungsbuch [Campbell biology – Secondary level, workbook]*. Munich, Germany: Pearson Schule.

Campbell, N. A., Reece, J. B., & J, M. (2006). *Biologie [Biology]* (Vol. 6.). Munich, Germany: Pearson Studium.

Fenner, A. (2013). *Schülervorstellungen zur Evolutionstheorie: Konzeption und Evaluation von Unterricht zur Anpassung durch Selektion [Students conceptions* *about the theory of evolution: Design and evaluation of a teaching sequence about adaption through selection]* (doctoral dissertation). Justus-Liebig-University, Giessen, Germany. Retrieved from http://geb.uni-giessen.de/geb/volltexte/2013/9250/ (accessed 01 March 2017)

Green, D. (1982). *Probability concepts in 11-16 year old pupils. Report of research sponsored by the Social Science Research Council.* Loughborough, England: CAMET, University of Technology.

Klymkowsky, M. W., Underwood, S. M., & Garvin-Doxas, R. K. (2010). Biological Concepts Instrument (BCI): A diagnostic tool for revealing student thinking. *arXiv preprint* arXiv:1012.4501.

Robson, R. L., & Burns, S. (2011). Gain in student understanding of the role of random variation in evolution following teaching intervention based on Luria-Delbruck experiment. *Journal of Microbiology & Biology Education: JMBE, 12*, 3–7.

4.7.3 The Randomness and Probability test in the context of Mathematics (RaProMath)

Note: * = correct answer; contact author for guide of free-response items

M01 (Source: cf. Garfield, 2003)
Two containers (A and B) are filled with the following numbers of red and blue marbles:

Containers	Red	Blue
A	6	4
B	60	40

Each container is shaken vigorously. **Which container offers the highest chance of pulling out a blue marble?**

- □ Container A.
- □ Container B.
- □ *There are equal chances of getting a blue marble from container A and container B.

M02 (Source: cf. Green, 1982)
In an experiment 12 coins are tossed together in the air and land on a table. It is possible that one of the following results may occur:

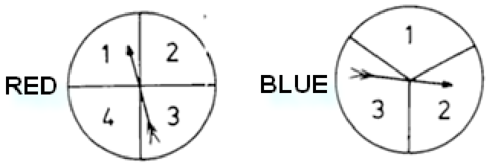
Result 1: Two heads and 10 tails
Result 2: Five heads and seven tails
Result 3: Six heads and six tails
Result 4: Seven heads and five tails

On which result do you bet <u>within a single round</u>? Explain your answer:

M03 (Source: cf. Green, 1982)
A small, round disc is red on one side and green on the other. The disc is thrown in the air with the red side upwards, turns several times and lands on a table. **Mark the right statement:**

- □ Only the red side will be facing upwards
- □ Only the green side will be facing upwards.
- □ *Either the red or green side will be facing upwards.

M04 (Source: cf. Green, 1982)
In the following diagram you see two wheels of fortune (red and blue) with their arrows at rest. You can spin the arrow of either wheel, and win $50 if it lands on a three.



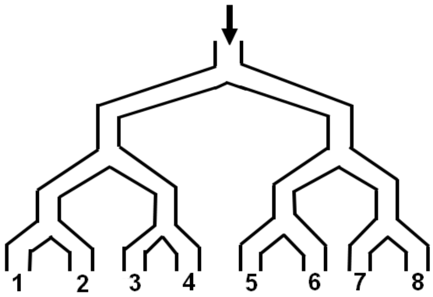
Which wheel should you choose to maximize your chance to win?

- □ The red wheel.
- □ *The blue wheel.
- □ The chance is equally high with both discs.

M05 (Source: *EvoVis* project)
Anna and Moritz are gambling by throwing a dice six times. Anna receives $2 from Moritz if a 5 or 6 comes up, while Anna has to pay Moritz $1 if a 1, 2, 3, or 4 comes up.
Mark if the chance of profit is:

- □ Higher for Anna than for Moritz.
- □ Higher for Moritz than for Anna.
- □ *Equally high for Anna and Moritz.
- □ Impossible to judge.

M06 (Source: cf. Green, 1982)
A marble is dropped into an apparatus with forked channels, illustrated below.



Mark where the marble will come out:

- □ Channel 1 or 8.
- □ Channel 3, 4, 5 or 6.
- □ Channel 1, 3, 5 or 7.
- □ *Channel 1, 2, 3, 4, 5, 6, 7 or 8.
- □ Channel 1, 2, 7 or 8.

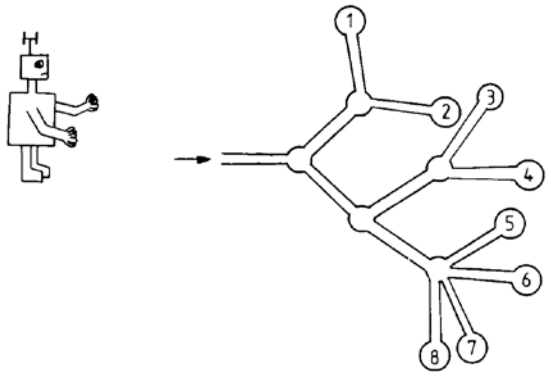
M07 (Source: Garfield, 2003)
When two dice are simultaneously thrown it is possible that one of the following two results occurs:
Result 1: A 5 and a 6 are obtained.
Result 2: A 5 is obtained twice.
Select the response that you agree with the most:

- □ *There is more chance of obtaining result 1.
- □ There is more chance of obtaining result 2.
- □ The chances of obtaining each of these results are equal.

M08 (Source: Green, 1982)
A mathematics class has 13 boys and 16 girls in it. Each pupil's name is written on a slip of paper. All the slips are put in a hat. The teacher picks out one slip without looking.
Thick the correct sentence:

- □ The name is more likely to be a boy than a girl.
- □ *The name is more likely to be a girl than a boy.
- □ It is just as likely to be a girl as a boy.

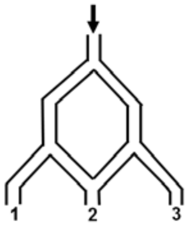
M09 (Source: cf. Green, 1982)
A robot is put in a maze and begins to explore it. At every junction the robot must choose a path, and at the end of every path there is a trap (see picture).



Mark the trap (or traps) where the robot is most likely to be trapped:

- □ *Trap 1
- □ *Trap 2
- □ Trap 3
- □ Trap 4
- □ Trap 5
- □ Trap 6
- □ Trap 7
- □ Trap 8

M10 (Source: cf. Green, 1982)
A marble is dropped into an apparatus with forked channels, illustrated below.



Mark where the marble will come out:

- □ Channel 1.
- □ Channel 2 or 3.
- □ Channel 1 or 3.
- □ *Channel 1, 2 or 3.
- □ Channel 2.
- □ Channel 3.
- □ Channel 1 or 2.

M11 (Source: cf. Garfield, 2003)
Five sides of a faire cube are black and the other is white. The cube is thrown six times and the following results are possible:
Result 1: A black side will come up in five throws and the white side in one throw.
Result 2: A black side will come up in all six throws.
Choose the correct statement:

- □ *Result 1 is more likely than result 2.
- □ Result 2 is more likely than result 1.
- □ Results 1 and 2 are equally likely.

M12 (Source: cf. Jones et al., 1997)
A container is filled with five green, three red and two blue marbles then vigorously shaken. A red marble is blindly pulled out and then put back into the container. The container is shaken again. Then another marble is pulled out. **Choose the correct statement**:

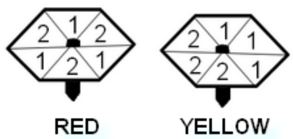
- □ *The marble pulled out is most likely to be green.
- □ The marble pulled out is most likely to be red.
- □ The marble pulled out is most likely to be blue.
- □ The marble pulled out is equally likely to be green, red or blue.

M13 (Source: cf. Garfield, 2003)
If three faire dice are thrown simultaneously, one of the following results may occur:
Result 1: Three 5s may come up
Result 2: Two 5s and a 3 may come up
Result 3: A 5, 3 and 6 may come up
Choose the answer you agree with most strongly:

- □ Result 1 has the highest chance.
- □ Result 2 has the highest chance.
- □ *Result 3 has the highest chance.
- □ The chance of all three results is equally high.

M14 (Source: cf. Jones et al., 1997)
A container is filled with five green, three red and two yellow marbles then vigorously shaken. One of the marbles is blindly pulled out and examined. It is green. The marble is put back into the container, which is shaken again then another marble is pulled out. **What is this marble's color? Explain your answer:**

M15 (Source: cf. Green, 1982)
Segments of two symmetrical six-sided spinning tops (one red and one yellow) are marked with the numbers 1 and 2, as illustrated below. You win $50 if one of them lands resting on the side of a segment marked with a 2.



Which spinning top provides the best chance of winning?'

- □ The red one.
- □ The yellow one.
- □ *The chances are equally high with both spinning tops.

M16 (Source: *EvoVis* project)
If a fair coin is tossed, the likelihood of getting ‘tails’ (i.e. the reverse side facing up) is ½. In three consecutive throws the result is ‘tails’. **Which of the four statements applies for the next throw?**

- □ ‘Heads’ (the obverse side facing up) is most likely
- □ ‘Tails’ is most likely.
- □ *‘Heads’ and ‘tails’ are equally likely.
- □ More information is needed to answer the question.

M17 (Source: *EvoVis* project)
When throwing a faire dice a number between 1 and 6 will face upwards. When is the result of the next throw predictable?

- □ After 100 throws.
- □ After 10,000 throws.
- □ *The result is never predictable.
- □ After 500 throws.
- □ After 50,000 throws.

M18 (Source: cf. Green, 1982)
Two containers (A and B) are filled with the following numbers of red and blue marbles:

Containers	Red	Blue
A	12	4
B	20	10

Each container is shaken vigorously. You want to pull out a blue marble. **Which statement is correct?**

- □ There is a higher chance of pulling a blue ball from container A than from container B.
- □ *There is a higher chance of pulling a blue ball from container B than from container A
- □ There are equal chances of pulling a blue ball from container A and container B.

M19 (Source: Garfield, 2003)
Half of all newborns are girls and half are boys. Hospital A records an average of 50 births a day. Hospital B records an average of 10 births a day. **On a particular day, which hospital is more likely to record 80% or more female births?**

- □ Hospital A (with 50 births a day).
- □ *Hospital B (with 10 births a day).
- □ The two hospitals are equally likely to record such an event.

M20 (Source: cf. Herget et al., 2009)
You have five white, five black and five grey marbles. **Describe how a container has to be filled so that the likelihood of pulling out a white marble is 3/10:**

M21 (Source: cf. Weber and Mathea, 2008)
In a roulette wheel there are 17 segments with numbers from 1 to 17. You win if you spin and the ball lands in a segment with an even number, and lose otherwise. **Mark the correct statement:**

- □ *You are more likely to lose than to win.
- □ You are more likely to win than to lose.
- □ Winning and losing are equally likely.

M22 (Source: cf. Jones et al., 1997)
A container is filled with five green, three red and two yellow marbles and vigorously shaken. One of the marbles is blindly pulled out and examined. It is red. The marble <u>is not put back into</u> the container. The container is shaken again and another marble is pulled out. **Mark the correct statement:**

- □ *The marble pulled out is most likely to be green.
- □ The marble pulled out is most likely to be red.
- □ The marble pulled out is most likely to be yellow.
- □ The marble pulled out is equally likely to be green, red or yellow.

M23 (Source: *EvoVis* project)
Many adults study lottery statistics every week to forecast the next round's winning numbers. The following pictures show lottery coupons with three sets of forecasts.

1	2	3	4	5	~~6~~	7
8	~~9~~	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	~~25~~	26	27	28
~~29~~	30	31	32	33	~~34~~	35
36	37	38	~~39~~	40	41	42
43	44	45	46	47	48	49

Lottery coupon A

1	2	3	4	5	6	7
~~8~~	9	10	11	12	13	14
15	16	17	~~18~~	19	20	21
22	23	24	25	~~26~~	~~27~~	~~28~~
29	30	31	32	33	34	35
36	37	38	39	40	41	42
43	44	45	46	47	48	~~49~~

Lottery coupon B

1	2	3	4	5	~~6~~	~~7~~
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	~~28~~
~~29~~	30	31	32	33	~~34~~	35
36	37	38	39	40	41	42
~~43~~	44	45	46	47	48	49

Lottery coupon C

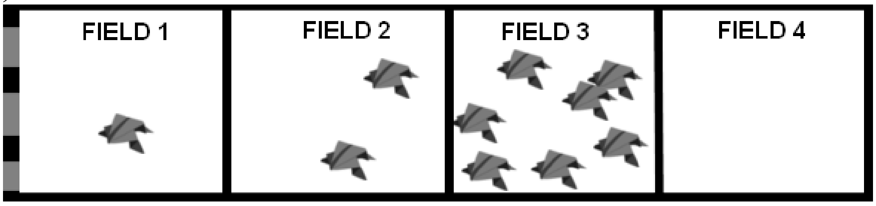
Which of these lottery coupons could win?

- □ Lottery coupon A
- □ Lottery coupon B
- □ Lottery coupon C
- □ Lottery coupon A and C
- □ *Lottery coupon A, B, or C

M24 (Source: cf. Garfield, 2003)
When you throw a coin five times possible sequences of 'heads' (H, obverse side facing up) and 'tails' (T, reverse side facing up) include the following:
Sequence 1: H H H T T
Sequence 2: T H H T H
Sequence 3: H T H H H
Sequence 4: H T H T H
Which of these sequences is most likely?

- ◻ Sequence 1
- ◻ Sequence 2
- ◻ Sequence 3
- ◻ Sequence 4
- ◻ *Sequence 1, 2, 3 or 4

M25 (Source: cf. Eichler and Vogel, 2012)
Class 4B records the lengths of 10 paper frogs' jumps and which of four zones they land in, as illustrated below.



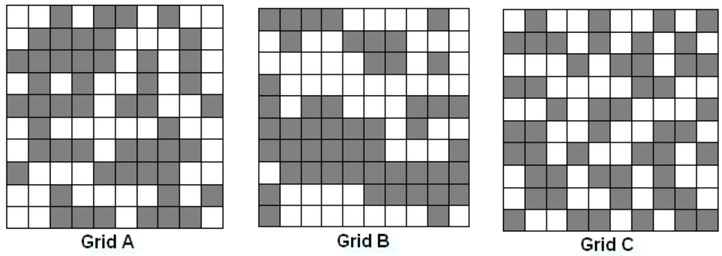
Where could an eleventh frog land?

- ◻ Field 1
- ◻ Field 2
- ◻ Field 3
- ◻ Field 4
- ◻ *Field 1, 2, 3 or 4

M26 (Source: *EvoVis* project)
In an experiment, the life span of 10,000 light bulbs was examined; 80% lasted longer than 100 hours, while 5% stopped working within 20.4 hours. **Mark how long a freshly installed light bulb will probably last:**

- ◻ More than 100 hours.
- ◻ At most 20.4 hours.
- ◻ Between 20.4 and 100 hours.
- ◻ *The life span of the light bulb cannot be predicted.

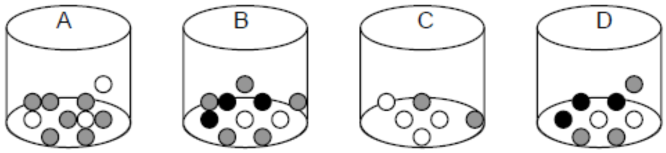
M27 (Source: cf. Falk and Konold, 1997)
The following diagram shows three 10x10 grids, each with 50 white and 50 grey squares.



Mark which of the grids may have originated from random placement of white and grey squares:

- □ Grid A
- □ Grid B
- □ Grid C
- □ Grid A and C
- □ *Grid A, B and C

M28 (Source: cf. Herget et al., 2009)
Four containers are placed on a table, each containing different numbers of uniformly sized marbles, with varying proportions of colors, as shown in the following picture. If you pull out a white marble, you will win.



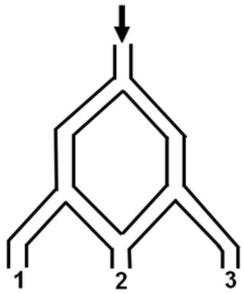
Which container provides the best chance of winning?

- □ Container A.
- □ Container B.
- □ *Container C.
- □ Container D.

M29 (Source: *EvoVis* project)
A coin is tossed five times and every time it lands with the head facing upwards. **Mark the correct sentence:**

- □ Next time the coin will land with the head facing upwards.
- □ Next time the coin will land with the tail (reverse side) facing upwards.
- □ *Next time the coin will land with either the head or tail facing upward.

M30 (Source: cf. Green, 1982)
Numerous marbles are dropped into an apparatus with forked channels, illustrated below.



With which statement do you agree most strongly?

- □ The same numbers of marbles will come out from channels 1, 2 and 3.
- □ *Twice as many marbles will come out from channel 2 than from channel 1 or 3.
- □ Half of the marbles will come out from channel 1, while the other half will come out from channel 3.

M31 (Source: cf. Green, 1982)
To find out whether a thumb tack (drawing pin) more often lands on its back  or side , a seminar leader empties a pack of 100 on a table. 68 fall on their back, and 32 on their side. The experiment is repeated another three times by the seminar leader. The results are:

1. Back, 64; side, 36.
2. Back, 70; side, 30.
3. Back, 66; side, 34.

If the seminar leader carried out the experiment once more, the following results would be possible:

Result 1. Back, 36; side, 64.
Result 2. Back, 63; side, 37.
Result 3. Back, 50; side, 50.
Result 4. Back, 85; side, 16.

Which of the results is most likely?

- □ Result 1
- □ *Result 2
- □ Result 3
- □ Result 4
- □ Result 1, 2, 3 and 4

M32 (Source: cf. Weber and Mathea, 2008)
Three hunters (Adam, Ben and Chris) shoot at a duck simultaneously. **Decide for each of the following cases whether or not the duck has a chance to survive:**

	Duck has a chance to survive	Duck has no chance to survive
Adam, Ben and Chris have hit rates of 30%, 50% and 20%, respectively.	□*	□
Adam, Ben and Chris have hit rates of 40%, 50% and 30%, respectively.	□*	□
All three hunters have a hit rate of 20%	□*	□

M33 (Source: cf. Green, 1982)
A container is filled with unknown numbers of green, red and yellow marbles. You blindly pull out a marble and record the color. Then the marble is put back into the container and the container is vigorously shaken. You repeat the process four times, and every time the marble is red. **If another marble is taken what color will it be?**

- □ Red.
- □ Green.
- □ Yellow.
- □ *Red, green or yellow.
- □ Green or yellow.

References

Eichler, A., & Vogel, M. (2012). Basic modelling of uncertainty: young students' mental models. *ZDM Mathmatics Education, 44*, 841–854.

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review, 104*, 301–318.

Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal 2*, 22–38.

Green, D. (1982). *Probability concepts in 11-16 year old pupils. Report of research sponsored by the Social Science Research Council.* Loughborough, England: CAMET, University of Technology.

Herget, W., Kösters, E., & Merziger, P. (2009). *Stochastisches Denken – Lehren und Lernen [Stochastic thinking – Teaching and learning]*. Retrieved from http://disk.mathematik.uni-halle.de/lehrerseite/Stochastik_Reader_mit_Aufgaben_und_Loesungen_0909.pdf (accessed 01 March 2017)

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children‘s thinking in probability. *Educational Studies in Mathematics, 32*, 101–125.

Weber, F., & Mathea, B. (2008). *SINUS-TRANSFER: Zufällige Erscheinungen - Hinweise zur Umsetzung der Leitidee Daten und Zufall im rheinland-pfälzischen Rahmenlehrplan Mathematik [SINUS-TRANSFER: Random phenomena – Tips for implementation of the central idea data and coincidence of the curriculum mathematics in Rhineland-Palatinate]*. Mainz, Germany: Ministerium für Bildung, Wissenschaft, Jugend und Kultur.

4.7.4 Academic self-concept items

(Adapted from Braun et al., 2008)

1. I can see the connections and inconsistencies in …
2. I can give an overview of the topic of …
3. I can clearly present complicated issues of …
4. Now I see myself in the position to process a typical question of …
5. I can work out the contradictions and similarities of learning content (e.g., contradictions between different models or methods) of the subject area of …

References

Braun, E., Gusy, B., Leidner, B., & Hannover, B. (2008). Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). [The Berlin evaluation instrument for selfapproximated, student competence]. *Diagnostica, 54*, 30–42.

4.7.5 Item parameter estimates for the RaProEvo test

Table 4.8 Item parameter estimates for the RaProEvo test

Item	Discrimination (a)	Difficulty (b_1)
ovE01	0.54	0.523
ovE02	0.21	-0.567
ovE03	0.40	-0.727
ovE07	0.23	-1.366
ovE11	0.38	-1.118
ovE12	0.42	-0.420
ovE15	0.51	2.012
ovE17	0.40	0.003
adE04	0.19	-0.643
adE09	0.19	-1.366
rpE07	0.23	-1.366
rpE13	0.38	0.952
rpE16	0.35	-2.128
nsE06	0.35	-0.049
nsE10	0.33	-1.236
nsE16	0.35	-2.128
nsE18	0.33	0.208
peE08a	0.44	-0.641
peE08b	0.44	-0.223
peE19a	0.35	1.502
peE19b	0.34	3.291

Note. The first two letters stand for: *ov* origin of variation, *ad* accidental death (single events), *rp* random phenomena, *ns* process of natural selection, and *ps* probability of events. *E* represents the content of evolutionary theory, while *01* to *19* indicates the item number in the RaProEvo test and the last letter stands for *a* item 1 and *b* item 2 within a similar item task.

4.7.6 Item parameter estimates for the RaProMath test

Table 4.9 Item parameter estimates for the RaProMath test

Item	Discrimination (a)	Difficulty (b_1)
seM02	0.37	2.378
seM03	0.17	-2.923
seM06	0.35	-1.447
seM10	0.21	1.322
seM14	0.21	2.600
seM25	0.38	0.988
seM29	0.31	-1.954
seM33	0.37	-0.808
rpM17	0.11	-1.757
rpM23	0.42	-1.442
rpM26	0.41	1.191
rpM27	0.29	-0.310
prM01	0.41	-0.988
prM04	0.17	-2.922
prM08	0.23	-1.593
prM12	0.38	-0.648
prM15	0.44	-0.725
prM16	0.26	-1.087
prM18	0.31	-0.504
prM20	0.51	0.987
prM21	0.33	-0.806
prM22	0.40	-1.311
prM28	0.26	-2.908
peM05	0.34	1.741
peM07	0.25	3.105
peM09	0.29	2.102
peM11	0.20	0.159
peM13	0.22	2.717
peM24	0.40	-0.436
peM32	0.61	0.115
srM19	0.15	2.678
srM30	0.37	-0.505
srM31	0.28	0.737

Note. The first two letters stand for: *se* single events, *rp* random phenomena, *pr* probability as ratio, *pe* probability of events, and *sr* sample reasoning. *M* represents the content of mathematics, while *01* to *40* indicates the original item number.

5 STUDY 2: EVOSKETCH: SIMPLE SIMULATIONS FOR LEARNING RANDOM AND PROBABILISTIC PROCESSES IN EVOLUTION, AND EFFECTS OF INSTRUCTIONAL SUPPORT ON LEARNERS' CONCEPTUAL KNOWLEDGE[2]

Abstract

Students' knowledge of scientific principles of evolution is often inadequate, despite its recognized importance for understanding biology. Moreover, difficulties associated with underlying abstract concepts such as randomness and probability can hinder successful learning of evolutionary concepts. However, learning abstract concepts can be supported by visualizations, particularly (reportedly) simulations together with appropriate instructional support. Therefore, we have developed interactive, web-based simulation software called EvoSketch in efforts to help learners grasp the nature and importance of random and probabilistic processes in evolutionary contexts. Here we report an investigation of EvoSketch simulations' effectiveness. We compared time-on-task, perceived cognitive load directly after interventions, and knowledge test performance on three occasions of students using the simulations (with and without additional instructional support) and others using text-based learning of randomness and probability. In total, 267 German secondary school students participated in the study. Significant between-intervention differences were detected in students' perceived cognitive load and time-on-task. Further, learners using EvoSketch without additional support obtained higher RaProEvo (Randomness and Probability Test in the Context of Evolution) scores in follow-up tests than those using the text-based approach. Our results indicate that EvoSketch is an effective tool for learning and teaching concepts of randomness and probability in evolutionary contexts, particularly for fostering long-term understanding. However, use of the simulations together with additional interpretative support (worked example) or reflective support (reflective questions) did not increase students' performance, relative to the text-based approach. Possible reasons for this are discussed, and recommendations are made to incorporate such interventions in several lessons.

Keywords:

Evolution education, Threshold concepts, Randomness and probability, Web-based simulations, Instructional support, Secondary school students

---

[2] This is a pre-peer reviewed version of the following article: Fiedler, D., Tröbst, S., Großschedl, J. & Harms, U. (minor revisions, 08/2018). EvoSketch: Simple simulations for learning random and probabilistic processes in evolution, and effects of instructional support on learners' conceptual knowledge. *Evolution: Education and Outreach*. This article may be used for non-commercial purposes in accordance with BioMed Central (BMC) Terms and Conditions for Use of Self-Achieved Versions.

5.1 Introduction

Learners have well-documented problems with understanding and learning key scientific concepts, including energy (e.g., Hartley, Momsen, Maskiewicz, & D'Avanzo, 2012; Herrmann-Abell & DeBoer, 2018; Opitz, Neumann, Bernholt, & Harms, 2017; Wernecke, Schwanewedel, & Harms, 2018), genetics (e.g., Schmiemann, Nehm, & Tornabene, 2017; Tsui & Treagust, 2004, 2010; Venville, Gribble, & Donovan, 2005), and evolution (e.g., Gregory, 2009; Rector, Nehm, & Pearl, 2013, Rosengren, Brem, Evans, & Sinatra, 2012). A shared aspect of these scientific concepts is that spatial and/or temporal dimensions of associated processes and structures prevent their direct perception. Hence, they can only be understood on an imaginary level, like all concepts beyond humans' perceptual (especially visible) dimensions (Lakoff, 1987; Lakoff & Johnson, 1980). For instance, random mutations in DNA are important sources of variation in the key evolutionary process of natural selection (Heams, 2014). However, mutations are not visible to the naked human eye, although they can be visualized technologically (e.g., using DNA sequencing techniques). The consequent lack of possibility for students to observe these phenomena in everyday situations may result in misunderstanding of the importance of random processes in evolution (Garvin-Doxas & Klymkowsky, 2008). Furthermore, students tend to frequently misunderstand general abstract concepts that underlie biological processes like randomness and probability (Garvin-Doxas & Klymkowsky, 2008; Mead & Scott, 2010). Therefore, it may be essential to address underlying abstract concepts to overcome problems in learning evolution, and appropriate visualization could make these concepts tangible (Tibell & Harms, 2017).

Thus, researchers involved in the *EvoVis*-project (*EvoVis: Challenging Threshold Concepts in Life Science - enhancing understanding of evolution by visualization*) have developed interactive, web-based simulation software, called EvoSketch, which allows learners to explore random and probabilistic phenomena associated with the process of natural selection. The software generates a line (representing a reproducing organism) that is replicated by the user for 20 generations.

The main aim of the presented study was to test the effectiveness of EvoSketch for teaching and learning the roles of randomness and probability in evolutionary contexts (i.e., mutation and selection, respectively). A further aim was to identify the optimal kind of additional instructional support (if any) to use.

5.2 Rationale

5.2.1 Learning Evolution and Threshold Concepts

Evolution education research indicates several difficulties for learning the essential tenants of evolutionary theory (e.g., Bishop & Anderson, 1990; Gregory, 2009;

Kampourakis & Zogza, 2008). Moreover, students hold diverse alternative conceptions about evolutionary key and core concepts (e.g., Baalmann, Frerichs, Weitzel, Gropengießer, & Kattmann, 2004; Beggrow & Nehm, 2012; Bishop & Anderson, 1990; Nehm & Schonfeld, 2008; Opfer, Nehm, & Ha, 2012; Shtulman, 2006; Spindler & Doherty, 2009; Yates & Marek, 2015). Additionally, many words in science lessons such as adaptation or fitness also appear in other contexts or everyday language with slightly different meanings. This can confuse students and lead to misused scientific terminology (Rector, Nehm, & Pearl, 2013; To, Tenebaum, & Hogh, 2017). Thus, teachers and instructors should target these alternative conceptions and meanings, and openly address them to the students to cause cognitive conflicts; hence conceptual change is likely to occur (e.g., Sinatra, Brem, & Evans, 2008; Posner, Strike, Hewson, & Gertzog, 1982).

However, current research also mentioned learning difficulties with those evolutionary concepts that are strongly related to underlying abstract concepts like randomness and probability, so-called threshold concepts (Mead & Scott, 2010; Ross et al., 2010). Threshold concepts are described as conceptual gateways that, once passed, open up a new way of thinking and are distinguished from “key” or “core” concepts as they are more than mere building blocks toward understanding within a discipline (Meyer & Land, 2003, 2006). Thus, Tibell and Harms (2017) concluded that complete understanding of evolutionary theory might require the understanding of underlying threshold concepts such as randomness, probability, spatial scale, and temporal scale.

Research reveals that students particularly struggle with the importance and nature of randomness (Garvin-Doxas & Klymkowsky, 2008; Robson & Burns, 2011). The term is often used in everyday language to explain that a phenomenon is without order, predictability or pattern, while scientists use the term randomness to suggest unpredictability without referring to purposelessness (Wagner, 2012; Buiatti & Longo, 2013; Mead & Scott, 2010). Moreover, the notion of randomness in evolution is rather specific by speaking about events (e.g., mutations or genetic drift) that are independent of an organisms’ need or the directionality provided by the process of natural selection (Mead & Scott, 2010; Heams, 2014). Thus, mutations are called random because it cannot be predicted precisely where and when a mutation will appear, and mutations are not directed to an organisms’ adaptation (Heams, 2014). In contrast, natural selection itself can be described as a probabilistic process, if the process of selection is defined as individuals’ probabilities to survive and reproduce in a specific environment depending on their specific traits (Tibell & Harms, 2017). Therefore, a clear understanding of randomness and probability is essential for understanding evolution.

5.2.2 Computer Simulations and Instructional Support

Computer simulations can be effective tools to handle such intangible nature of scientific concepts (Ainsworth & VanLabeke, 2004; Plass et al., 2012). They also allow students to visualize processes occurring at spatial scales and temporal scales that are difficult or impossible to observe directly (Rutten, van Joolingen, & van der Veen, 2012). Moreover, simulations have several advantages over reading textbooks or attending lectures, including opportunities for students to explore theoretical situations, interact with a simplified version of the focal process(es), and/or change time-scales of events (van Berkum & de Jong, 1991).

However, research on simulation-based learning has revealed that learners may encounter difficulties during the learning process, for two contrasting reasons (de Jong & van Joolingen, 1998). One is that simulations can involve complex learning environments, which may overwhelm the learner due to the high amount of information that is conveyed and must be processed (Wouters & van Oostendorp, 2013). In stark contrast, minimizing guidance (and thus reducing the amount of information) may reduce the effectiveness of simulation based-learning (Rutten et al., 2012). Therefore, instructional support may be needed to provide suitable learning environments and overcome students' learning difficulties (Kombartzky, Ploetzner, Schlag, & Metz, 2010; Urhahne & Harms, 2006).

Several kinds of support may be provided in different phases of the learning process in efforts to enhance simulation-based learning (Zhang, Chen, Sun, & Reid, 2004). Interpretative support, given before the interaction, can provide scaffolding for learners to activate prior knowledge, and generate appropriate hypotheses. One way to provide effective interpretative support is to offer accessible domain-specific background information (Reid, Zhang, & Chen, 2003). Another, shown to have positive effects on learning outcomes, is to provide worked examples (Spanjers, Wouters, van Gog, & van Merriënboer, 2011; Yaman, Nerdel, & Bayrhuber, 2008).

Experimental support is provided during an interaction and can scaffold learners' process of scientific inquiry during simulation-based learning by helping them to design verifiable experiments, predict and observe the outcomes, and draw appropriate conclusions. Effective experimental support for knowledge acquisition may include gradual, cumulative introductions to handle a simulation and/or requests for learners to predict and describe the outcome (Urhahne & Harms, 2006; Wang, Wu, & Hsu, 2017).

Reflective support is provided after an interaction and may foster learners' integration of their discoveries. Such support scaffolds the integration of new information arising from discoveries after learners' interaction with a simulation. It involves promoting reflective processes, which may be done through a reflective assignment tool or opportunities to discuss the results (Eckhardt, Urhahne, Conrad, & Harms, 2013; Zhang et al., 2004).

5.2.3 Simulations to Support Students' Understanding of Evolution and Threshold Concepts

Evolutionary and threshold concepts might be tangible through appropriate simulations. Even though, biology is one of the major subjects with research on dynamic visualizations such as simulations, the topic of evolution is seldom mentioned (Lee & Tsai, 2013; Rutten et al., 2012). Moreover, although the number of available online educational videos increases; they often lack explanations regarding underlying threshold concepts, and if mentioned, they are communicated orally only (Bohlin, Göransson, Höst,

& Tibell, 2017). Simulations that explicitly visualize threshold concept such as randomness, and build connections to the probabilistic processes of natural selection may help students to overcome their problems with these concepts.

In fact, there are a few computer simulations available for free and used in evolution education such as Evolve (Price & Vaughn, 2010), Avida-ED (Pennock, 2007, 2018), or Evolution Readiness Activities (Concord Consortium, 2018). The conducted research studies indicated positive learning gains after using these simulations (Horwitz, McIntyre, Lord, O'Dwyer, & Staudt, 2013; Speth, Long, Pennock, & Ebert-May, 2009; Soderberg & Price, 2003). Nevertheless, they were designed to focus on evolutionary concepts without focusing on particular underlying threshold concepts such as randomness and probability. For instance, the activities of Evolution Readiness focus on the process of (natural) selection, variation within species (without referring to the origins of variation), and inheritance of various traits (Horwitz et al., 2013). This also counts for Evolve, which is designed to focus on the effects of selection, genetic drift, and migration of a population over time without modeling mutations or their random nature (Soderberg & Price, 2003). In contrast, Avida-ED includes random mutations occurring in the organisms' genome, while students can also observe evolution in action (Speth et al., 2009). Still, the above mentioned simulations do not imply the underlying threshold concepts such as randomness or probability.

5.2.4 EvoSketch Software

EvoSketch is project-developed interactive, web-based simulation software (free of charge and available online as English version at http://learninglabs.se/evolution/randomlineEN/ or German version at http://learninglabs.se/evolution/evosketchde/) that allows learners to explore random and probabilistic phenomena associated with the process of natural selection. The software (which can be used on various electronic devices, such as smartphones, tablets, laptops and desktop computers) generates a line (representing a reproducing organism) that is replicated by the user for 20 generations.

Every generation consists of four replications of a parent line (representing a reproducing organism), drawn with a mouse or finger, resulting in four offspring lines (Figure 5.1). Since copying errors inevitably occur while drawing, each replication varies and drifts slightly to the right or left of the parent line. These shifts in offspring lines represent the concept of the origin of variation, and hence random processes in evolution. After each generation has been completed by drawing four replications, one of the four offspring lines is selected (by the software) to continue the parent line, and thus represents the next reproducing "organism" in the simulation. The selected line is closest to a point (indicated by the red dot in Figure 5.1) indicating optimal fitness for the offspring in the surrounding environment. Thus, the organism represented by the selected line has the highest probability to survive and reproduce, and there is selective pressure on the line to move towards the point (probabilistic processes). After 20 generations the line will normally be shifted either to the right or the left due to the combination of copying error and selection. Users are guided through an EvoSketch exercise by an accompanying worksheet (EvoSketch Worksheet). This worksheet consists of two introductory texts explaining random processes (specifically, mutations) in evolution (128 words), and the probabilistic process of natural selection (98 words). Both texts are directly followed by a task asking learners to make predictions, run the simulation and/or observe the outcome, and explain the outcome (predict-observe-explain strategy; White & Gunstone, 1992). An example of such a task is available as supplementary material accompanying the online article (see Supplemental Material 5.8.1).

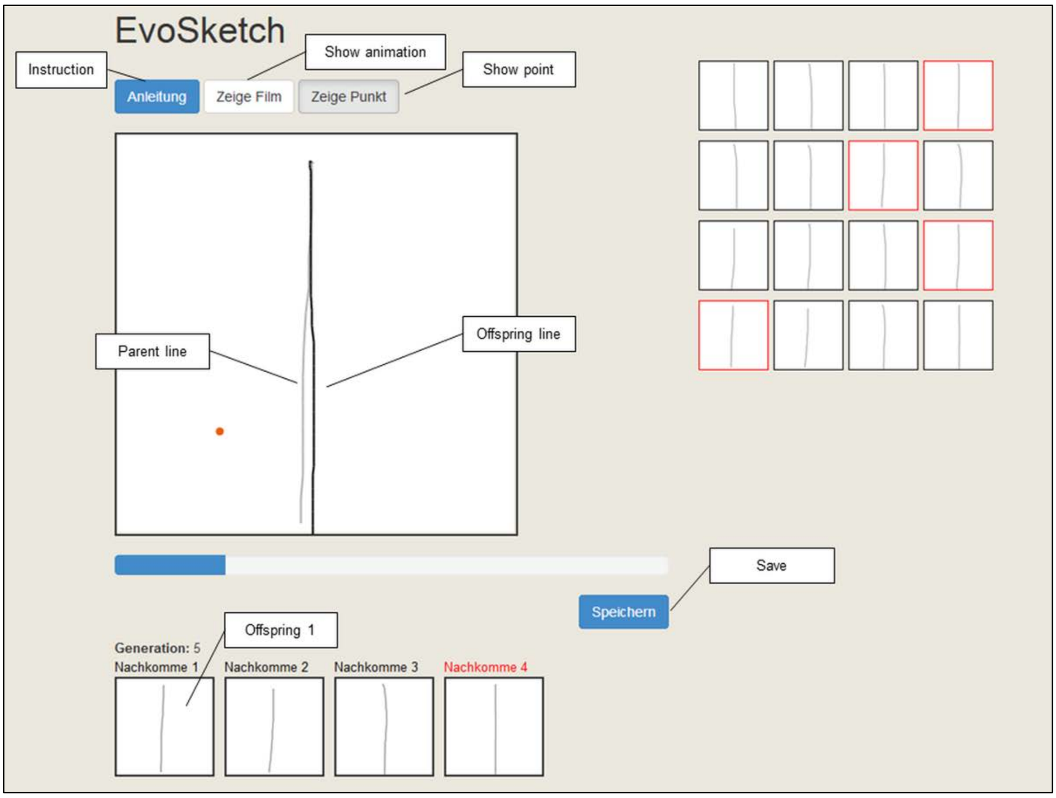


Figure 5.1 Screen display of an EvoSketch simulation (German version). The main box shows the offspring line (black) drawn, with a mouse or finger, from the parent line (grey). After saving the line it is visualized in one of the four offspring boxes (offspring 1 to 4). To the right are displayed all offspring lines that have been generated and saved so far. The red framed boxes show the offspring selected as parent lines for successive generations.

5.2.5 Research Aim

Since EvoSketch software provides integrated experimental support (EvoSketch Worksheet tasks), we addressed the potential utility of additional interpretative and reflective instructional support in this study. As described in the following sections, we used time-on-task, perceived cognitive load (PCL) of the learners, and knowledge test performance on three occasions to evaluate the effectiveness of EvoSketch for teaching and learning the roles and importance of randomness and probability in evolutionary contexts. We then compared pre-, post- and follow-up performance scores of students who learned with EvoSketch, with and without additional instructional support, to those of students who used text-based learning of the same topics.

5.3 Methods

5.3.1 Interventions

As already mentioned, the main aim of this study was to assess the effectiveness of EvoSketch for fostering students' conceptual knowledge of randomness, probability, and evolution. An additional aim was to identify which type of instructional support (if any) most effectively promotes learning with EvoSketch. For these purposes, we assigned sets of students to the following four kinds of intervention: text-based, simulation-based, simulation-based with interpretative support, and simulation-based with reflective support. Students assigned to all four groups received an overview of the topic of evolution by means of a short, standardized introductory text (cf. Neubrand, Borzikowsky, & Harms, 2016) to reactivate prior knowledge. The students of each group subsequently individually addressed the following worksheets and tasks:

Text-based intervention (hereafter, text). Learners of this intervention group worked with a worksheet (the two introductory EvoSketch Worksheet texts mentioned above) and a Powerpoint presentation on the roles of randomness (specifically, mutation) and probability (specifically, selection) in evolution. The presentation explains the evolution of wingless flies from normal flies on the Kerguelen Islands through random and probabilistic processes. Afterwards, learners were asked to answer three questions regarding information given in the presentation, and two questions regarding evolution (see Supplemental Material 5.8.2).

Simulation-based intervention (hereafter, simulation). Learners in this group were asked to follow the instructions of the EvoSketch Worksheet (mentioned above in section EvoSketch Software). They started by reading the introductory text on the topic of randomness in evolution and worked through the first task. During this task, they also progressed through the EvoSketch simulations. They then read the second text on the role of probability in evolution and addressed the second task, regarding selective pressure (indicated by distance from the red point in their simulation). Learners in this group did not receive additional instructional support.

Simulation-based intervention with interpretative support (hereafter, sim-interpret). This was identical to the simulation intervention, except that learners were provided interpretative support in the form of a worked example on the roles of randomness and probability in evolution before starting to work with the simulation. Worked examples consist of a problem followed by a worked-out solution, normally presented in a step-by-step format to the learner (Renkl, 2005). We used a worked example created by Neubrand et al., 2016, with revised and supplementary sections added in efforts to increase the focus on randomness and probability aspects, and thus establish helpful connections to EvoSketch simulations. A comparison of the original and revised worked example is available as supplementary material accompanying the online article (see Supplemental Material 5.8.3).

Simulation-based intervention with reflective support (hereafter, sim-reflect). The last group of learners, the sim-reflect intervention group, also worked through the mentioned EvoSketch Worksheet. However, in contrast to the simulation and sim-interpret groups, learners received reflective support in the form of reflective questions after each task while working with the simulation (e.g., "Describe briefly why the line has shifted this way").

5.3.2 Participants

The sample consisted of 14 classes from nine comprehensive schools (German *Gemeinschaftsschulen*) in northern Germany. In total, 267 tenth grade students aged between 14 and 18 years (M = 15.6 years, SD = 0.6 years; 47.19% female) participated in the study. Students of each class were randomly assigned to one of the four intervention groups: text (n = 43), simulation (n = 70), sim-interpret (n = 79), and sim-reflect (n = 77). The study was conducted during regular science lessons between November 2016 and March 2017. All students were informed that participation was voluntary and that their results would not affect their final grades. Students had received no formal instruction on evolutionary theory before. Nevertheless, we assume that they had some fragmentary knowledge on topics related to aspects of evolutionary theory (e.g., genetics), although evolutionary theory is not specifically included in the German curriculum before the tenth grade (Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany, 2005).

5.3.3 Instruments

The following instruments were used to acquire data concerning the effectiveness of the interventions and potentially influential variables.

Randomness and Probability Test in the context of Evolution (RaProEvo). RaProEvo is a test instrument designed to measure students' conceptual knowledge of randomness and probability in evolutionary contexts (Fiedler, Tröbst, & Harms, 2017). It comprises 21 items (16 multiple-choice, three open response and two matching items) that focus on five aspects in which randomness and probability play important roles: the origin of variation, accidental death, random phenomena, the process of natural selection, and the probability of events. The items are scored dichotomously, and we used a reduced set of 19 items (excluded the two matching items) with internal consistency (Cronbach's Alpha) ranging from .44 and .63.

Conceptual Inventory of Natural Selection (CINS). CINS is a diagnostic test designed to assess students' understanding of evolution through natural selection (Anderson, Fisher, & Norman, 2002). It consists of 20 multiple-choice questions that focus on common misconceptions pertaining to 10 key conceptual aspects of natural selection, variation, and speciation. The inventory is structured so that each of the 10 concepts is assessed once in items 1 to 10 (CINS-A) and once again in items 11 to 20 (CINS-B). We used the CINS-A and CINS-B sets of items in the pretests and posttests, respectively, to minimize the influence of pretest on posttest scores and students' fatigue by reading the same items. All items are dichotomously scored, and Cronbach's Alpha ranged from .23 and .40, suggesting that effects in this study may be somewhat underestimated due to lower-than-desired reliability.

General Biological Content Knowledge Test (GBCK). The GBCK test was used to control for differences in students' existing prior knowledge of the subject (Neubrand et al., 2016). It is designed to measure tenth-grade students' existing content knowledge of biological topics included in up to tenth grade curricula. It consists of 19 dichotomously scored items (16 multiple-choice items, two matching items, and one open response item). The results obtained with our students indicate that the test has an internal consistency (Cronbach's Alpha) of .36, lower than the level (.51) reported by Neubrand et al. (2016) in applications with other samples of tenth-grade German students.

Students' General Language Proficiency (C-Test). C-Tests are designed to measure students' general language proficiency (Eckes & Grotjahn, 2006), which may affect their performance in other diagnostic test instruments (Härtig, Heitmann, & Retelsdorf, 2015). Therefore, we assessed our students' general language proficiency using C-tests based on two texts, each including 20 words with missing letters (Wockenfuß & Raatz, 2006). Since learners' ability to read items or texts and produce answers is highly relevant in a study such as this, the responses were screened for both orthographical and grammatical errors. The students' answers were dichotomously scored, and Cronbach's Alpha of the test was found to be .78.

Perceived Cognitive Load (PCL). Cognitive load can affect learning (Sweller, 1994), but it can be reduced by providing instructional support for learning with simulations (Leutner, 1993). Therefore, students' PCL during the intervention was assessed using a 5-point rating scale instrument (Urhahne, 2002), consisting of eight items that allow differentiation of participants' PCL with a Cronbach's Alpha of .87.

Self-Reported Effort. Scores obtained by takers of any test are likely to depend on the effort they expend while taking it (Wise & Kong, 2005). Thus, students' self-reported test-taking effort was appraised on one 10-point scale item (OECD, 2010), after they completed both the pre- and post-tests.

5.3.4 Procedure

Prior to the intervention (day 1), every student took pretests consisting of the targeted randomness, probability and evolutionary knowledge tests (RaProEvo and CINS-A), and instruments designed to capture information on the control variables: general biological knowledge (GBCK), language proficiency (C-Test), self-reported test-taking effort, and demographic data (i.e., age, sex, and biology grade). Roughly two weeks later (day 2: intervention day) every student of each intervention group worked alone through their own EvoSketch Worksheet on a single laptop. Laptops were all of the same type and provided by the Leibniz Institute of Science and Mathematics Education (IPN). Students of all intervention groups had 45 minutes to complete their worksheet tasks (intervention). On average, learners spent 30 minutes (*SD* = 7 minutes; range 13 to 52 minutes) completing their tasks. Immediately after completing their worksheet, students took posttests, consisting of the knowledge tests (RaProEvo and CINS-B) and items asking about their self-reported test-taking effort and PCL during the learning process. Roughly eight school weeks later (day 3), students took follow-up tests consisting of the targeted knowledge tests (RaProEvo and CINS-B). The study was conducted by the first author, with support from a university student who set-up and removed the laptops on the second day.

5.3.5 Statistical Analysis

We analyzed the CINS-B and RaProEvo responses with generalized linear mixed models featuring a logistic link function, crossed random effects for participants and items, and an additional random effect for class (Baayen, Davidson, & Bates, 2008). The random effects for participants and items were included to account for differences in participants' general ability and items' general difficulty, respectively. The random effect for class also controlled for possible discrepancies in average ability between classes. To uncover systematic effects of the experimental conditions on the development of students' knowledge, dummy-coded variables for intervention, assessment and their interaction were incorporated as fixed effects in the models. Text served as the reference category for intervention, while posttest and pretest, respectively, served as the reference category for the CINS-B and RaProEvo assessments. This approach ensured simultaneous generalization of significant effects to new samples of both participants and items (Raaijmakers, Schrijnemakers, & Gremmen, 1999). As measures of effect size for fixed effects, we computed odds ratios. We used the lme4-package (Bates, Maechler, & Bolker, 2011) for the statistical computing environment R 3.0.0 (R Core Team, 2013) for all these statistical analyses.

5.4 Results

5.4.1 Baseline Equivalence

One-way analyses of variance (ANOVAs) were conducted to detect possible significant differences between intervention groups in pretest performance (i.e., CINS-A and RaProEvo scores) or the control variables (demographic variables and either C-Test or GBCK scores). Values of these variables for each of the groups are listed in Table 1. The ANOVA results indicated that the groups only significantly differed in C-Test performance ($F(3, 243) = 3.14$, $p = .026$, partial $\eta^2 = .04$), and post-hoc Bonferroni tests indicated there were no significant between-group differences (*p*s > .060). Thus, the random assignment of learners to the four intervention groups caused no apparent bias in terms of any of these variables.

Table 5.1 Age (years), Biology Grade, Control Variable Scores (GBCK C-Test), and Pretest Performance Scores (RaProEvo, CINS-A) of the Four Intervention Groups (Means, with Standard Deviations in Parentheses).

Intervention	*n*	Age	Biology Grade	GBCK	C-Test	RaProEvo	CINS-A
Text	39	15.39	2.51	6.41	13.05	10.08	3.51
		(0.55)	(0.73)	(2.61)	(2.88)	(2.51)	(1.68)
Simulation	66	15.62	2.61	5.86	11.58	9.29	3.45
		(0.74)	(0.91)	(2.31)	(3.34)	(2.92)	(1.84)
Sim-interpret	70	15.57	2.77	6.23	12.90	9.93	3.90
		(0.61)	(0.75)	(2.31)	(2.78)	(2.57)	(1.87)
Sim-reflect	69	15.65	2.77	5.90	12.74	9.57	3.84
		(0.66)	(0.75)	(2.10)	(2.77)	(2.78)	(1.78)
Total	244	15.58	2.65	6.07	12.52	9.68	3.70
		(0.65)	(0.85)	(2.30)	(2.99)	(2.72)	(1.81)

Note. The total sample is smaller than $N = 267$ because 23 participants were absent during the pretests. *GBCK* = general biological content knowledge; *C-Test* = general language proficiency; *RaProEvo* = conceptual knowledge of randomness and probability in evolutionary context; *CINS-A* = conceptual inventory of natural selection.

5.4.2 Time-on-Task

Differences in knowledge acquisition could be due to differences in the time learners spend on tasks in their interventions, and one-way ANOVA indicated a significant effect of intervention on time-on-task: $F(3, 216) = 16.63$, $p < .001$, partial $\eta^2 = .19$. Post-hoc Bonferroni tests revealed that learners in the text group worked significantly longer than learners in the simulation intervention group (adjusted $M = 1822.70$ seconds, $SD = 388.51$ seconds, $n = 37$; and adjusted $M = 1587.46$ seconds, $SD = 388.51$ seconds, $n = 59$, $p = .026$, $d = 0.61$, respectively). Furthermore, simulation intervention learners spent significantly less time with the material than the sim-interpret (adjusted $M = 2033.33$ seconds, $SD = 388.53$ seconds, $n = 63$, $p < .001$, $d = 1.15$) and sim-reflect intervention (adjusted $M = 2001.64$ seconds, $SD = 388.48$ seconds, $n = 61$, $p < .001$, $d = 1.07$) learners.

5.4.3 Perceived Cognitive Load (PCL)

Since cognitive load can influence learners' knowledge acquisition (Sweller, 1994), the students' PCL was measured directly after each intervention, and one-way ANOVA indicated that there were significant differences between intervention groups: $F(3, 215) = 3.68$, $p = .013$, partial $\eta^2 = .04$. Post-hoc Bonferroni tests showed that average PCL was higher in the text intervention group (adjusted $M = 1.35$, $SD = 0.61$, $n = 37$) than in the simulation (adjusted *M* = 0.97, *SD* = 0.61, *n* = 59, *p* = .021, *d* = 0.62) and sim-reflect intervention (adjusted M = 0.96, SD = 0.62, n = 61, p = .017, d = 0.62) groups. However, no significant differences in this respect between the other pairs of interventions were detected (in all remaining cases, p > .05).

5.4.4 Self-Reported Effort

Test performance may depend on the test-taking effort, as low effort is likely to result in test scores underrepresenting learners' true level of knowledge (Wise & Kong, 2005). Thus, we applied repeated-measures ANOVA to investigate differences between intervention groups in self-reported test-taking effort. The results showed a significant main effect of self-reported test-taking effort: $F(1,229) = 30.86$, $p < .001$, partial $\eta^2 = .12$. Repeated contrasts also revealed that learners self-reportedly spent significantly more effort in the pretests (adjusted *M* = 7.04, *SD* = 1.83) than in the posttests (adjusted *M* = 6.25, *SD* = 2.14, *n* = 233; $p < .001$, *d* = 0.40). Nevertheless, no significant main effect of group or interaction effect between group and effort was detected: $F(3,229) = 1.46$, $p = .227$, partial $\eta^2 = .02$ and $F(3,229) = 0.47$, $p = .701$, partial $\eta^2 = .01$, respectively.

In addition, Pearson's correlation coefficients were calculated to assess relationships between effort (pre- and post-test) and test performance (RaProEvo and CINS-A/B scores). A significant positive association was detected between pretest effort and RaProEvo pretest performance ($r_s = .14$, $p = .031$, $n = 236$). Significant positive relationships were also found between posttest effort and both RaProEvo and CINS-B posttest scores ($r_s = .18$, $p = .007$, $n = 240$, and $r_s = .14$, $p = .028$, $n = 240$, respectively).

5.4.5 Intervention Effects on CINS-B and RaProEvo Scores

CINS-B. As already stated, students' responses to the CINS-B items were analyzed with a generalized linear mixed model featuring a logistic link function, crossed random effects for participants and items, and a random effect for class. Dummy-coded variables for intervention (with text as the reference category) and assessment (with posttest as the reference category) were included as fixed effects. Moreover, a fixed effect for students' CINS-A pretest performance was included as a covariate. A main effect for CINS-A (*b* = 0.12, *SE* = 0.02, $p < .001$, *OR* = 1.12) was detected, but no other significant fixed effects (Table 5.2). Thus, no significant differences were detected, at posttest or follow-up, between intervention groups in understanding of evolution through natural selection. Inclusion of the GBCK results, C-Test scores, and time-on-task data as further covariates did not alter this pattern of results.

RaProEvo. Students' RaProEvo performance was explored with a similar model, but with pretest as the reference category for assessment. No significant main effects of intervention were detected, indicating that there were no substantial differences between intervention groups in conceptual knowledge of randomness and probability in evolutionary context at the outset of the study (Table 5.2). Similarly, there was no significant general improvement in students' performance across assessments. However, a significant interaction revealed that students in the simulation group outperformed students in the text group at follow-up, b = 0.34, SE = 0.15, p = .02, OR = 1.40. Incorporation of the GBCK results, C-Test scores, and time-on-task data as covariates did not change this pattern of results.

Table 5.2 Intervention Effects on CINS-B and RaProEvo Scores

Effects	**CINS-B**			**RaProEvo**		
	b	***SE***	***OR***	***b***	***SE***	***OR***
Fixed						
Intercept	-1.16***	0.22	0.31	0.10	0.26	1.10
Simulation	0.14	0.15	1.15	-0.21	0.15	0.80
Sim-interpret	-0.01	0.15	1.00	-0.01	0.15	0.99
Sim-reflect	-0.02	0.15	0.98	-0.14	0.15	0.87
Post	-	-	-	-0.09	0.11	0.92
Follow-up	0.25	0.16	1.28	-0.04	0.12	0.96
CINS-A	0.12***	0.02	1.12	-	-	-
Simulation*Post	-	-	-	0.04	0.15	1.04
Sim-interpret*Post	-	-	-	-0.04	0.14	0.96
Sim-reflect*Post	-	-	-	0.09	0.14	1.09
Simulation*Follow-up	-0.25	0.20	0.78	0.34*	0.15	1.10
Sim-interpret*Follow-up	-0.09	0.19	0.91	0.09	0.15	1.40
Sim-reflect*Follow-up	-0.01	0.20	0.99	0.16	0.15	1.17
	Var	*SD*		*Var*	*SD*	
Random						
$\text{Participants}_{\text{(intercept)}}$	0.08	0.28		0.30	0.55	
$\text{Items}_{\text{(intercept)}}$	0.28	0.52		0.94	0.97	
$\text{Class}_{\text{(intercept)}}$	0.02	0.13		0.05	0.23	

* $p < .05$, *** $p < .001$

Note. CINS-A/B = pretest/posttest of the conceptual inventory of natural selection; *RaProEvo* = pretest of conceptual knowledge of randomness and probability in evolutionary context;

5.5 Discussion

The main aim of this study was to assess the effectiveness of EvoSketch simulations for improving students' knowledge of the importance and roles of randomness and probability in evolutionary contexts, and their evolutionary knowledge. Since instructional support may reportedly improve the effectiveness of simulations, and EvoSketch Worksheets provide experimental support, we also examined and compared effects of interpretative and reflective support (a worked example and reflective questions, respectively) on learning with EvoSketch.

We found significant differences between intervention groups in both PCL and time spent on the material. Learners in the simulation with additional support (sim-interpret and sim-reflect) groups worked significantly longer on their tasks than learners in the simulation (without additional support) group. However, these groups did not differ in PCL. Students in the text group spent an intermediate amount of time on their worksheets, but reported a significantly higher PCL than students of the simulation and sim-reflect groups. The high PCL of the text group could have resulted from aspects of the intervention. These students received a text in a Powerpoint presentation, and had to answer three questions regarding topics covered in this text and two regarding evolution in a broader sense. This may have caused them to perceive a higher cognitive load than students of the other intervention groups, who were not asked these questions.

Further, students of all groups self-reportedly spent more effort in test-taking on the pretest than on the posttest (directly after the intervention), but no significant differences in these respects were detected among the groups. This may explain why mean posttest scores were lower, but mean follow-up test scores were higher, than pretest scores. Moreover, RaProEvo test scores indicated that learners in the simulation intervention group (but not those in the simulation with additional support groups) acquired significantly more knowledge between the pre- and follow-up tests than text learners. Thus, EvoSketch seems to be an effective tool for fostering students' conceptual knowledge of randomness and probability in evolutionary contexts.

The secondary school students who participated in the study had little prior knowledge of the focal topics. This is potentially problematic as learners may be overwhelmed by high amounts of information conveyed in simulations (Rutten et al., 2012; Wouters & van Oostendorp, 2013). The improvements in delayed knowledge acquisition of the simulation (without additional support) group, relative to the text-based learners, indicates that EvoSketch is not too abstract for fostering learners' knowledge about randomness and probability. However, it may not foster broad evolutionary knowledge.

Moreover, additional instructional support in either the interpretative or reflective forms did not lead to improvement in the performance of simulation-based learners relative to text group learners. This may have been because the students were overwhelmed by the high amount of additional information provided in these interventions. Moreover, the high amount of new information could have deterred learners with low interest, and reduced motivation (Amabile, Hill, Hennessey, & Tighe, 1994; Pintrich & Schrauben, 1992). Participants did not receive any credit for their test performance, and their results did not influence their final grade. Thus, their inherent learning motivation was likely correlated with motivation to address the large amount of material (e.g., worked example and large numbers of test items), thereby introducing a substantial random behavioral response factor in the posttest results (e.g., Meijer, 2003). Accordingly, results of correlation analyses indicated a positive correlation between self-reported effort and posttest scores. Moreover, students reported lower test-taking effort in the posttests than in the pretests.

Another factor that may have affected the results, particularly CINS scores, may have been the limited duration of the learning session (the intervention time was roughly 45 minutes). Decades of evolution education research have shown that the theory of evolution presents severe problems to learners, which have not been effectively solved by teaching strategies applied to date (e.g., Kampourakis & Zogza, 2008; Rosengren et al., 2012). Introducing abstract, counter-intuitive concepts in addition to these problems (particularly in a brief intervention) may partly explain the generally weak between-intervention differences in students' learning.

It should also be noted that the results are limited by the low reliability of the knowledge test instruments (RaProEvo, CINS, and GBCK). The GBCK test's reliability was not expected to be high because it covers a large range of biological topics. Nevertheless, its internal consistency was unsatisfactorily low. The internal consistency of the CINS instrument was similarly low, possibly because the tenth grade students had not received formal instruction on evolutionary theory before the intervention, and may have been overstrained by the complexity of the presented items. The internal consistency of the RaProEvo instrument was higher, but still not satisfactory.

Nevertheless, our findings may be useful for refining EvoSketch simulations and implementing EvoSketch as a tool to aid learning evolution in school sessions. Adequate knowledge of evolution, and particularly related abstract concepts such as randomness and probability, is essential for students to critically address numerous issues associated with their environment and everyday life. We recommend increasing the intervention timeframe to incorporate interventions on several days or weeks to increase students' understanding of randomness and probability as well as evolutionary knowledge. Further, deep learning is often more strongly supported by small-group learning than individual learning (Dori & Belcher, 2005; Springer, Stanne, & Donovan, 1999). Thus, EvoSketch could be used in group settings with each individual working initially on their own through simulations, and subsequently discussing observations in small groups. Such discussion could also be extended to class discussions with the teacher. Learners with little prior knowledge could also receive additional instructional support through worked examples or reflective prompts.

In conclusion, EvoSketch seems to be an effective tool for learning and teaching concepts of randomness and probability in evolutionary contexts, and seems to have positive long-term effects on understanding these concepts. In particular, simulation-based learning of randomness and probability appeared to promote delayed recall of knowledge more effectively than text-based learning.

5.6 Acknowledgments

The study reported here was part of the Swedish-German cooperation project *"EvoVis: Challenging Threshold Concepts in Life Science – enhancing understanding of evolution by visualization"*, funded by the Swedish Research Council. We thank the EvoVis team, particularly Andreas Göransson, for developing the simulation software. Special thanks are also due to the participating teachers for providing access to classes, the students who participated in the study, and the university students Denise Buchmann and Charlotte

Hölterhoff for their help with the laptops. We are also very grateful to Dr John Blackwell for language review.

5.7 References

Ainsworth, S., & VanLabeke, N. (2004). Multiple forms of dynamic representation. *Learning and Instruction, 14*, 241–255. doi:10.1016/j.learninstruc.2004.06.002

Amabile, T. M., Hill, K. G., Hennessey, B. A., & Tighe, E. M. (1994). The Work Preference Inventory: assessing intrinsic and extrinsic motivational orientations. *Journal of Personality and Social Psychology, 66,* 950–967. doi:10.1037/0022-3514.66.5.950

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

Baalmann, W., Frerichs, V., Weitzel, H., Gropengießer, H., & Kattmann, U. (2004). Schülervorstellungen zu Prozessen der Anpassung–Ergebnisse einer Interviewstudie im Rahmen der Didaktischen Rekonstruktion. [Students conceptions of adaptation processes - Results of an interview study within the framework of didactic reconstruction]. *Zeitschrift für Didaktik der Naturwissenschaften, 10*, 7–28.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. doi:10.1016/j.jml.2007.12.005

Bates, D. M., Maechler, M., & Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes.* R package version 0.999999-2.

Beggrow, E. P., & Nehm, R. H. (2012). Students' mental models of evolutionary causation: natural selection and genetic drift. *Evolution: Education and Outreach, 5*, 429–444. doi:10.1007/s12052-012-0432-z

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

Bohlin, G., Göransson, A., Höst, G. E., & Tibell, L. A. (2017). A conceptual characterization of online videos explaining natural selection. *Science & Education, 26*, 975–999. doi:10.1007/s11191-017-9938-7

Buiatti, M., & Longo, G. (2013). Randomness and multilevel interactions in biology. *Theory in Biosciences, 132*, 139–158. doi:10.1007/s12064-013-0179-2

Concord Consortium (2018). *Evolution readiness activities [Simulation software]*. Retreived from https://learn.concord.org/evolution-readiness (accessed 15 February 2018) de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*, 179–201. doi:10.3102/00346543068002179

Dori, Y. J., & Belcher, J. (2005). How does technology-enabled active learning affect undergraduate students' understanding of electromagnetism concepts? *The Journal of the Learning Sciences, 14*, 243–279. doi:10.1207/s15327809jls1402_3

Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing, 23*, 290–325. doi:10.1191/0265532206lt330oa

Eckhardt, M., Urhahne, D., Conrad, O., & Harms, U. (2013). How effective is instructional support for learning with computer simulations? *Instructional Science, 41*, 105–124. doi:10.1007/s11251-012-9220-y

Garvin-Doxas, K., & Klymkowsky, M. W. (2008). Understanding randomness and its impact on student learning: lessons learned from building the Biology Concept Inventory (BCI). *CBE-Life Sciences Education, 7*, 227–233. doi:10.1187/cbe.07-08-0063

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

Härtig, H., Heitmann, P., & Retelsdorf, J. (2015). Analyse der Aufgaben zur Evaluation der Bildungsstandards in Physik - Differenzierung von schriftsprachlichen Fähigkeiten und Fachlichkeit. [Analysis of tasks for the evaluation of the educational standards in physics - differentiation of literary skills and professionalism]. *Zeitschrift für Erziehungswissenschaft, 18,* 763–779. doi:10.1007/s11618-015-0646-2

Hartley, L. M., Momsen, J., Maskiewicz, A., & D'Avanzo, C. (2012). Energy and matter: differences in discourse in physical and biological sciences can be confusing for introductory biology students. *BioScience, 62*, 488–496. doi:10.1525/bio.2012.62.5.10

Heams, T. (2014). Randomness in biology. *Mathematical Structures in Computer Science, 24*, e240308. doi:10.1017/S096012951200076X

Herrmann-Abell, C. F., & DeBoer, G. E. (2018). Investigating a learning progression for energy ideas from upper elementary through high school. *Journal of Research in Science Teaching, 55*, 68-93. doi:10.1002/tea.21411

Horwitz, P., McIntyre, C. A., Lord, T. L., O'Dwyer, L. M., & Staudt, C. (2013). Teaching 'evolution readiness' to fourth graders. *Evolution: Education and Outreach, 6*, 1–21. doi:10.1186/1936-6434-6-21

Kampourakis, K., & Zogza, V. (2008). Students' intuitive explanations of the causes of homologies and adaptations. *Science & Education, 17*, 27–47. doi:10.1007/s11191-007-9075-9

Kombartzky, U., Ploetzner, R., Schlag, S., & Metz, B. (2010). Developing and evaluating a strategy for learning from animations. *Learning and Instruction, 20*, 424–433. doi:10.1016/j.learninstruc.2009.05.002

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago, IL: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lee, S. W.-Y., & Tsai, C.-C. (2013). Technology-supported learning in secondary and undergraduate biological education: observations from literature review. *Journal of Science Education and Technology, 22*, 226–233. doi:10.1007/s10956-012-9388-6

Leutner, D. (1993). Guided discovery learning with computer-based simulation games: Effects of adaptive and non-adaptive instructional support. *Learning and Instruction, 3*, 113–132. doi:10.1016/0959-4752(93)90011-N

Mead, L. S., & Scott, E. C. (2010). Problem concepts in evolution part II: Cause and chance. *Evolution: Education and Outreach, 3*, 261–264. doi:10.1007/s12052-010-0231-3

Meijer, R. R. (2003). Diagnosing item score patterns on a test using item response theory-based person-fit statistics. *Psychological Methods, 8*, 72–87. doi:10.1037/1082-989x.8.1.72

Meyer, J. H., & Land, R. (2003). Threshold concepts and troublesome knowledge: linkages to ways of thinking and practising within the disciplines. In C. Rust (Ed.), *Improving Student Learning: Theory and Practice Ten Years On* (pp. 412–424). Oxford, United Kingdom: Oxford Centre for Staff and Learning Development (OCSLD).

Meyer, J. H., & Land, R. (2006). Threshold concepts and troublesome knowledge: an introduction. In J. H. Meyer & R. Land (Eds.), *Overcoming barriers to student understanding: Threshold concepts and troublesome knowledge* (pp. 3–18). Abington, United Kingdom: Routledge.

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

Neubrand, C., Borzikowsky, C., & Harms, U. (2016). Adaptive prompts for learning evolution with worked examples - Highlighting the students between the "Novices" and the "Experts" in a classroom. *International Journal of Environmental and Science Education, 11*, 6774–6795.

OECD, Organisation for Economic Co-operation and Development (2010). *PISA computer-based assessment of student skills in science*. Paris, France: OECD.

Opfer, J. E., Nehm, R. H., & Ha, M. (2012). Cognitive foundations for science assessment design: Knowing what students know about evolution. *Journal of Research in Science Teaching, 49*, 744–777. doi:10.1002/tea.21028

Opitz, S. T., Neumann, K., Bernholt, S., & Harms, U. (2017). How do students understand energy in biology, chemistry, and physics? Development and validation of an assessment instrument. *Eurasia Journal of Mathematics Science and Technology Education, 13*, 3019–3042. doi:10.12973/eurasia.2017.00703a

Pennock, R. T. (2007). Learning evolution and the nature of science using evolutionary computing and artificial life. *McGill Journal of Education, 42,* 211–224.

Pennock, R. T. (2018). *Avida-ED [Educational application]*. Retreived from https://avida-ed.msu.edu/avida-ed-application/ (accessed 15 February 2018)

Pintrich, P. R., & Schrauben, B. (1992). Students' motivational beliefs and their cognitive engagement in classroom academic tasks. In D. H. Schunk & J. L. Meece (Eds.), *Student perceptions in the classroom* (pp. 149–183). Hillsdale, NJ: Erlbaum.

Plass, J. L., Milne, C., Homer, B. D., Schwartz, R. N., Hayward, E. O., Jordan, T., . . . Barrientos, J. (2012). Investigating the effectiveness of computer simulations for chemistry learning. *Journal of Research in Science Teaching, 49*, 394–419.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: *Toward a theory of conceptual change. Science Education, 66,* 211–227. doi:10.1002/sce.3730660207

Price, F., & Vaughn, V. G. (2010). *Evolve [Simulation software]*. Retreived from https://bioquest.org/BQLibrary/library_details.php?product_id=2367 (accessed 15 February 2018)

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "The language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41*, 416–426. doi:10.1006/jmla.1999.2650

R Core Team (2013). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rector, M., Nehm, R., & Pearl, D. (2013). Learning the Language of Evolution: Lexical Ambiguity and Word Meaning in Student Explanations. *Research in Science Education, 43*, 1107–1133. doi:10.1007/s11165-012-9296-z

Reid, D. J., Zhang, J., & Chen, Q. (2003). Supporting scientific discovery learning in a simulation environment. *Journal of Computer Assisted Learning, 19*, 9–20. doi:10.1046/j.0266-4909.2003.00002.x

Renkl, A. (2005). The worked-out examples principle in multimedia learning. In Mayer, R.E. (Ed.), *The Cambridge Handbook of Multimedia Learning.* Cambridge: Cambridge University Press.

Robson, R. L., & Burns, S. (2011). Gain in student understanding of the role of random variation in evolution following teaching intervention based on Luria-Delbruck experiment. *Journal of Microbiology & Biology Education: JMBE, 12*, 3–7. doi:10.1128/jmbe.v12i1.272

Rosengren, K. S., Brem, S. K., Evans, E. M., & Sinatra, G. M. (2012). *Evolution challenges: Integrating research and practice in teaching and learning about evolution*. New York, NY: Oxford University Press.

Ross, P. M., Taylor, C. E., Hughes, C., Whitaker, N., Lutze-Mann, L., Kofod, M., & Tzioumis, V. (2010). Threshold concepts in learning biology and evolution. *Biology International, 47*, 47–52.

Rutten, N., van Joolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education, 58*, 136–153. doi:10.1016/j.compedu.2011.07.017

Schmiemann, P., Nehm, R. H., & Tornabene, R. E. (2017). Assessment of Genetics Understanding. *Science & Education, 26*, 1161–1191. doi:10.1007/s11191-017-9925-z

Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss (Jahrgangsstufe 10) (Beschluss der Kultusministerkonferenz vom 16.12.2004) [National educational standards for the subject of biology concerning the 10th grade]*. Munich, Germany: Luchterhand.

Shtulman, A. (2006). Qualitative differences between naïve and scientific theories of evolution. *Cognitive Psychology, 52*, 170–194. doi:10.1016/j.cogpsych.2005.10.001

Sinatra, G. M., Brem, S. K., & Evans, E. M. (2008). Changing minds? Implications of conceptual change for teaching and learning about biological evolution. *Evolution: Education and Outreach, 1*, 189–195. doi:10.1007/s12052-008-0037-8

Soderberg, P., & Price, F. (2003). An examination of problem-based teaching and learning in population genetics and evolution using EVOLVE, a computer simulation. International *Journal of Science Education, 25*, 35–55. doi:10.1080/09500690110095285

Spanjers, I. A., Wouters, P., van Gog, T., & van Merriënboer, J. J. (2011). An expertise reversal effect of segmentation in learning from animated worked-out examples. *Computers in Human Behavior, 27,* 46–52. doi:10.1016/j.chb.2010.05.011

Speth, E. B., Long, T. M., Pennock, R. T., & Ebert-May, D. (2009). Using Avida-ED for teaching and learning about evolution in undergraduate introductory biology courses. *Evolution: Education and Outreach, 2*, 415–428. doi:10.1007/s12052-009-0154-z

Spindler, L., & Doherty, J. (2009). Assessment of the teaching of evolution by natural selection through a hands-on simulation. *Teaching Issues and Experiments in Ecology, 6,* 1–20.

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research, 69*, 21–51. doi:10.3102/00346543069001021

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4,* 295–312. doi:10.1016/0959-4752(94)90003-5

Tibell, L. A., & Harms, U. (2017). Biological principles and threshold concepts for understanding natural selection. *Science & Education, 26*, 953–973. doi:10.1007/s11191-017-9935-x

To, C., Tenenbaum, H. R., & Hogh, H. (2017). Secondary school students' reasoning about evolution. *Journal of Research in Science Teaching, 54*, 247–273. doi:10.1002/tea.21347

Tsui, C. Y., & Treagust, D. F. (2004). Conceptual change in learning genetics: An ontological perspective. *Research in Science & Technological Education, 22*, 185–202. doi:10.1080/0263514042000290895

Tsui, C. Y., & Treagust, D. F. (2010). Evaluating secondary students' scientific reasoning in genetics using a two-tier diagnostic instrument. *International Journal of Science Education, 32*, 1073–1098. doi:10.1080/09500690902951429

Urhahne, D. (2002). *Motivation und Verstehen [Motivation and understanding]*. Münster, Germany: Waxmann Verlag.

Urhahne, D., & Harms, U. (2006). Instruktionale Unterstützung beim Lernen mit Computersimulationen. [Instructional support for learning with computer simulations]. *Unterrichtswissenschaft, 34,* 358–377.

van Berkum, J. J., & de Jong, T. (1991). Instructional environments for simulations. *Education and Computing, 6,* 305–358. doi:10.1016/0167-9287(91)80006-J

Venville, G., Gribble, S. J., & Donovan, J. (2005). An exploration of young children's understandings of genetics concepts from ontological and epistemological perspectives. *Science Education, 89,* 614–633. doi:10.1002/sce.20061

Wagner, A. (2012). The role of randomness in Darwinian evolution. *Philosophy of Science, 79*, 95–119. doi:10.1086/663239

Wang, J.-Y., Wu, H.-K., & Hsu, Y.-S. (2017). Using mobile applications for learning: Effects of simulation design, visual-motor integration, and spatial ability on high school students' conceptual understanding. *Computers in Human Behavior, 66*, 103–113. doi:10.1016/j.chb.2016.09.032

Wernecke, U., Schwanewedel, J., & Harms, U. (2018). Metaphors describing energy transfer through ecosystems: Helpful or misleading? *Science Education, 102*, 178–194. doi:10.1002/sce.21316

White, R. T., & Gunstone, R. F. (1992). *Probing understanding.* London, United Kingdom: Routledge.

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*, 163–183. doi:10.1207/s15324818ame1802_2

Wockenfuß, V., & Raatz, U. (2006). Über den Zusammenhang zwischen Testleistung und Klassenstufe bei muttersprachlichen C-Tests [About the relationship between test performance and grade level in native language C-tests]. *Der C-Test: Theorie, Empirie, Anwendungen/The C-Test: Theory, Empirical Research, Applications*, 211–242.

Wouters, P., & van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education, 60*, 412–425. doi:10.1016/j.compedu.2012.07.018

Yaman, M., Nerdel, C., & Bayrhuber, H. (2008). The effects of instructional support and learner interests when learning using computer simulations. *Computers & Education, 51*, 1784–1794. doi:10.1016/j.compedu.2008.05.009

Yates, T. B., & Marek, E. A. (2015). A study identifying biological evolution-related misconceptions held by prebiology high school students. *Creative Education, 6*, 811–834. doi:10.4236/ce.2015.68085

Zhang, J., Chen, Q., Sun, Y., & Reid, D. J. (2004). Triple scheme of learning support design for scientific discovery learning based on computer simulation: experimental research. *Journal of Computer Assisted Learning, 20*, 269–282. doi:10.1111/j.1365-2729.2004.00062.x

5.8 Supplemental Material

5.8.1 Example of an EvoSketch Worksheet task

Where could a point be placed that has influenced the development of your line?

a. Make a prediction:

The point	□ will appear at the top right.
	□ will appear at the top left.
	□ will appear at the bottom right.
	□ will appear at the bottom left.

b. Now press the button *Show point*

c. Describe where the point appeared in your animation:

The point	□ appeared at the top right.
	□ appeared at the top left.
	□ appeared at the bottom right.
	□ appeared at the bottom left.

5.8.2 Text intervention Worksheet tasks

1. Name the factors referred to as "mechanisms of evolution" in the "wingless flies" text and briefly explain them.
2. Indicate why the wingless flies on the islands have selective advantages compared to the normal (winged) flies.
3. Explain who or what "decides" whether a mutation is positive or negative. Argue whether the same mutation can be sometimes positive and sometimes negative.
4. Explain why albinos are rare in nature.
5. Justify why sexual reproduction has advantages over asexual reproduction when environmental conditions change.

5.8.3 Comparison of the original and revised worked example

Table 5.3 Comparison of the original and revised worked example

	Original Worked Example[1] **"How do species originate?"**	**Revised Worked Example** **"Why does a line change over the c ourse of 20 generations?"**
Solution steps	(1) Looking at differences (2) Locking at the chances of survival and reproduction (3) Looking at the consequences on biological fitness.	(1) Looking at differences (2) Locking at the chances of survival and reproduction (3) Looking at the consequences on biological fitness.
Type of prompting	Novice/expert prompts only, transition from novice to expert prompts	Expert prompts only
Example 1: Conditional factors of natural selection	Lemurs on Madagascar: a) variation of individuals in a species, b) heredity, and c) differential reproduction and survival	Evolution of the line (EvoSketch) a) origin of variation (mutation and recombination; focus on the first one as random process), b) variation of individuals in a population, c) heredity, and d) differential reproduction and survival (as probabilistic processes)
Example 2: Selection	Peppered moth	Peppered moth (no change)
Example 3: Speciation	Lemurs on Madagascar: a) reproductive isolation b) allopatric speciation	----- (excluded)

[1] cf. Neubrand, Borzikowsky, & Harms, 2016; German version: Neubrand, 2017

References

Neubrand, C. (2017). *Evolution unterrichten: Untersuchung von Aspekten erfolgreichen Lehrens und Lernens mit Fokus auf Beispielaufgaben und vorwissensangepasste Impulse [Teaching evolution: Investigate aspects of successful teaching and learning with focus on worked examples and knowledge adapted prompts]* (doctoral dissertation). Kiel University, Kiel, Germany.

Neubrand, C., Borzikowsky, C., & Harms, U. (2016). Adaptive prompts for learning evolution with worked examples - Highlighting the students between the "Novices" and the "Experts" in a classroom. *International Journal of Environmental and Science Education, 11*, 6774–6795.

6 STUDY 3: IS STATISTICAL REASONING RELEVANT FOR EVOLUTION EDUCATION?[3]

Abstract

Although a rich body of research has explored student's alternative conceptions of evolution and their relationships to understanding and acceptance of evolution, much less work has focused on the role of statistical thinking and reasoning. Thus, we aim to examine the relevance of statistical reasoning for understanding and accepting evolution by focusing on the relationship between these concepts and the extent to which statistical reasoning can predict understanding and acceptance of evolution. We recruited a large sample (N = 538) of undergraduate students enrolled in an evolution-focused introductory biology course to complete a suite of instruments that assessed students' statistical reasoning, understanding, and acceptance of evolution. Our results indicate that statistical reasoning is related to evolution understanding and acceptance. Furthermore, findings of the regression analyses revealed that statistical reasoning explains a fair proportion of variance of students' understanding and acceptance. Thus, our work provides new insight into the relationship of evolution understanding and acceptance. Moreover, we suggest that improving statistical reasoning could be a valuable addition to evolution instruction.

Keywords:

Evolution education, Statistical reasoning, Randomness and probability, Acceptance of evolution

[3] This is a pre-peer reviewed version of the following article: Fiedler, D., Sbeglia, G., Nehm, R. H., & Harms, U. (accepted with major revisions, 07/2018). How strongly does statistical reasoning influence knowledge and acceptance of evolution? *Journal of Research in Science Teaching (JRST).* This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

6.1 Introduction

Over the last decades, a rich body of research in evolution education deals with students' acceptance of evolution and its relationship to understanding the theory of evolution. Although the relationship between these two aspects is still elusive with epmirical research indicating positive connections (e.g., Barnes, Evans, Hazel, Brownell, & Nesse, 2017; Cofré, Cuevas, & Becerra, 2017; Deniz, Donnelly, & Yilmaz, 2008; Großschedl, Konnemann, & Basel, 2014; Kim & Nehm, 2011) as well as no connections (e.g., Cavallo & McCall, 2008; Coleman, Stears, & Dempster, 2015; Sinatra, Southerland, McConaughy, & Demastes, 2003), accepting evolution is intended to be a goal of evolution education (Smith & Siegel, 2016).

Moreover, the theory of evolution is the integrative framework of modern biology, and learning its essential tenets is broadly considered as a fundamental feature of science literacy (Nationale Akademie der Wissenschaften Leopoldina, 2017; Next Generation Science Standards, 2012). Still, many, including university students and biology teachers, not only lack an understanding of the processes of evolution, but also resist in accepting evolutionary theory as the best scientific explanation for similarities among organisms, biological diversity, and various features and processes of our world (Berkman & Plutzer, 2011; J. D. Miller, Scott, & Okamoto, 2006). In fact, empirical studies also indicate that teachers and students still struggle with teaching and learning evolution, respectively, and have various alternative conceptions (e.g., Kampourakis & Nehm, 2014; Rosengren, Brem, Evans, & Sinatra, 2012; To, Tenenbaum, & Hogh, 2017; Zabel & Gropengiesser, 2011). But, even though learners' knowledge of a particular concept is argued to be a relevant factor for acceptance of it (Smith & Siegel, 2016), other researchers propose several additional factors to be important for understanding and accepting evolution (e.g., Ha, Haury, & Nehm, 2012; Nadelson & Hardy, 2015). In this context, the general idea of conceptual ecology is often used as a theoretical framework to describe the influence of specific factors on understanding and acceptance (e.g., Athanasiou & Papadopoulou, 2015; Demastes, Good, & Peebles, 1995; Großschedl et al., 2014). Although the early framework focused particularly on the cognitive domain factors that govern conceptual change processes (Posner, Strike, Hewson, & Gertzog, 1982), the idea has been applied to acceptance and implements also cognitive, affective, and contextual domain factors (e.g., Deniz et al., 2008; Sinatra et al., 2003). Nevertheless, many of these domain factors remain to be investigated.

For instance, understanding (and through this acceptance) of evolution might be influenced by statistical reasoning and thinking, since recent research indicates learning difficulties with those aspects of evolution that are strongly related to underlying abstract concepts such as randomness and probability (e.g., Mead & Scott, 2010; Ross et al., 2010). In fact, students in the life sciences often have deficits in mathematical and statistical knowledge and reasoning (e.g., Jungck, 1997; Klatzky, Geiwitz, & Fischer, 1994; Konold, 1989). Additionally, a rich body of research in statistics education investigates students' understanding and misinterpretations regarding probability, chance, and randomness (e.g., Ben-Zvi & Garfield, 2008; Kahneman, Slovic, & Tversky, 1982; Shaughnessy, 2003). One of the major findings is that even people who can correctly calculate probabilities tend to apply faulty reasoning when asked about an uncertain event (Garfield & Ahlgren, 1988). Moreover, misunderstandings of randomness are mostly manifested during kindergarten and continuous to be hold throughout adulthood (K. E. Metz, 1998). These faulty understandings, or alternative conceptions of learners, are not only troublesome for statistics education, but are also relevant for explanations of scientific phenomena.

Indeed, students in the life sciences need a firm grounding in statistical concepts (Garfield, 2003). A clear understanding of randomness and probability is essential for understanding both evolution and molecular/cellular biology (Lenormand, Roze, & Rousset, 2009; Mead & Scott, 2010). Moreover, Tibell and Harms (2017) suggested a two-dimensional framework connecting principles and key concepts of evolution with underlying general abstract concepts such as randomness, probability, spatial scale, and temporal scale. They propose that a complete understanding of evolution does not only require the development of knowledge concerning both principles and abstract concepts, but also the ability to freely navigate through this two-dimensional framework (Tibell & Harms, 2017).

Nevertheless, randomness is often difficult to understand because of its different meanings in diverse contexts. In our everyday language, the notion of randomness is used to refer to situations that are without order, predictability or a pattern (Bennett, 1998). Hence, random processes are purposeless and directionless (Mead & Scott, 2010). In contrast, scientists (and mathematicians) use the term random to suggest unpredictability (also given our current state of knowledge), but do not mean a purposelessness (Buiatti & Longo, 2013). Moreover, the evolutionary notion of randomness is often quite specifically by referring to events (e.g., mutations, genetic drift) that are independent of an organisms' need or the directionality provided by natural selection in the process of adaptation (Eble, 1999; Millstein, 2000).

In contrast, probability is the likelihood of a particular outcome and is assigned a numerical value between zero and one (Feller, 1968). The closer a probability value is to one, the more likely the outcome. Students in evolution education have to deal with probabilities and probabilistic equations such as Hardy-Weinberg-Equilibrium (Masel, 2012). Additionally, natural selection itself can be described as a probabilistic process, because which organism may survive to produce more offspring and which may not survive or may produce less offspring depends on the hereditary traits in relation to the surrounding environment (Buiatti & Longo, 2013; Mayr, 2001).

Thus, a firm grounding of understanding randomness and probability in different contexts is crucial for understanding evolution (Tibell & Harms, 2017). In fact, recent research in biology education starts to take factors into account that are connected to mathematics such as statistical thinking, quantitative reasoning or emotions about mathematics (e.g., Chiel, McManus, & Shaw, 2010; Hester, Buxner, Elfring, & Nagy, 2014; A. M. Metz, 2008; Schuchardt & Schunn, 2016; Stanhope et al., 2017; Wachsmuth, Runyon, Drake, & Dolan, 2017). Still, until now, there is no empirical research – at least to our knowledge – that examined the relationship between statistical reasoning, understanding, and accepting evolution. Therefore, the presented research aims to examine these relationships by focusing on the question if statistical reasoning can predict evolutionary knowledge and acceptance of evolution. Thus, our study asks: To what extent does statistical reasoning predict understanding and acceptance of evolution?

6.2 Methods

6.2.1 Participants and Course Description

Students participating in this study were enrolled in a semester-long introductory biology course (BIO 201) at Stony Brook University, which is taken by both biology majors and non-science majors. The prerequisites for this course are high school biology and freshman-level math, neither of which typically include statistical instruction or a particular focus on randomness and probability. The course content is aligned with the five core concepts of biological literacy listed in the American Association for the Advancement of Science's Vision and Change policy document (AAAS, 2011). Thus, evolution is one of these core concepts and a constant theme throughout the course. Randomness and probability were discussed briefly in so far as they relate to evolution but these topics did not receive focused attention. The course is taught by two evolutionary biologists. Enrollment in spring semester 2017 was a total of 538 undergraduates (56.1% females) with a mean age of 19.4 years (*SD* = 1.6 years; range from 17 to 32 years). The race/ethnicity of participants was 40.5% Asian, 8.1% African American, 33.9% European American, 11.4% Hispanic of any race, and 6.2% others. Most students were juniors (37.4%) or seniors (38.5%), with 58% of the students planning to major in biology. For 73.4% of students in our sample, English was their first language. This study was approved by the university's Institutional Review Board (IRB), and students in this study consented to participate and have de-identified data published.

6.2.2 Instruments

To investigate the relationships among statistical reasoning, understanding, and acceptance of evolution, we requested students at the end of their course to complete online diagnostic tests consisting of instruments that assess their conceptual knowledge of randomness and probability, evolutionary knowledge, and acceptance of evolution. The specifically used instruments are outlined below.

Randomness and Probability test in the context of Evolution/Mathematics (RaProEvo/RaProMath). The RaProEvo and RaProMath are two distinct instruments to measure students' conceptual knowledge of randomness and probability in the context of evolution (RaProEvo) and mathematics (RaProMath; Fiedler, Tröbst, & Harms, 2017). The RaProEvo instrument consists of 21 items (16 multiple-choice, three free-response, and two matching items), while the RaProMath contains 33 items (30 multiple-choice, and three free-response items). The validity of each instrument was established by expert ratings, statistical analyses of student responses based on IRT, and criterion-related validity measures (Fiedler et al., 2017). All items were designed to have one correct answer, except for two items in the RaProEvo that were partial-credit scored (0 = no credit, 1 = partial-credit, 2 = full credit). For the presented study, the accuracy of measurement calculated by Expected A Posteriori/Plausible Value (EAP/PV) reliability that can be interpreted like Cronbach's alpha (Wu, Adams, Wilson, & Haldane, 2007), was 0.77 for the RaProEvo and 0.74 for the RaProMath.

Conceptual Assessment of Natural Selection (CANS). The CANS determines how well students understand the basic process of natural selection in a deep way (Kalinowski, Leonard, & Taper, 2016). The CANS consists of 24 multiple-choice items that assess five concepts related to natural selection (i.e., variation, selection, inheritance, mutation, and evolution as the interaction between the above-named concepts). The instrument is divided into four animal or plant contexts (i.e., anteaters, bowhead whales, saguaro cacti, and mosquitoes) and validity of the instrument was established by an expert panel, student interviews, and statistical analyses of student responses based on Item Response Theory (Kalinowski et al., 2016). The EAP/PV reliability of the instrument based on our sample was 0.82.

Inventory of Student Evolution Acceptance (I-SEA). The I-SEA measures students' evolution acceptance along three specific aspects such as microevolution, macroevolution, and human evolution (Nadelson & Southerland, 2012). This measure consists of 24 Likert-scale items (eight items per aspect) that employed a five-option response format (strongly disagree, disagree, undecided, agree, and strongly agree). The instrument includes items such as 'I think that new species evolve from a lot of small changes occurring over relatively long periods of time' (macroevolution), 'All groups of organisms will continue to change' (microevolution), and 'I think that humans and apes share an ancient ancestor' (human evolution). The validity of the I-SEA was established by a group of biology teachers, science teacher educators, and college biology faculties as well as classical psychometric statistical analyses of student responses. The I-SEA yielded an EAP/PV reliability of 0.93 for our sample.

Demographic data. Additionally, we also asked students' demographic variables, i.e., gender, age, race/ethnicity, class level, class plan, and English writing and reading skills.

6.2.3 Data Analysis

Students who did not respond to any of the instruments mentioned above were excluded from the analyses, resulting in a dataset of $n = 448$ cases. We employed Rasch analysis by using ACER ConQuest item response modeling software (version 4; Wu et al., 2007), because the Rasch model provides a means for converting ordinal rating scale or partial-credit data into interval data that can be used for quantitative analyses (Bond &

Fox, 2015; Wright & Mok, 2000). For further analyses, person parameters were estimated by calculating weighted maximum likelihood estimation (WLE) values.

As baseline equivalence, we verify the psychometric distinction of RaProEvo and RaProMath. The two instruments are regarded to measure two distinct competencies, but the psychometric distinction was initially verified in a small sample size with students of different ages (Fiedler et al., 2017). To check this distinction in our larger sample, we compared a two-dimensional model (RaProEvo vs. RaProMath) against a one-dimensional model (RaProEvo & RaProMath), applied χ^2-tests (Bentler, 1990), and additional information-based criteria such as Akaike's information criterion (AIC; Akaike, 1981) and Bayes's information criterion (BIC; Wilson, de Boeck, & Carstensen, 2008). Both AIC and BIC criteria do not enable tests of the significance of differences between models, but smaller values indicated better fits of the data to the model (Wilson et al., 2008).

Afterwards, we determine if personal characteristics were related to statistical reasoning, evolutionary knowledge, and acceptance of evolution by calculating correlation analyses, t-tests, and one-way analyses of variance (ANOVAs).

We then conducted several regression analyses to determine if statistical reasoning alone or in combination with other relevant variables is predictive for evolutionary knowledge and acceptance of evolution. At first, we examined if RaProEvo alone or in combination with RaProMath serves as a predictor for evolutionary knowledge and acceptance. We choose RaProEvo as a predictor of the first step due to two reasons: firstly, RaProEvo measures conceptual knowledge of randomness and probability in the same context as acceptance is assessed, while secondly, our first study also indicated higher correlations of knowledge and acceptance with RaProEvo than RaProMath. In a third step, we included personal characteristics to the model. For acceptance of evolution, we also included CANS as a fourth step to the model.

6.3 Results

6.3.1 Baseline Equivalence

To examine if conceptual knowledge of randomness and probability in the context of evolution is psychometrically distinct to the context of mathematics, we compared a two-dimensional to a one-dimensional partial-credit model fitted to the data of 448 student responses. Information-based criteria values (AIC and BIC) indicate that the two-dimensional model provides a better fit to the data (see Table 6.1), while the results of the χ^2-test also revealed that the two-dimensional model significantly outperformed the one-dimensional model: χ^2 $(2, n = 448) = 30.06$, $p < .001$.

Table 6.1 Dimensionality test results

No. of dimensions	Deviance (no. of free parameters)	AIC	BIC
One	21476.14 (54)	21584.14	21814.06
Two	21446.09 (56)	21558.09	21796.52

Note. AIC = Akaike's Information Criterion; BIC = Bayes' Information Criterion.

Additionally, we generated a joint Wright map from the two-dimensional model (Figure 6.1) to analyze the internal structure of the RaProEvo and RaProMath (Boone & Rogan, 2005). In such a map, the distributions of persons and items of the instruments (or more strictly person ability and item difficulty estimates) are plotted along the same dimension (conventionally to the left and right, respectively) and can be directly compared. Items of equivalent difficulty are located at the same position on the scale (e.g., E07 and M30; Figure 6.1), and persons at the same position or height on the scale as a particular item have a 50% chance of answering that item correctly, while those located above and below an item have a 50% higher and lower chance of answering it correctly, respectively. Our joint Wright map reveals that a typically respondent of our sample would answer most of the RaProEvo items and two-thirds of the RaProMath items.

6.3.2 Personal Characteristics

Findings of our correlation analyses (see Table 6.2) indicated that RaProEvo test score was significantly related to higher age, class level, class plan, and self-rated English reading and writing skills, while I-SEA was only significant positive related to self-rated English reading and writing skills. In contrast, RaProMath test score was significantly negatively related to age, meaning that the older the students, the lower their RaProMath scores.

Table 6.2 Intercorrelations between personal characteristics and statistical reasoning (RaProEvo, RaProMath), evolutionary knowledge, and acceptance of evolution (n = 423)

				English skills	
Measures	Age	Class level	Class plan	Reading	Writing
RaProEvo	-.07	.10*	.14**	.17**	.15**
RaProMath	-.12*	.07	.04	.09	.07
Knowledge evolution	-.09	.01	.01	.07	.09
Acceptance evolution	-.05	.09	.06	.22***	.19***

Note. RaProEvo = conceptual knowledge of randomness and probability in evolution; *RaProMath* = conceptual knowledge of randomness and probability in mathematics; *Class level:* 1 = freshman, 2 = sophomore, 3 = junior, 4 = senior, 5 = entry level masters; *Class plan:* 1 = Non-STEM, 2 = Non-Biology STEM, 3 = Biology STEM; *English skills Reading/Writing:* six-point rating scale (0 = *very poor*, 5 = *excellent*).
* $p < .05$; ** $p < .01$; *** $p < .001$.

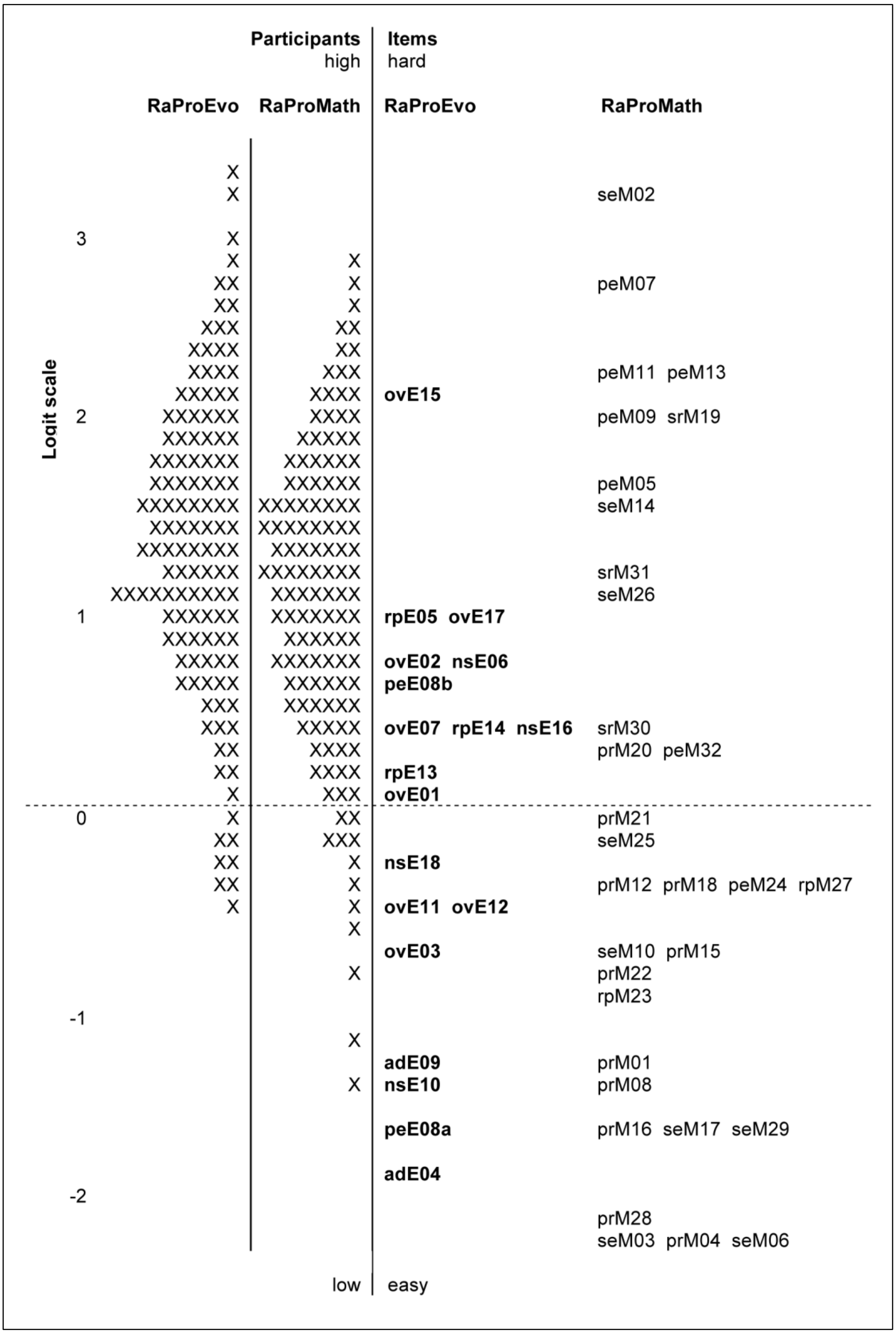


Figure 6.1 Joint Wright map of students' responses to items linked to the RaProEvo (bold; *n* = 448; 19 items) and RaProMath (*n* = 448; 32 items) test. Abilities of persons who took the test are displayed on the left and the difficulty of the items on the right. Each X indicates 3.4 individuals in our sample. The first two letters stand for: *ov* origin of variation, *ad* accidental death (single event), *rp* random phenomena, *se* single event, *ns* process of natural selection, *pe* probability of events, *pr* probability as ratio, and *sr* sample reasoning (after Fiedler et al., 2017). *E01* to *E18* indicates the item number in the RaProEvo test, while *M01* to *M32* represents the item number in the RaProMath test. The last letter stands for *a* item 1 and *b* item 2 within a similar task.

Our findings of the *t*-tests regarding RaProEvo indicated that, on average, males (M = 1.25, SD = 1.04, n = 179) had significantly higher scores compared to females (M = 0.98, SD = 0.88, n = 179; $t(431) = 2.90$, $p = .004$). Moreover, males (M = 3.88, SD = 2.17, n = 179) had also significantly higher acceptance scores than females (M = 3.32, SD = 1.99, n = 179; $t(361.43) = 2.69$, $p = .007$). In contrast, we detected no significant differences in RaProEvo and evolutionary knowledge scores.

At least, race/ethnicity had a main effect on RaProEvo, RaProMath, and I-SEA (see Table 6.3). European Americans had higher scores compared to Asians, African Americans, or Hispanics of any race. Additionally, African Americans scored lower on acceptance of evolution compared to Hispanics of any race. Nevertheless, no differences were detected among the other comparisons or regarding scores of evolutionary knowledge.

Table 6.3 Effects of ethnicity on conceptual knowledge of randomness and probability (RaProEvo, RaProMath), evolutionary knowledge, and acceptance of evolution (n = 437)

	Race/ethnicity (n)							**Post-hoc Bonferroni mean difference**[a]			
Measures	**1** (176)	**2** (33)	**3** (151)	**4** (48)	**5** (29)	***df***	***F***	**1–3**	**2–3**	**2–4**	**3–5**
RaProEvo	1.23 (1.00)	1.08 (0.94)	1.64 (0.93)	1.18 (0.87)	1.36 (0.81)	4,432	5.31***	-0.41**	-0.56*	-0.10	0.46**
RaProMath	1.03 (1.03)	0.39 (0.83)	1.35 (0.90)	0.74 (0.86)	1.08 (0.78)	4,432	6.32***	-0.32*	-0.66**	-0.05	0.62**
Knowledge evolution	0.86 (1.16)	0.61 (1.28)	1.12 (1.24)	0.70 (1.49)	0.94 (1.11)	4,432	1.98	-0.26	-0.51	-0.09	0.42
Acceptance evolution	3.09 (1.92)	2.28 (1.68)	4.33 (2.08)	3.68 (2.06)	3.43 (2.06)	4,432	11.59***	-1.23***	-2.05***	-1.40*	0.65

Note. RaProEvo = conceptual knowledge of randomness and probability in evolution; *RaProMath* = conceptual knowledge of randomness and probability in mathematics; *Race/ethnicity*: 1 = Asian, 2 = African American, 3 = European American, 4 = Hispanic of any race, 5 = others.
[a] all other comparisons with $p > .05$.
* $p < .05$; ** $p < .01$; *** $p < .001$.

6.3.3 Regression analyses

At first, we conducted a correlation analyses of all instruments to determine preliminary relationships between our variables of interest and if they overly correlated. The results revealed significant positive moderate to strong correlations among the instruments (see Table 6.4). Still, there remained unique variance to each instrument, indicating that the instruments assessing unique aspects of personal perspectives.

Table 6.4 Descriptive statistics and intercorrelations of the instruments ($n = 448$)

	raw scores		WLE values		Intercorrelation		
Measures	***M***	***SD***	***M***	***SD***	**RaProMath**	**CANS**	**I-SEA**
RaProEvo	15.07	3.17	1.36	0.96	.57***	.50***	.42***
RaProMath	21.50	4.34	1.09	0.95	---	.49***	.39***
CANS	15.89	4.95	0.91	1.23		---	.35***
I-SEA	4.41	0.53	3.56	2.10			---

Note. RaProEvo = Randomness and Probability test in the context of Evolution; RaProMath = Randomness and Probability test in the context of Mathematics; *CANS* = Conceptual Assessment of Natural Selection; *I-SEA* = Inventory of Student Evolution Acceptance.
*** $p < .001$.

Indeed, our regression analyses of evolutionary knowledge (see Table 6.5) indicates that conceptual knowledge of randomness and probability in evolution account for 21.9% variance, $F(1,417) = 117.14$, $p < .001$. The addition of RaProMath to the model increased the variance explained to 27.7%, $F(1,416) = 33.44$, $p < .001$. Thus, conceptual knowledge of randomness and probability in mathematics can explain 5.8% more variance in evolutionary knowledge. In contrast, the inclusion of personal characteristics changed the variance explained to 28.3% without being significant, $F(1,406) = 0.32$, $p = .976$. Overall, statistical reasoning can be regarded as a predictor for evolutionary knowledge and explains a great proportion of variance.

Concerning our regression analyses of acceptance of evolution (see Table 6.5), findings indicate that conceptual knowledge of randomness and probability also accounts for 15.3% variance of acceptance, $F(1,417) = 75.36$, $p < .001$. Moreover, the addition of RaProMath to the model again increased the variance explained to 19.3%, $F(1,416) = 20.73$, $p < .001$, and thus, can explain 4.0% additional variance. The addition of personal characteristics in step 3 increased the variance to additional 7.7%, $F(1,406) = 4.31$, $p < .001$, resulting in a model that explained 27.1% variance of acceptance of evolution. In this model, acceptance of evolution is significant positive influences by statistical reasoning (RaProEvo and RaProMath) but significant negative by gender, English reading skills, and race/ethnicity. In other words, acceptance of evolution increase with higher scores in RaProEvo and RaProMath, but decreases significantly for females compared to males, and Asians, as well as African Americans, compared to European Americans (the reference group). At least, the inclusion of evolutionary knowledge to the model increased the variance explained to a final proportion of 28.0%, $F(1,405) = 5.07$, $p = .025$. The significant predictors for this model are the same as before (i.e., statistical reasoning, gender, English reading skills, and race/ethnicity) including evolutionary knowledge.

Table 6.5 Summary of hierarchical (or sequential) regression analyses for variables explaining evolution knowledge and acceptance of evolution (*n* = 419)

	Knowledge evolution				Acceptance evolution			
Predictor(s)	***B***	***SE B***	**β**	***R²***	***B***	***SE B***	**β**	***R²***
Step 1				.22				.15
RaProEvo	0.61	0.06	.47***		0.86	0.10	.39***	
Step 2				.28				.19
RaProEvo	0.39	0.07	.30***		0.54	0.12	.25***	
RaProMath	0.39	0.07	.30***		0.54	0.12	.25***	
Step 3				.28				.27
RaProEvo	0.40	0.07	.31***		0.42	0.12	.19***	
RaProMath	0.39	0.07	.30***		0.47	0.12	.21***	
Age	0.01	0.03	.01		-0.04	0.06	-.03	
Gender	0.06	0.11	.02		-0.50	0.19	-.12**	
Class level	-0.01	0.01	-.04		-0.01	0.01	-.02	
Class plan	-0.07	0.08	-.04		0.06	0.14	.02	
English reading skills	-0.11	0.11	-.08		0.38	0.19	.16*	
English writing skills	0.06	0.10	.05		-0.09	0.17	-.04	
Race/ethnicity								
Asian	0.02	0.13	.01		-0.80	0.22	-.19***	
African American	0.01	0.22	.01		-1.42	0.37	-.18***	
Hispanic of any race	0.07	0.19	.02		-0.13	0.32	-.02	
others	-0.01	0.22	-.01		-0.50	0.39	-.06	
Step 4				--				.28
RaProEvo	--	--	--		0.35	0.12	.16**	
RaProMath	--	--	--		0.40	0.12	.18**	
Age	--	--	--		-0.04	0.06	-.03	
Gender	--	--	--		-0.51	0.19	-.12**	
Class level	--	--	--		0.00	0.01	-.01	
Class plan	--	--	--		0.07	0.14	.02	
English reading skills	--	--	--		0.40	0.19	.17*	
English writing skills	--	--	--		-0.10	0.17	-.05	
Race/ethnicity								
Asian	--	--	--		-0.81	0.22	-.19***	
African American	--	--	--		-1.43	0.37	-.18***	
Hispanic of any race	--	--	--		-0.14	0.32	-.02	
others	--	--	--		-0.50	0.38	-.06	
Knowledge evolution	--	--	--		0.19	0.08	.11**	

Knowledge evolution: R^2 = .219 for step 1; ΔR^2 = .058 for step 2; ΔR^2 = .006 for step 3.
Acceptance evolution: R^2 = .153 for step 1; ΔR^2 = .040 for step 2; ΔR^2 = .077 for step 3; ΔR^2 = .009 for step 4.
* $p < .05$; ** $p < .01$; *** $p < .001$.
Note. B = unstandardized regression coefficient; *SE B* = standard error; β = standardized regression coefficient; *RaProEvo* = conceptual knowledge or randomness and probability in evolution; *RaProMath* = conceptual knowledge of randomness and probability in mathematics; *Gender:* 0 = male, 1 = female; *Class level:* 1 = freshman, 2 = sophomore, 3 = junior, 4 = senior, 5 = entry level masters; *Class plan:* 1 = Non-STEM, 2 = Non-Biology STEM, 3 = Biology STEM; *English skills Reading/Writing:* six-point rating scale (0 = *very poor*, 5 = *excellent*); Race/ethnicity was coded as a dummy variable with European American as the omitted or reference group.

6.4 Discussion

Acceptance of evolution has been associated with multiple variables including understanding of evolution (e.g., Großschedl et al., 2014; Rice, Clough, Olson, Adams, & Colbert, 2015), understanding of nature of science (e.g., Akyol, Tekkaya, Sungur, & Traynor, 2012; Dunk, Petto, Wiles, & Campbell, 2017; Kim & Nehm, 2011), thinking disposition (e.g., Athanasiou, Katakos, & Papadopoulou, 2012; Deniz et al., 2008), and religious factors (e.g., Athanasiou & Papadopoulou, 2015; Deniz, Çetin, & Yılmaz, 2011). Still, statistical reasoning seems to be a neglected, but relevant factor for understanding and accepting evolution.

Our baseline results (dimensionality check) indicated that students' conceptual knowledge of randomness and probability in the context of evolution and mathematics are two distinct, but related competencies. Thus, we could replicate the findings provided in the literature showing that German university students' conceptual knowledge of randomness and probability in the context of evolution and mathematics are also two distinct, but related competencies (Fiedler et al., 2017). Moreover, our joint Wright map revealed comparable patterns regarding the distribution of the items. Most students could answer items concerning the process of natural selection (board probabilistic aspects of the process). Items regarding the origin of variation were distributed along the scale, but seem to be a bit more difficult for this sample of students compared to the German sample (Fiedler et al., 2017). Nevertheless, the patterns in the mathematical context (RaProMath) indicate similarities to the German sample. Most students could answer questions concerning probability as a ratio, while only high performers correctly answered items regarding the probability of events. In contrast, single event items were distributed across the scale.

Incorrect answers on items of the RaProMath test reflect three major statistical reasoning misconceptions concocted to understanding randomness and probability: the outcome orientation (i.e., making yes or no decisions based in intuitive models of a single events probability rather than looking at the series of events; Konold, 1989), the representative misconceptions (i.e., estimation of the probability of uncertain events based on the degree to which this sample reflects the population; Hirsch & O'Donnell, 2001; Kahneman et al., 1982), and the equiprobabillity bias (i.e., interpretation of events as equally likely due to insensitivity of prior events or the sample size; Lecoutre, 1992). Moreover, even though RaProMath and RaProEvo items are not directly comparable, we suggest that these faulty reasoning patterns may be an underlying explanation for faulty reasoning patterns in the RaProEvo items. Some items (e.g. E02, E08, E09, E13) are more connected to the well-known bias of statistical reasoning mentioned above, while others can be interpreted as an extension of underlying faulty intuitive models of probability (e.g., items on the process of natural selection; E06, E10, E16, E18).

Besides, concerning our regression analyses conceptual knowledge of randomness and probability in the context of mathematics (RaProMath), and acceptance of evolution revealed gender differences between the scores, resulting in males scoring higher than females. Gender was consistently identified as an important factor associated with performance on math assessments (e.g., Brown & Josephs, 1999; Hyde, Fennema, & Lamon, 1990; H. Miller & Bichsel, 2004) although more recent studies indicate that this gender gap is disappearing (e.g., Guiso, Monte, Sapienza, & Zingales, 2008; Hyde, Lindberg, Linn, Ellis, & Williams, 2008). Thus, our result might somehow be unexpected in this sense, but still confirm recent results of Flanagan and Einarson (2017), who also found gender differences in undergraduate biology students' math performance. In contrast, gender differences regarding acceptance are more diverse. While research indeed found gender differences regarding the MATE instrument (e.g., Großschedl et al., 2014; Kim & Nehm, 2011), such effects were not yet found in the I-SEA instrument (Nadelson & Hardy, 2015). Still, our results indicate that male students showed higher acceptance scores in contrast to their female counterparts. But other personal characteristic such as race/ethnicity was also significantly associated with conceptual knowledge of randomness and probability in mathematics (RaProMath) and acceptance of evolution. In fact, research proposed that ethnicity may account for more variance in science attitude and achievement differences than gender (e.g., Catsambis, 1995; Greenfield, 1996). Gender is maybe mediated through ethnicity effects.

Nevertheless, conceptual knowledge of randomness and probability in the context of both evolution and mathematics were predictive of both knowledge and acceptance of evolution. The explained variance of evolutionary knowledge predicted by statistical reasoning was in a range of 22% to 28%, while the explained variance of acceptance of evolution was between 15% and 19%. Although there is not much research about the actual connection between statistical reasoning and understanding evolution, there is a rich body of research of acceptance of evolution. Regarding this literature, our proportions of explained variance are somewhat similar to other significant factors explaining variance in acceptance of evolution such as the understanding of nature of science or thinking disposition (see Figure 6.2). Indeed, statistical reasoning is situated among the other relevant factors. In addition, recent research indicates students struggle with randomness and probability in the context of evolutionary theory (e.g., Garvin-Doxas & Klymkowsky, 2008; Mead & Scott, 2010). Thus, our research not only indicates new insights into the relationship of understanding and accepting evolution. We also state that statistical reasoning might be a neglected but relevant factor for understanding and accepting evolution, and thus, would be a valuable addition to evolution instruction.

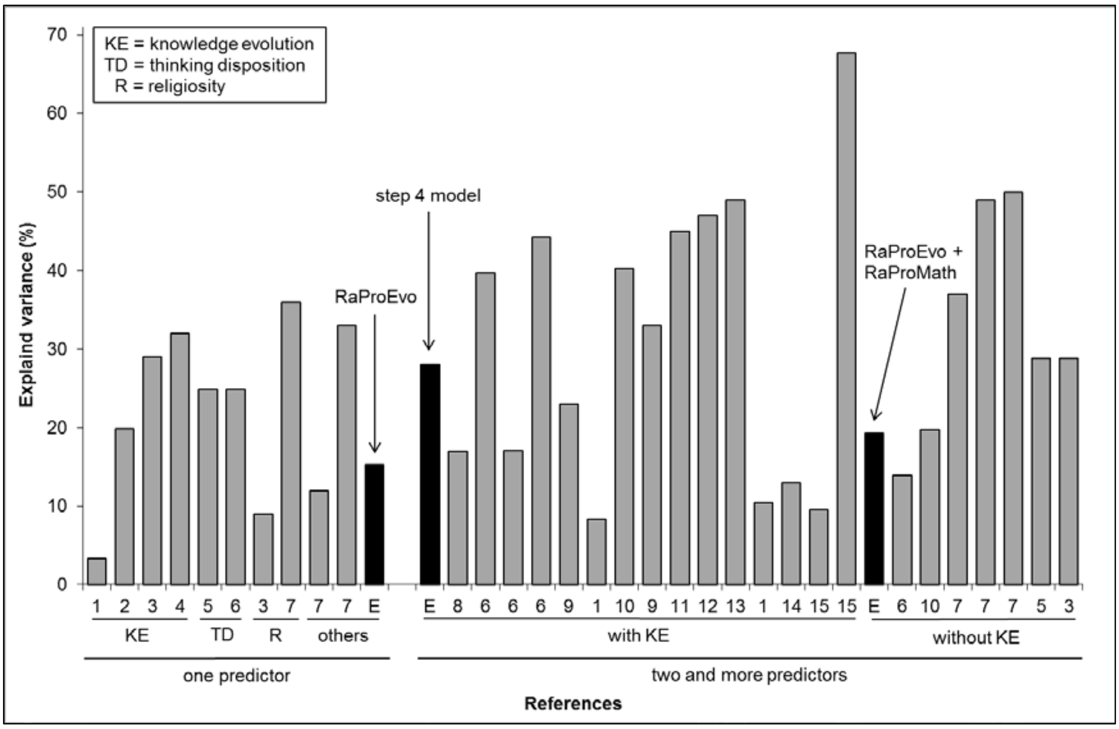


Figure 6.2 Empirical research studies investigating predictors of acceptance of evolution (grey) and our findings of the regression analyses investigating statistical reasoning alone or in combination with other variables as a possible predictor (black). Numbers appearing more than once are indicating different cohorts or pre/posttest data within a study. A list of the exact values and predictors as well as a reference list connected to the reference numbers are available as supplementary material accompanying the online article (see Supplemental Material 6.8.1 and 6.8.2).

6.4.1 Limitations

As it is often the case with studies that use data of a single momentum, we could only indicate that there might be somehow a connection between statistical reasoning, understanding, and acceptance of evolution. Nevertheless, we are not able to state which variable might influence the other – or if they influence each other at all. Thus, more research expanding our cross-sectional data is needed to examine how the concepts are related to each other and which concept does influence the other.

Moreover, we measured statistical reasoning at the end of an introductory biology course, but do not have any data that shows how high students' statistical reasoning was at the beginning of such a course. This means that we could not detect if there is already a change in students' statistical reasoning by attending an introductory biology course.

At least, this study was conducted in a single class at a particular institution. The patterns that exist here may not apply to other institutions or classes. Examining whether these patterns persist among different classes and at other institutions with different students is an essential next step.

6.5 Conclusions

In this study, we had explored the relationships among students' statistical reasoning (in two contexts), understanding, and acceptance of evolution. We found that statistical reasoning is significant positive connected to evolutionary knowledge and acceptance of evolution. Personal characteristics such as gender, English reading skills, and race/ethnicity were also significantly associated with the outcome variable acceptance of evolution measured by the I-SEA. Still, statistical reasoning seems to serve as a reliable predictor of both evolutionary knowledge and acceptance of evolution. Thus, we would like to extend research in evolution education by implementing understanding of underlying abstract concepts such as randomness and probability as significant factors, which can improve understanding of evolution, and through this might also enlighten the acceptance of evolution.

6.6 Acknowledgments

The study reported here was part of the Swedish-German cooperation project *"EvoVis: Challenging Threshold Concepts in Life Science – enhancing understanding of evolution by visualization"*, funded by the Swedish Research Council. We are very thankful to the students who participated in this study.

6.7 References

Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics, 16,* 3–14. doi:10.1016/0304-4076(81)90071-3

Akyol, G., Tekkaya, C., Sungur, S., & Traynor, A. (2012). Modeling the interrelationships among pre-service science teachers' understanding and acceptance of evolution, their views on nature of science and self-efficacy beliefs regarding teaching evolution. *Journal of Science Teacher Education, 23*, 937–957. doi:10.1007/s10972-012-9296-x

American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: A call to action.* Washington, DC: AAAS.

Athanasiou, K., Katakos, E., & Papadopoulou, P. (2012). Conceptual ecology of evolution acceptance among Greek education students: the contribution of knowledge increase. *Journal of Biological Education, 46*, 234–241. doi:10.1080/00219266.2012.716780

Athanasiou, K., & Papadopoulou, P. (2015). Evolution Theory Teaching and Learning: What Conclusions Can We Get From Comparisons of Teachers' and Students' Conceptual Ecologies in Greece and Turkey? *Eurasia Journal of Mathematics, Science & Technology Education, 11*, 841–853. doi:10.12973/eurasia.2015.1443a

Barnes, M. E., Evans, E. M., Hazel, A., Brownell, S. E., & Nesse, R. M. (2017). Teleological reasoning, not acceptance of evolution, impacts students' ability to learn natural selection. *Evolution: Education and Outreach, 10*, 1–12. doi:10.1186/s12052-017-0070-6

Ben-Zvi, D., & Garfield, J. (2008). Introducing the emerging discipline of statistics education. *School Science and Mathematics, 108*, 355–361. doi:10.1111/j.1949-8594.2008.tb17850.x

Bennett, D. J. (1998). *Randomness.* Cambrigde, United Kingdom: Harvard University Press.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.

Berkman, M. B., & Plutzer, E. (2011). Defeating creationism in the courtroom, but not in the classroom. *Science, 331*, 404–405. doi:10.1126/science.1198902

Bond, T., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). New York, NY: Routledge.

Boone, W., & Rogan, J. (2005). Rigour in quantitative analysis: The promise of Rasch analysis techniques. *African Journal of Research in Mathematics, Science and Technology Education, 9*, 25–38.

Brown, R. P., & Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology, 76*, 246. doi:10.1037//0022-3514.76.2.246

Buiatti, M., & Longo, G. (2013). Randomness and multilevel interactions in biology. *Theory in Biosciences, 132*, 139–158. doi:10.1007/s12064-013-0179-2

Catsambis, S. (1995). Gender, race, ethnicity, and science education in the middle grades. *Journal of Research in Science Teaching, 32*, 243–257. doi:10.1002/tea.3660320305

Cavallo, A. M., & McCall, D. (2008). Seeing may not mean believing: examining students' understandings & beliefs in evolution. *The American Biology Teacher, 70*, 522–530. doi:10.1662/0002-7685-70.9.522

Chiel, H. J., McManus, J. M., & Shaw, K. M. (2010). From biology to mathematical models and back: teaching modeling to biology students, and biology to math and engineering students. *CBE-Life Sciences Education, 9*, 248–265. doi:10.1187/cbe.10-03-0022

Cofré, H., Cuevas, E., & Becerra, B. (2017). The relationship between biology teachers' understanding of the nature of science and the understanding and acceptance of the theory of evolution. *International Journal of Science Education, 39*, 2243–2260. doi:10.1080/09500693.2017.1373410

Coleman, J., Stears, M., & Dempster, E. (2015). Student teachers' understanding and acceptance of evolution and the nature of science. *South African Journal of Education, 35*, 1–9. doi:10.15700/saje.v35n2a1079

Demastes, S. S., Good, R. G., & Peebles, P. (1995). Students' conceptual ecologies and the process of conceptual change in evolution. *Science Education, 79*, 637–666.

Deniz, H., Çetin, F., & Yılmaz, I. (2011). Examining the relationships among acceptance of evolution, religiosity, and teaching preference for evolution in Turkish preservice biology teachers. *Reports of the National Center for Science Education, 31*.

Deniz, H., Donnelly, L. A., & Yilmaz, I. (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*, 420–443. doi:10.1002/tea.20223

Dunk, R. D., Petto, A. J., Wiles, J. R., & Campbell, B. C. (2017). A multifactorial analysis of acceptance of evolution. *Evolution: Education and Outreach, 10*, 1–8. doi:10.1186/s12052-017-0068-0

Eble, G. J. (1999). On the dual nature of chance in evolutionary biology and paleobiology. *Paleobiology, 25*, 75–87. doi:10.1666/0094-8373(1999)025<0075:OTDNOC>2.3.CO;2

Feller, W. (1968). *An introduction to probability theory and its applications* (Vol. 1). London, United Kingdom: John Wiley & Sons.

Fiedler, D., Tröbst, S., & Harms, U. (2017). University students' conceptual knowledge of randomness and probability in the contexts of evolution and mathematics. *CBE-Life Sciences Education, 16,* ar38. doi:10.1187/cbe.16-07-0230

Flanagan, K., & Einarson, J. (2017). Gender, Math Confidence, and Grit: Relationships with Quantitative Skills and Performance in an Undergraduate Biology Course. *CBE-Life Sciences Education, 16,* ar47. doi:10.1187/cbe.16-08-0253

Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal, 2*, 22–38.

Garfield, J. B., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education, 19*, 44–63. doi:10.2307/749110

Garvin-Doxas, K., & Klymkowsky, M. W. (2008). Understanding randomness and its impact on student learning: lessons learned from building the Biology Concept Inventory (BCI). *CBE-Life Sciences Education, 7*, 227–233. doi:10.1187/cbe.07-08-0063

Greenfield, T. A. (1996). Gender, ethnicity, science achievement, and attitudes. *Journal of Research in Science Teaching, 33*, 901–933. doi:10.1002/(Sici)1098-2736(199610)33:8<901::Aid-Tea5>3.0.Co;2-#

Großschedl, J., Konnemann, C., & Basel, N. (2014). Pre-service biology teachers' acceptance of evolutionary theory and their preference for its teaching. *Evolution: Education and Outreach, 7*, 1–16. doi:10.1186/s12052-014-0018-z

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). *Culture, Gender, and Math. Science, 320*, 1164.

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

Hester, S., Buxner, S., Elfring, L., & Nagy, L. (2014). Integrating quantitative thinking into an introductory biology course improves students' mathematical reasoning in biological contexts. *CBE-Life Sciences Education, 13*, 54–64. doi:10.1187/cbe.13-07-0129

Hirsch, L. S., & O'Donnell, A. M. (2001). Representativeness in statistical reasoning: Identifying and assessing misconceptions. *Journal of Statistics Education, 9*. doi:10.1080/10691898.2001.11910655

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological Bulletin, 107*, 139.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science, 321*, 494–495.

Jungck, J. R. (1997). Ten equations that changed biology: mathematics in problem-solving biology curricula. *BioScience, 23*, 11–36.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge, United Kingdom: Cambridge University Press.

Kalinowski, S. T., Leonard, M. J., & Taper, M. L. (2016). Development and validation of the Conceptual Assessment of Natural Selection (CANS). *CBE-Life Sciences Education, 15*, ar64.

Kampourakis, K., & Nehm, R. H. (2014). History and philosophy of science and the teaching of evolution: students' conceptions and explanations. In M. Mathews (Ed.), *International Handbook of Research in History, Philosophy and Science Teaching* (pp. 377–399). Dodrecht, The Netherlands: Springer.

Kim, S. Y., & Nehm, R. H. (2011). A Cross-Cultural Comparison of Korean and American Science Teachers' Views of Evolution and the Nature of Science. International *Journal of Science Education, 33*, 197–227. doi:10.1080/09500690903563819

Klatzky, R. L., Geiwitz, J., & Fischer, S. (1994). Using statistics in clinical practice: A gap between training and application. In M. S. Bogner (Ed.), *Human error in medicine* (pp. 123–140). Boca Raton, FL: CRC Press

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction, 6*, 59–98. doi:10.1207/s1532690xci0601_3

Lecoutre, M.-P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics, 23*, 557–568. doi:10.1007/BF00540060

Lenormand, T., Roze, D., & Rousset, F. (2009). Stochasticity in evolution. *Trends in Ecology & Evolution, 24*, 157–165. doi:10.1016/j.tree.2008.09.014

Masel, J. (2012). Rethinking Hardy–Weinberg and genetic drift in undergraduate biology. *BioEssays, 34*, 701–710. doi:10.1002/bies.201100178

Mayr, E. (2001). *What evolution is.* New York, NY: Basic books.

Mead, L. S., & Scott, E. C. (2010). Problem concepts in evolution part II: Cause and chance. *Evolution: Education and Outreach, 3*, 261–264. doi:10.1007/s12052-010-0231-3

Metz, A. M. (2008). Teaching statistics in biology: using inquiry-based learning to strengthen understanding of statistical analysis in biology laboratory courses. *CBE-Life Sciences Education, 7*, 317–326. doi:10.1187/cbe.07-07-0046

Metz, K. E. (1998). Emergent understanding and attribution of randomness: Comparative analysis of the reasoning of primary grade children and undergraduates. *Cognition and Instruction, 16*, 285–265.

Miller, H., & Bichsel, J. (2004). Anxiety, working memory, gender, and math performance. *Personality and Individual Differences, 37*, 591–606. doi:10.1016/j.paid.2003.09.029

Miller, J. D., Scott, E. C., & Okamoto, S. (2006). Public acceptance of evolution. *Science, 313,* 765–766. doi:10.1126/science.1126746

Millstein, R. L. (2000). Chance and macroevolution. *Philosophy of Science, 67*, 603–624. doi:10.1086/392857

Nadelson, L. S., & Hardy, K. K. (2015). Trust in science and scientists and the acceptance of evolution. *Evolution: Education and Outreach, 8*, 1–9. doi:10.1186/s12052-015-0037-4

Nadelson, L. S., & Southerland, S. (2012). A more fine-grained measure of students' acceptance of evolution: development of the Inventory of Student Evolution Acceptance—I-SEA. International *Journal of Science Education, 34*, 1637–1666. doi:10.1080/09500693.2012.702235

Nationale Akademie der Wissenschaften Leopoldina. (2017). *Evolutionsbiologische Bildung in Schule und Universität [Teaching evolutionary biology at school and university]*. Deutsche Akademie der Naturforscher Leopoldina e.V. Nationale Akademie der Wissenschaften, Halle (Saale). Retreived from https://www.leopoldina.org/uploads/tx_leopublication/2017_Stellungnahme_Evolutionsbiologie.pdf (accessed 03 April 2018)

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States.* Washington, DC: The National Academies Press.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*, 211–227. doi:10.1002/sce.3730660207

Rice, J. W., Clough, M. P., Olson, J. K., Adams, D. C., & Colbert, J. T. (2015). University faculty and their knowledge & acceptance of biological evolution. *Evolution: Education and Outreach, 8*, 1–15. doi:10.1186/s12052-015-0036-5

Rosengren, K. S., Brem, S. K., Evans, E. M., & Sinatra, G. M. (2012). *Evolution challenges: Integrating research and practice in teaching and learning about evolution.* New York, NY: Oxford University Press.

Ross, P. M., Taylor, C. E., Hughes, C., Kofod, M., Whitaker, N., Lutze-Mann, L., & Tzioumis, V. (2010). Threshold concepts: Challenging the way we think, teach and learn in biology. In J. H. F. Meyer, R. Land, & C. Baillie (Eds.), *Threshold Concepts and Transformational Learning* (Vol. 1, pp. 165–177). Rotterdam, The Netherlands: SensPublishers.

Schuchardt, A. M., & Schunn, C. D. (2016). Modeling scientific processes with mathematics equations enhances student qualitative conceptual understanding and quantitative problem solving. *Science Education, 100*, 290–320. doi:10.1002/sce.21198

Shaughnessy, J. M. (2003). Research on students' understandings of probability. In J. Kilpatric, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 216–226). Reston, VA: National Council of Teachers of Mathematics.

Sinatra, G. M., Southerland, S. A., McConaughy, F., & Demastes, J. W. (2003). Intentions and beliefs in students' understanding and acceptance of biological evolution. *Journal of Research in Science Teaching, 40*, 510–528. doi:10.1002/tea.10087

Smith, M. U., & Siegel, H. (2016). On the relationship between belief and acceptance of evolution as goals of evolution education. *Science & Education, 25*, 473–496. doi:10.1007/s11191-016-9836-4

Stanhope, L., Ziegler, L., Haque, T., Le, L., Vinces, M., Davis, G. K., . . . Belitsky, J. M. (2017). Development of a Biological Science Quantitative Reasoning Exam (BioSQuaRE). *CBE-Life Sciences Education, 16*, ar66.

Tibell, L. A., & Harms, U. (2017). Biological principles and threshold concepts for understanding natural selection. *Science & Education, 26*, 953–973. doi:10.1007/s11191-017-9935-x

To, C., Tenenbaum, H. R., & Hogh, H. (2017). Secondary school students' reasoning about evolution. *Journal of Research in Science Teaching, 54*, 247–273. doi:10.1002/tea.21347

Wachsmuth, L. P., Runyon, C. R., Drake, J. M., & Dolan, E. L. (2017). Do Biology Students Really Hate Math? Empirical Insights into Undergraduate Life Science Majors' Emotions about Mathematics. *CBE-Life Sciences Education, 16*, ar49.

Wilson, M., de Boeck, P., & Carstensen, C. H. (2008). Explanatory item response models: A brief introduction. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of Competencies in Educational Contexts* (pp. 91–120). Göttingen, Germany: Hogefe & Huber.

Wright, B. D., & Mok, M. (2000). Understanding Rasch measurement: Rasch models overview. *Journal of Applied Measurement, 1*, 83–106.

Wu, M., Adams, R., Wilson, M., & Haldane, S. (2007). *ACER ConQuest 2.0: General item response modelling software [computer program manual]*. Camberwell, Australia: Australian Council for Educational Research.

Zabel, J., & Gropengiesser, H. (2011). Learning progress in evolution theory: climbing a ladder or roaming a landscape? *Journal of Biological Education, 45*, 143–149. doi:10.1080/00219266.2011.586714

6.8 Supplemental Material

6.8.1 Predictors Explaining Variance of Acceptance

Table 6.6 Predictors explaining variance of acceptance of evolution indicated by published empirical studies (references listed in Table 6.8.2)

Reference	Explained variance (%)	Predictor(s) for acceptance of evolution
1	3.3	Knowledge evolution
2	19.8	Knowledge evolution
3	29.0	Knowledge evolution (posttest)
4	32.0	Knowledge evolution
5	24.8	Thinking disposition
6	24.8	Thinking disposition
3	9.0	Religiosity (pretest)
7	36.0	Religiosity
7	12.0	Political orientation
7	33.0	Trust in science and scientists
Authors	15.3	Randomness and probability in evolution
Authors	28.0	Knowledge evolution + randomness and probability in evolution + randomness and probability in mathematics + age(*) + gender + class level(*) + class plan(*) + English reading skills + English writing skills(*) + race/ethnicity
8	17.0	Knowledge evolution + nature of science
6	39.7	Knowledge evolution + religiosity (pretest)
6	17.1	Knowledge evolution + religiosity + nature of science (pretest)
6	44.2	Knowledge evolution + religiosity + nature of science (posttest)
9	23.0	Knowledge evolution + level education(*) + religiosity
1	8.3	Knowledge evolution + thinking disposition
10	40.2	Knowledge evolution + nature of science + religious beliefs + trust in science (German sample)
9	33.0	Knowledge evolution + level education(*) + religion + feeling of certainty
11	45.0	Knowledge evolution + nature of science + STEM influences(*) + influence of participants religious beliefs
12	47.0	Knowledge evolution + religiosity(*) + creationism + science + scientism(*) + gender + track(*) + class level
13	49.0	Knowledge evolution(*) + teleological reasoning(*) + prior educational exposure + religiosity + parental attitude towards evolution + parent's education level(*)
1	10.5	Knowledge evolution + TD + parents' education level
14	13.0	Knowledge evolution(*) + composite disposition + epistemological sophistication(*)
15	9.5	Knowledge evolution + class level(*) + class plan(*) + prior intro course(*) + seen trees before + learn to read trees(*) (pretest)
15	67.7	Knowledge evolution pretest(*) + Knowledge evolution posttest + acceptance pretest + class level + class plan(*) + prior intro course(*) + seen trees before(*) + learn to read trees(*) (posttest)
Authors	19.3	randomness and probability in evolution + randomness and probability in mathematics
6	13.9	Religiosity + nature of science (pretest)
10	19.7	Religious beliefs + trust in science(*) (Turkish sample)
7	37.0	Political orientation + trust in science and scientists
7	49.0	Religiosity + trust in science and scientists
7	50.0	Religiosity + political orientation + trust in science and scientists
5	28.8	Thinking disposition + religiosity
3	28.8	Thinking disposition + students' frequency of religious practices

Note. Authors = referring to values indicated in the original article

6.8.2 References of the Empirical Studies used in Figure 6.2

Table 6.7 References of the empirical studies used in figure 6.2 of the original article

#	References
1	Deniz, H., Donnelly, L. A., & Yilmaz, I. (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*, 420–443. doi:10.1002/tea.20223
2	Brown, J., & Scott, J. A. (2016). Measuring the Acceptance of Evolutionary Theory in Texas 2-Year Colleges. *Research in Higher Education Journal, 31*, 1–11.
3	Athanasiou, K., Katakos, E., & Papadopoulou, P. (2012). Conceptual ecology of evolution acceptance among Greek education students: the contribution of knowledge increase. *Journal of Biological Education, 46*, 234–241. doi:10.1080/00219266.2012.716780
4	Rice, J. W., Clough, M. P., Olson, J. K., Adams, D. C., & Colbert, J. T. (2015). University faculty and their knowledge & acceptance of biological evolution. *Evolution: Education and Outreach, 8*, 1–15. doi:10.1186/s12052-015-0036-5
5	Athanasiou, K., & Papadopoulou, P. (2015). Evolution Theory Teaching and Learning: What Conclusions Can We Get From Comparisons of Teachers' and Students' Conceptual Ecologies in Greece and Turkey? Eurasia Journal of Mathematics, *Science & Technology Education, 11*, 841–853. doi:10.12973/eurasia.2015.1443a
6	Athanasiou, K., & Papadopoulou, P. (2012). Conceptual ecology of the evolution acceptance among Greek education students: Knowledge, religious practices and social influences. *International Journal of Science Education, 34*, 903–924. doi:10.1080/09500693.2011.586072
7	Nadelson, L. S., & Hardy, K. K. (2015). Trust in science and scientists and the acceptance of evolution. *Evolution: Education and Outreach, 8*, 1–9. doi:10.1186/s12052-015-0037-4
8	Akyol, G., Tekkaya, C., Sungur, S., & Traynor, A. (2012). Modeling the interrelationships among pre-service science teachers' understanding and acceptance of evolution, their views on nature of science and self-efficacy beliefs regarding teaching evolution. *Journal of Science Teacher Education, 23*, 937–957. doi:10.1007/s10972-012-9296-x
9	Ha, M., Haury, D. L., & Nehm, R. H. (2012). Feeling of certainty: Uncovering a missing link between knowledge and acceptance of evolution. *Journal of Research in Science Teaching, 49*, 95–121. doi:10.1002/tea.20449
10	Graf, D., & Soran, H. (2011). Einstellung und Wissen von Lehramtsstudierenden zur Evolution – ein Vergleich zwischen Deutschland und der Türkei [Preservice teachers' attitude and knowledge of evolution - a comparison between Germany and Turkey]. In D. Graf (Ed.), *Evolutionstheorie - Akzeptanz und Vermittlung im europäischen Vergleich* (pp. 141–161). Berlin, Heidelberg: Springer.
11	Glaze, A. L., Goldston, M. J., & Dantzler, J. (2015). Evolution in the southeastern USA: Factors influencing acceptance and rejection in pre-service science teachers. *International Journal of Science and Mathematics Education, 13*, 1189–1209. doi:10.1007/s10763-014-9541-1.
12	Großschedl, J., Konnemann, C., & Basel, N. (2014). Pre-service biology teachers' acceptance of evolutionary theory and their preference for its teaching. *Evolution: Education and Outreach, 7*, 1–16. doi:10.1186/s12052-014-0018-z
13	Barnes, M. E., Evans, E. M., Hazel, A., Brownell, S. E., & Nesse, R. M. (2017). Teleological reasoning, not acceptance of evolution, impacts students' ability to learn natural selection. *Evolution: Education and Outreach, 10*, 1–12. doi:10.1186/s12052-017-0070-6
14	Sinatra, G. M., Southerland, S. A., McConaughy, F., & Demastes, J. W. (2003). Intentions and beliefs in students' understanding and acceptance of biological evolution. *Journal of Research in Science Teaching, 40*, 510–528. doi:10.1002/tea.10087
15	Gibson, J. P., & Hoefnagels, M. H. (2015). Correlations between tree thinking and acceptance of evolution in introductory biology students. *Evolution: Education and Outreach, 8*, 1–17. doi:10.1186/s12052-015-0042-7

7 STUDY 4: ITEM CONTEXT AFFECTS THE USE OF THRESHOLD CONCEPTS IN STUDENT EXPLANATION OF EVOLUTION BY NATURAL SELECTION[4]

Abstract

Earlier research on students' evolutionary explanation indicated item feature effects on the use of key concept among different contexts. Until now, there is no empirical research – at least to our knowledge - that examined item feature effects regarding the use of threshold concepts such as randomness, probability, temporal scale, and spatial scale. Thus, the aim of the presented study is to describe how students apply particular threshold concepts in their written explanations of evolutionary processes, and to characterize the relation between item feature and the expression of threshold concepts. A total of 247 university students from Sweden and Germany participated in this study. The findings indicate that students' written explanations of three evolutionary processes differ regarding both key concepts and threshold concepts. Overall, students most often express spatial scales but less often used randomness or probability in their written answers. Moreover, students seldom use a threshold concept across the three items. We discuss whether our findings are relevant to explain the well-documented variation in evolutionary explanations regarding item features. Additionally, we consider possible implications for teaching and learning as well as for assessment of threshold concepts.

Keywords:

Science education, Evolution, Natural selection, Assessment, Conceptual understanding, Threshold concepts, Undergraduates

[4] This is a prior pre-peer reviewed version of the following manuscript: Göransson, A., Orraryd, D., Fiedler, D., & Tibell, L. A. E. (in preparation). Item context affects the use of threshold concepts in student explanation of evolution by natural selection.

7.1 Introduction

Natural selection defined by Darwin 1859 as a major mechanism explaining evolution is central to biology (Dobzhansky, 1973). It is often formulated around three major principles: variation, selection, and inheritance (e.g., Lewontin, 1970; Tibell & Harms, 2017). Some scholars in science education describe these principles as *core concepts* of natural selection: (1) the presence and causes of variation, (2) inheritance of variation, and (3) differential reproduction and/or survival (Nehm & Ha, 2011). Knowing these concepts is often described as necessary and sufficient for explaining natural selection (Nehm & Ha, 2011). However, additional concepts, called *key concepts* (e.g., biotic potential, selection pressure, limited resources, competition, and change in population of distribution/frequency of traits or genes) are also often used to explain natural selection (Nehm & Reilly, 2007). While these concepts are central to explain and to understand natural selection, and thus, important to consider for the evaluation of student explanations, research indicates that learners tend to use other non-causal explanations of evolution such as need-based, teleological or anthropomorphic explanations (Gregory, 2009).

Although the aforementioned core and key concepts have been claimed to be causally central to explain natural selection (Opfer, Nehm, & Ha, 2012), we argue that some of the causal nature of these concepts are overlooked and are potential source for misconceptions (Tibell & Harms, 2017). For instance, differential survival does not mean that all individuals with higher fitness always survive. Rather, it is a probabilistic concept meaning that individuals that are better adapted to their surrounding environment have a higher probability of surviving than others. Abstract concepts such as randomness or spatial scale are also described as *threshold concepts* (Meyer & Land, 2003; Ross et al., 2010; Tibell & Harms, 2017). A threshold concept has been defined as a portal that once passed opens up a new and previously unavailable way of thinking leading to a transformed view of subject matter (Meyer & Land, 2003).

To understand evolution and that evolution applies to all contexts of organisms and traits, learners firstly need to understand that the function of a living organism is ultimately dependent on the genes and that genes in the form of DNA molecules are present in all living cells. But learner's first acquaintance with cells, genes, and DNA often occur rather later in school years, whereas in earlier school years, biology education tends to emphasize more on basic taxonomy and ecological concepts. Thus, learners are likely to lack the unity of life. Moreover, they rather experience different life forms and that taxa have a distinct nature or "essence" (Kalinowski, Leonard, & Andrews, 2010). But an integrated understanding of biology requires the understanding and ability to traverse spatial scales and levels of organization, in order to fully recognize the unifying principles of biology:

- All living organisms are composed of cells with organism being unicellular or multicellular.
- Cells are the basic unit of life.

- Cells only arise from pre-existing cells.
- Energy flow (conversion) occurs within cells.
- Traits are passed on from parent to offspring (from cell to cell) through genes that are located on chromosomes consisting of DNA.
- Changes in genes only occur through mutation and recombination. These processes are random with respect to adaptive value to the surrounding environment.

Therefore, traversing several spatial or organizational levels can be considered to be a threshold to cross in order to fully comprehend this unity of life and that evolution applies to all forms of life. Moreover, learners need to understand that life, although seemingly purposeful and designed, is to a large extent subject to random processes. In fact, research has shown that learners perceive random processes as inefficient (Garvin-Doxas & Klymkowsky, 2008). For instance, focusing on randomness in the origin of novel variation is likely important in order to overcome teleological explanations and misunderstandings such as new traits arise in response to a need somehow "sensed" by the organism. Perkins and Grotzer (2005) argue that learners often have a limited repertoire of causal models used in explanations of scientific phenomena. Learners use mainly simple linear causal models, which can be seen in typical student explanations such as the environment poses a pressure or a need to the organism or population/species. By adapting, the organism responds to this pressure or "need". What is needed in a learners' mind is a more complex causal model with several steps: random occurrence of new traits with the ensuing selection in the form of differential survival and reproduction. Even though learners might come to understand natural selection as this two-step processes (Mayr, 2002), other misunderstandings can occur. One typical misunderstanding is that only the ones with the novel trait survive, essentially leading to a population with no or little variation. Again, learners tend to use a simple deterministic causal model that differential survival means that all "unfit" die and all organisms with a higher fitness caused by a novel trait survive, so that the population changes in one generation. Thus, they conceptualize natural selection as an event and not as an ongoing process where a small variation in a trait gradually accumulates in a population over generations.

Another problem with causality lies in the temporal aspect. The temporal order of casually central events in evolution is often misunderstood. Typically, learners think that environmental factors cause the appearance of novel traits without explicitly mentioning mechanisms for the occurrence of this trait such as mutation or gene duplication. These novel traits are then inherited and enriched in the population due to selection (a typical Lamarckian explanation). Around 70 years ago, the temporal order of mutation and selection was an active field of research, since it was not self-evident that mutations should occur before the application of selection pressures. Only because of the experiments performed by Luria and Delbrück (1943) or the Lederbergs (1952), we know that mutations appear irrespective of their fitness to the environment. If learners understand this temporal order of events, it might be possible to avoide need-based reasoning to some extent.

Temporality is also connected to time scales. Evolutionary processes take place on vastly different time scales, which is often situated outside humans' perceptual range. Many events in evolution such as a specific mutation have a very low probability. However, given the enormous time span and large populations, such events are actually very likely to occur. Thus, time is an important concept for understanding evolution and natural selection. Previous studies have focused mostly on the issue of deep time and concluded that students have difficulties with large time scales (Catley & Novick, 2009). However, deep time needs to be placed in a context of evolutionary mechanisms. It is one thing to be able to place a number of important events in time such as origin of life, nucleated cells, and photosynthesis. But of greater concern is the ability to compare duration of events on different time scales. This is also related to the competency to work with large numbers and reasoning about proportional relationships (Cheek, 2012), and thus, is directly related to reasoning about evolutionary mechanisms and time (e.g., how unlikely events became likely given a large time frame). To reason with time, learners need to translate large time frames into generation numbers and connect these with population numbers, mutation frequency, and so on.

Therefore, to capture these neglected aspects of evolutionary understanding in student explanations, we use the two-dimensional conceptual framework described by Tibell and Harms (2017). In this framework, core concepts and key concepts (origin of variation, individual variation, inherited variation, differential survival, reproductive success, selection pressure, change in population, and limited survival) constitute the first dimension, while threshold concepts (randomness, probability, spatial scale, and temporal scale) represent the second dimension. They propose that complete understanding of evolution requires the development of knowledge concerning both the principles and abstract concepts, and the ability to freely navigate through this two-dimensional framework.

7.1.1 Aim and Research Questions

The aim of the presented study is to explore to what extent learners use threshold concepts in their explanations of evolutionary processes and to characterize the way these are expressed. The following research questions guided our study:

- How do students apply key concepts and threshold concepts in their written explanations of evolution by natural selection?
- What is the relation between item features and expression of threshold concepts?

7.2 Methods

7.2.1 Instrument

Since we were interested in how learners apply threshold concepts in evolutionary explanations, open response items were found to be most suitable, since we were not able to predict the answers in advance (requirement for multiple-choice items). Moreover, threshold concepts are not explicitly addressed in available multiple-choice test instruments (e.g., CINS: Anderson, Fisher, & Norman, 2002; MUM: Nadelson, & Southerland, 2009). Using an instrument with open response items has the advantage that learners generate their own spontaneous application of threshold concepts when explaining evolutionary phenomena. It is also known that open-ended tests provide more robust measures of student's knowledge, because recall of information rather than recognition is required (Opfer et al., 2012). Open response questions ask students to explain a phenomenon rather than just recall information. Students must construct an answer in which specific concepts are applied and integrated into an explanation. Thus, written explanations provided by the students should reflect their understanding.

An available instrument fulfilling the criteria above is the Open Response Instrument (ORI; Bishop & Anderson, 1990; Nehm & Reilly, 2007). The ORI has been found to be more valid than the commonly employed multiple-choice test (Nehm & Schonfeld, 2008). Thus, we used three of the items from the ORI:

- Explain why some bacteria have evolved a resistance to antibiotics (that is, the antibiotics no longer kill the bacteria).
- Cheetahs (large African cats) are able to run faster than 60 miles per hour when chasing prey. How would a biologist explain how the ability to run fast evolved in cheetahs, assuming their ancestors could run only 20 miles per hour?
- Cave salamanders (amphibian animals) are blind (they have eyes that are not functional). How would a biologist explain how blind cave salamanders evolved from ancestors that could see?

These three items are all framed in an evolutionary context and learners are expected to construct answers that explain how the changes occur referring to the process of natural selection. The problems are isomorphic in structure and expected to produce similar explanations. However, they differ in item features such as biological taxa, type of trait, and trait gain/loss. The first item (bacteria) is different from the second (cheetah) and third (salamander) regarding the type of organism (unicellular and prokaryotic organisms versus multicellular and eukaryotic organisms). The second and third item concerns animals that are probably more familiar to learners. In addition, the familiarity with the trait type should be higher for running speed in cheetahs and blindness in salamanders compared to drug resistance in bacteria, which is confined to subcellular components such as changes in proteins and enzymes. It is also worth noting that the cheetah and the salamander item involve evolutionary developmental changes affecting morphological and metabolic features, thus increasing the complexity of scientific explanation.

7.2.2 Participants

Data was collected by administering the test electronically to volunteers from universities in Sweden and Germany.

Swedish sample. The Swedish sample included 38 university students (*M* = 23.7 years, *SD* = 2.25). The students attended various education programs that can be divided into chemistry/biology related programs (*n* = 34 students) and primary teacher education (*n* = 4 students).

German sample. A total of 209 biology students from 21 German universities (*M* = 23.0 years, *SD* = 3.3) participated in our study. On average, they had attended 5.4 semesters (*SD* = 3.4) in tertiary education. A total of 97 students were biology majors, of whom 71 students attended undergraduate and 26 graduate courses. Similar, from the remaining 112 preservice biology teachers, a total of 52 students participated in undergraduate and 42 in graduate courses. Further, 18 preservice biology teachers underwent so called basic foundation courses (first state exam).

In Germany, biology tertiary education (after K-12) generally starts with a 3–4 years undergraduate course (leading to a Bachelor's degree) to equip students with a broad qualification by providing academic subject-specific foundations, methodological skills, and competences related to the professional field (KMK, 2010). Afterwards, a 1–2 years graduate course follows to prepare students with further subject and academic specialization with finally ending in a Master's degree (MKM, 2010). Apart from that, tertiary education for preservice teachers includes at least two focus subjects and can either be taken as a graded course (with Bachelor's and Master's degrees) or as a basic foundation course (concluding with the first state exam; KMK, 2010). In both graded and foundation course subject areas, subject didactics, and educational science components are coupled and supplemented with practical components (KMK, 2010).

7.2.3 Data Analysis

We used content analyses to address our research questions (Krippendorff, 2013). In this study, we are concerned with how learners construct written explanations of three different evolutionary phenomena (antibiotic resistant bacteria, fast running cheetah, and blind cave salamanders). The analysis of written explanations relies on the assumption that learners' explanation of a phenomenon is a representation of their conceptual knowledge.

Coding of variables. We established a coding scheme to score students' evolutionary explanations. The first set of variables *key concepts* covered the X key concepts mentioned in the introduction (i.e., origin of variation, individual variation, inherited variation, differential survival, reproductive success, selection pressure, change in population, and limited survival). In contrast, the second set of variables *threshold concepts* correspond to the four above-mentioned threshold concepts of the two-dimensional framework (i.e., randomness, probability, spatial scale, and temporal scale; Tibell & Harms, 2017). The coding schema was used to quantify the presence (coded as 1) or absence (coded as 0) of the key and threshold concepts. For piloting of the coding schema, two raters independently analyzed an at least 10% overlap of items in both samples (Krippendorff, 2013). Interrater reliability was computed by calculating Guliford's *G* (Holley & Guilford, 1964), which performs more consistently for variables with low presence of a concept (Xu & Lorber, 2014). Variables lacking a satisfactory reliability of $G > 0.7$ were discussed. Variable definitions were refined in the final codebook and students' responses checked by recoding of these variables. The final reliabilities of each variable are available as supplemental material accompanying the online article (see Supplemental Material 7.8.1).

Statistical analysis. The data was exported from MaxQDA version 12 to IBM SPSS statistics software 24 for further analysis. We used the pooled data of all tertiary students from both countries as our main sample. The remaining upper secondary student responses served as a comparison group for exploring education level effects on concept usage. Cochran's *Q* test was used to determine whether there were significant differences in the proportions of participants using a specific concept across the different items (bacteria, cheetah and salamander). This test is used for comparing a dichotomous outcome variable in k related samples such as testing for differences in pass/fail frequency on different test items (Siegel & Castellan, 1988). Cochran's *Q* test was performed for each concept (key and threshold), and Bonferroni correction was used to adjust for the number of comparisons within each sample (12), a=0.05/12=0.004. If a significant effect of item context was found, subsequent pairwise comparisons were performed (Dunn's post-hoc test, non-parametric) using built-in alpha-adjustment for multiple comparisons.

7.3 Results

Our findings indicate that the overall occurrence of threshold concepts varied in students' written explanations (see Figure 7.1). The threshold concepts randomness and probability were least used. Roughly only a third of all students mentioned these concepts at least once across the three items. In contrast, temporal scale had a slightly higher presence, while spatial scale was the most often found concept. Comparing key concepts and threshold concepts, threshold concepts were generally less frequent in students' explanations.

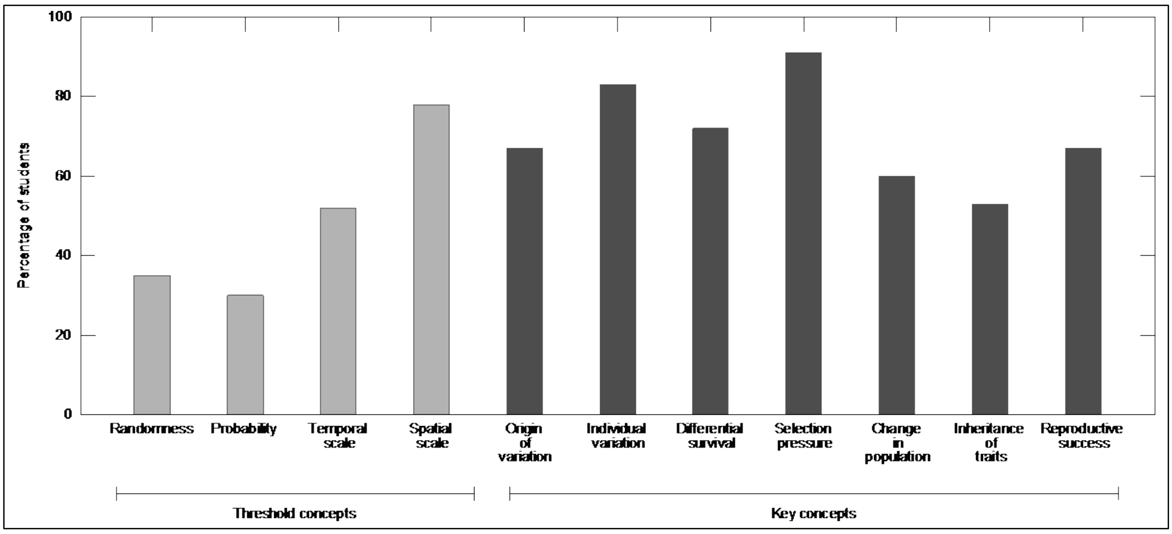


Figure 7.1 Overall frequencies of key and threshold concepts in students' explanations. The graph shows percentage of respondents using a concept at least once across the three items.

The consistency of used concepts across the three items was generally low (see Table 7.1). Only a minority of the students applied a specific concept across all three explanations. Moreover, there was no single item that consistently elicited overall more key or threshold concepts (see Figure 7.2 and 7.3). Rather, each concept varied in frequency depending on the item. However, lexical statistics (cf. (http://countwordsworth.com/sentences) of the explanations showed that the length of the students' answers did not differ significantly in terms of number of words, sentences or sentence length (see Table 7.2). Hence, the inconsistent application of concepts across the items does not appear to be an artefact of answer length.

Table 7.1 Consistency of concept application

	Concept used in at least		
Concepts	**1 item**	**2 items**	**3 items**
Threshold concepts			
Randomness	19%	7,2%	1,9%
Probability	22%	4,1%	0,6%
Temporal scale	29%	14%	3,1%
Spatial scale	38%	28%	3,8%
Key concepts			
Origin of variation	35%	15%	9.7%
Individual variation	25%	26%	20%
Differential survival	31%	24%	9,1%
Selection pressure	19%	26%	35%
Limited survival	22%	26%	35%
Change in population	28%	15%	6,9%
Inheritance of traits	29%	14%	1,9%
Reproductive success	27%	22%	8,2%

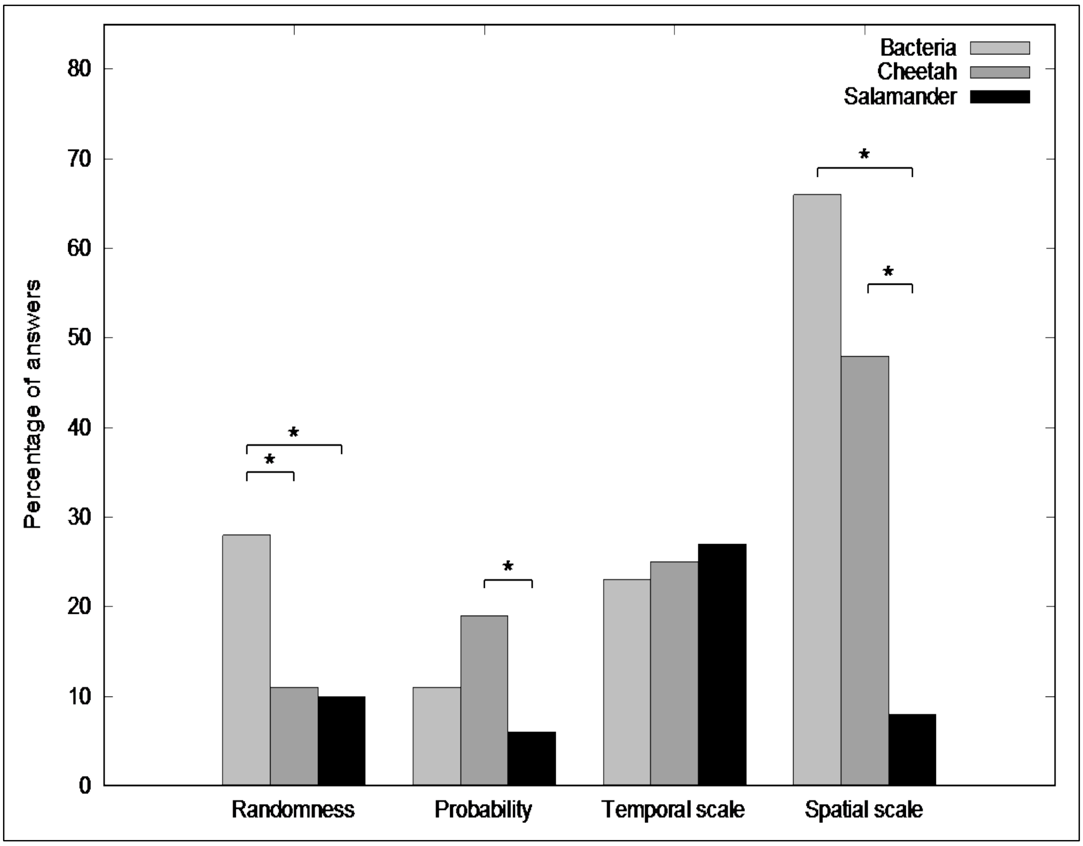


Figure 7.2 Threshold concept frequency across the three items ($N = 247$ students from Sweden and Germany). Asterisk denotes significant differences on Dunn's post hoc test.

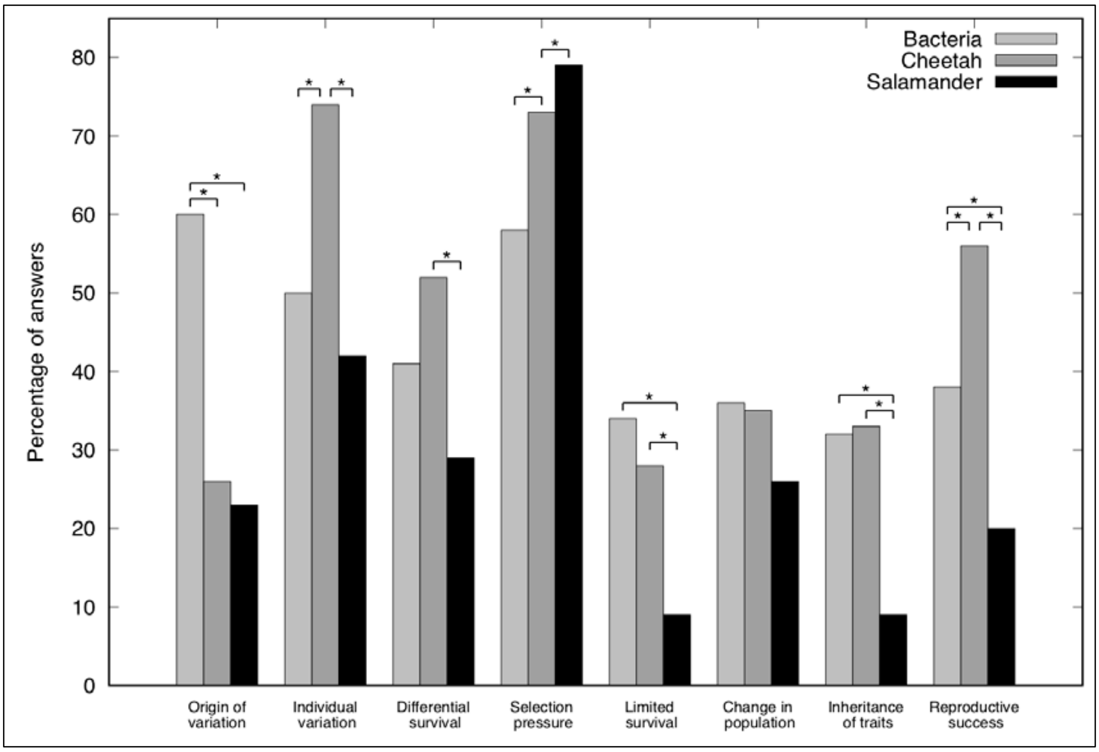


Figure 7.3 Key concept usage across the three items ($N = 247$ students from Sweden and Germany) Asterisk denotes significant differences on Dunn's post hoc test.

Table 7.2 General lexical statistics for collected responses

Item	Average word count per answer	Average sentence count per answer	Average sentence length
Sweden			
Bacteria	25	1.5	17
Cheetah	27	1.6	17
Salamander	23	1.3	18
Germany			
Bacteria	41	2.7	15
Cheetah	50	2.8	18
Salamander	46	2.8	17

7.3.1 Effects of Item Features

To explain the lack of consistency in the use of key and threshold concepts, we analyzed the concept application across the three items (see Figure 7.2 and 7.3). This analysis revealed significant differences in concept usage between the items for all variables except change in population (key concept) and temporal scale (threshold concept). Not only did the concept frequencies vary with the item but there was a variation in which item elicited the highest concept frequency depending on the concept. Lexical statistics for the answers (see Table 7.2) confirmed that there were no large differences in answers lengths in terms of word counts, number of sentences or sentence lengths depending on the item. In addition, these variables did not appear to vary with the item position.

Interestingly, we found a similar pattern of item context effect on threshold concepts (see Figure 7.4), even after data separation according to the two nationalities (i.e., Sweden and Germany). Thus, there could be an item feature effect that is more general and not an artefact of the nationality (e.g., different curricula, textbooks).

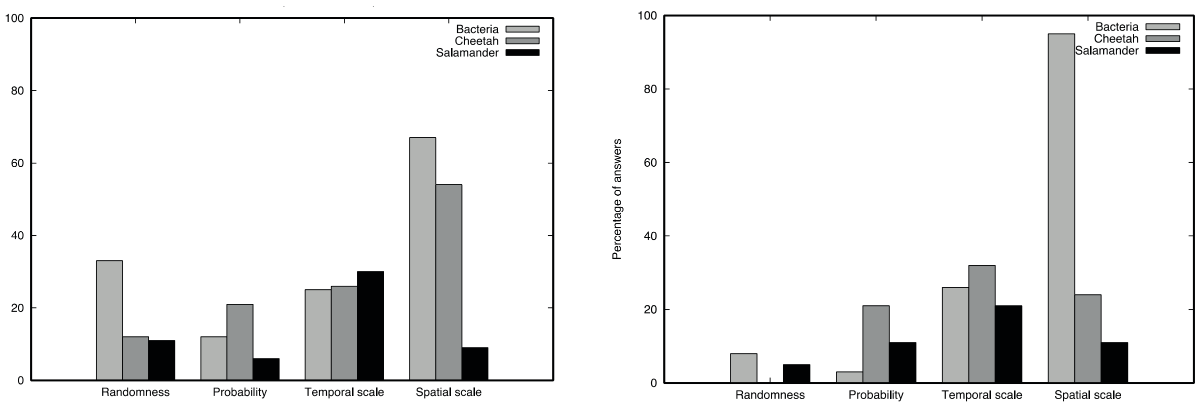


Figure 7.4 Threshold concept frequency across the three items with the German sample left ($n = 209$ students) and the Swedish sample right ($n = 38$ students).

7.3.2 Relation between Key Concepts and Threshold Concepts

While the variation was used in the form of the key concept of individual variation in 40% or more of the answers to the items, randomness was present in substantially lower frequencies. Thus, all answers mentioning variation did not include the random nature of variation. The presence of individual variation was mirrored to a certain extent by the variables differential survival and reproductive success, but this variable was lower than individual variation, indicating that not all students made connections between individual variation and differential fitness in terms of survival and reproduction. Survival and reproduction are not deterministic within a population but probabilistic. Interestingly, the pattern of probability usage in the items mirrored the pattern of individual variation, differential survival, and reproductive success. However, probability occurred in substantially lower frequencies than any of the aforementioned key concepts variables relevant for probability.

Differential survival and reproductive success leads to a change in genotype and phenotype proportions within a population hence change in population. Since differential survival and reproduction takes place over time it is interesting to compare these variables with the threshold concept temporal scale. Temporal scale was present to roughly the same extent within answers to the three items, but less than change in population. This indicates that all students mentioning change in population did not include the aspect that it takes place over time (over generation).Overall, these two concepts were most frequently used in the cheetah item, and least used in the salamander item. Limited survival did not exhibit the same pattern. It was significantly more employed in the bacteria item compared to the salamander item. The cheetah item elicited lower occurrences than in the bacteria item, but still significantly higher than in the salamander item. There was no significant difference between bacteria and cheetah, although the frequency was higher in bacteria than in cheetah. The connections made between the survival/reproduction and environmental factors (biotic and abiotic) were reflected in the variable selection pressure. This was on average the most used concept in the answers, although there were differences across the items. The result of the selection processes (change in population) was mentioned on average in a third of the answers, with no significant difference between the three items, although it occurred at a lower extent in the salamander item.

Inheritance of traits was most frequent and the bacteria and cheetah item, with approximately the same frequencies. The salamander item showed a much lower (significantly) frequency of this concept compared to the former two items. Thus, the tendency to include more of the lower organizational levels in bacteria item measured as spatial scale and origin of variation was not mirrored in terms of inheritance.

7.3.3 Threshold Concepts

While it is known from earlier research that key concept usage tends to vary with item features, the role of item features in relation to threshold concepts has not been described.

Therefore, we performed a more fine-grained analysis of how the found threshold concepts were expressed and in what context.

Randomness. Randomness was most frequent for the bacteria item and significantly different from the other two items. An analysis of the contexts where randomness was found revealed a relation to mutations and appearance of novel traits (see Table 7.3). In the bacteria and cheetah context, randomness was mostly associated with mutations. For the salamander context, randomness was slightly more associated with appearance of novel traits than with mutations. Very few of the answers mentioned randomness in the context of genetic drift or random death, which was in fact completely confined to the salamander item.

Table 7.3 Contexts of randomness usage expressed as percentage of randomness codes ($N = 247$ students)

Randomness codes	Bacteria	Cheetah	Salamander
Random mutation	71%	69%	42%
Random appearance of trait	29%	31%	50%
Random drift	0%	0%	8%
Random death	0%	0%	4%

Note. Percentages do not add up to 100% in all cases since some overlap between codes exists.

Probability. Probability was most frequently expressed in the cheetah item (see Table 7.4). To a large extent probability was mentioned as probability of survival in all items. Probability reasoning in answers to the bacteria item was mostly tied to survival probability and to some extent to occurrence of novel traits. In the cheetah item probability was also frequently mentioned as chance of catching prey, thus leading to increased chance of survival as well as chance of having and providing for offspring.

Table 7.4 Contexts of probability expressed as percentage of probability codes ($N = 247$ students)

Probability codes	Bacteria	Cheetah	Salamander
Survival probability	44%	61%	53%
Chance of catching prey	N/A	35%	6%
Reproduction probability	15%	30%	29%
Probability of novel trait	22%	4%	12%
Mutation probability	22%	0%	0%
Probability of inheritance	11%	11%	12%
Chance of providing for offspring	N/A	4%	0%

Note. Percentages do not add up to 100% in all cases since some overlap between codes exists.

Temporal scale. The most common expression of time was that events occur over time such as "adaptations takes time", but with no specific time frame specified (see Table 7.5). The most frequently used time scale was relative time expressed as generations. Shorter time scales (<days) were in principle completely absent in the

sample, while longer time scales such as years was found in low frequencies , betony in the cheetah and salamander item. There were few mentions of the significance of time scales for accumulation of e.g. several mutations in a population. The frequency of mutations over time (i.e., given enough time beneficial mutations will likely occur) was most frequent in the bacteria item, although for a low frequency of the respondents. A number of the respondents also mentioned time in terms of generations (most frequent in salamander and cheetah item). Interestingly, generation time was mentioned exclusively in the bacteria item.

Table 7.5 Temporal scales expressed as percentage of temporal scale codes ($N = 247$ students).

Temporal scale codes	Bacteria	Cheetah	Salamander
Selection duration	11%	6%	2%
Accumulation of traits	3%	3%	2%
Reproduction rate	15%	11%	5%
Adaptation takes time	32%	50%	48%
Mutations over time	21%	6%	5%
Traits evolve over time	3%	14%	36%
Temporal scale linking			
Generation time affects rate of Evolution	2%	0%	0%
Unspecified time			
Over time	38%	59%	68%
Time frame	9%	3%	2%
Relative time			
Generation time	42%	0%	0%
Generations	14%	30%	32%
Absolute time			
Years	0%	9%	8%
Days	2%	0%	0%
Hours or shorter time	0%	0%	0%

Note. Percentages do not add up to 100% in all cases since some overlap between codes exists.

Spatial scale. Bacteria elicited far more mentions of spatial scale than the other contexts (see Figure 7.2). To explore possible causes for this pattern, we performed a detailed analysis of spatial scale categorized according to different organizational levels (see Table 7.6). While the variable spatial scale used in the initial analysis was used for instances where learners made connections between different organizational levels, the detailed analysis also looked into mentions of objects and processes on specific organizational levels regardless of whether connections were made to other levels. Overall, there was a clear tendency for answers to the bacteria item to include more of the lower organizational levels. Worth noting is the low frequency of gene expression in the salamander item compared to the other two items.

Table 7.6 Spatial scale level categories expressed as percentage (*N* = 247 students).

Scale level	Definition	Subcodes	Bacteria	Cheetah	Salamander
Sub-micro	Molecular/biochemical level. No direct experience possible. Imagination necessary.	Molecule	1%	0%	0%
		Protein	9%	0%	1%
		Gene	26%	27%	7%
		DNA	15%	1%	0%
		DNA-replication	2%	0%	0%
		Gene transfer	22%	N/A	N/A
Micro	Cellular/subcellular level. Visible under light/electron microscope.	Cell	4%	0%	0%
		Cell sio (single individual organism)	10%	N/A	N/A
		Organelle	0%	0%	0%
		Mitosis	6%	0%	0%
Macro	Biological structures visible to the naked eye.	Individual	34%	30%	26%
		Organ	0%	0%	0%
Super-macro	Higher level organization / biological units. Population and above.	Population	22%	15%	18%
		Species	5%	26%	34%
		Higher taxa	40%	1%	0%

Note. Percentages do not add up to 100% in all cases since some overlap between codes exists.

Spatial scale level linkages. We also performed a detailed categorization of the links between different levels of organization/spatial scales found in the answers (see Figure 7.5, 7.6, and 7.7). The pattern of linkages differed between the items. The bacteria and cheetah showed somewhat similar patterns of links with submicro, macro and supermacro levels although the micro level was only found in the bacteria item (see Figure 7.5 and 7.6). For the salamander item another pattern emerged where most links were made between the macro and supermacro levels, while very few links appeared to the sub-micro level (see Figure 7.7).

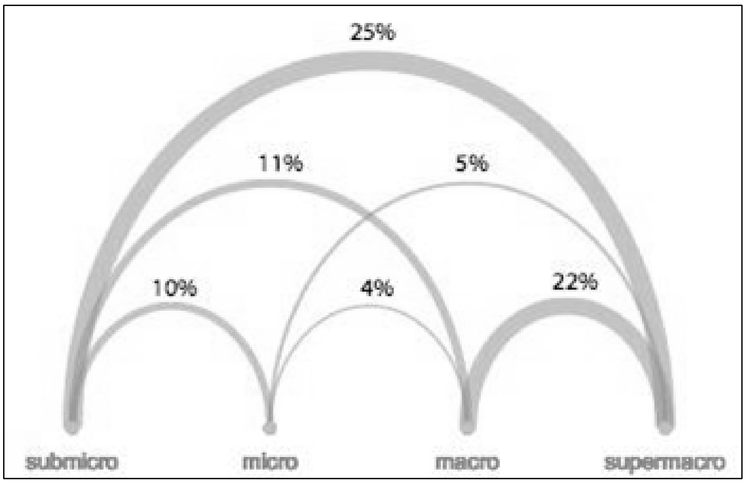


Figure 7.5 Frequency of spatial scale level linkages in bacteria item ($N = 247$ students). Thickness of lines depicts the number of linkages found.

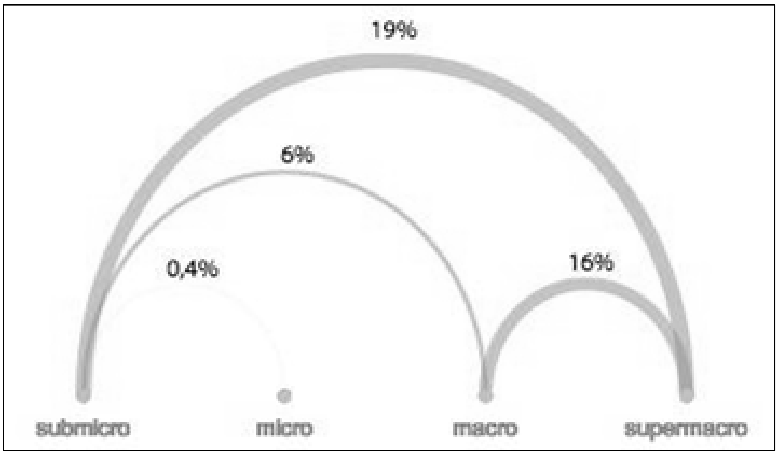


Figure 7.6 Frequency of spatial scale level linkages in cheetah item ($N = 247$ students). Thickness of lines depicts the number of linkages found.

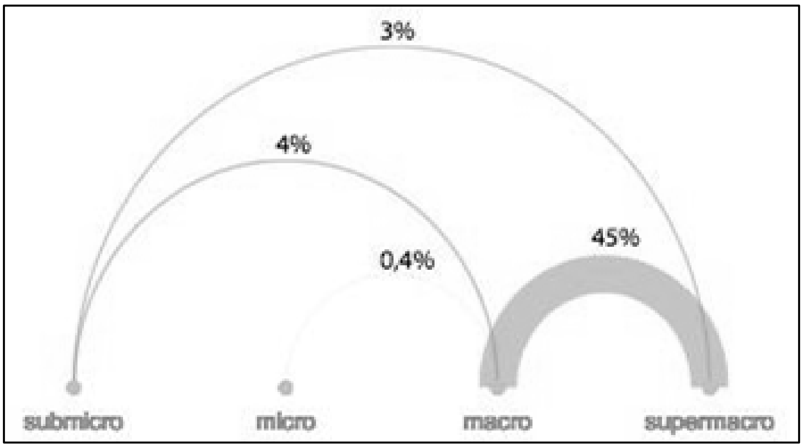


Figure 7.7 Frequency of spatial scale level linkages in salamander item ($N = 247$ students). Thickness of lines depicts the number of linkages found.

7.4 Discussion

Overall, our results indicate that the explanations for the three different evolutionary problems differ for both key concepts and threshold concepts. There was no immediately consistent pattern for all concepts related to item features (i.e., there was no item that consistently elicited fewer or more concepts across all analyzed concepts). Rather, the different items seem to sometimes elicit more or less specific concepts. Thus, if researchers or teachers are interested in learner's knowledge of for example the importance of randomness in evolution, the bacteria item seems to be preferable. In contrast, if one is interested in how well learners can abstract and apply evolutionary reasoning in general several items with different item features should be compared.

The effect of item features on student explanations of natural selection has previously been documented in a number of studies (Heredia, Furtak, & Morrison, 2016; Nehm & Ha, 2011). Heredia et al. (2016) found that animal and plant items had an effect on students' conceptions of natural selection. Animals were higher associated with random origin of traits and variation of individuals. However, previous studies only varied in aspects of trait gain/loss, but did not discriminate between quantitative and qualitative traits. Our results indicate a possible importance of trait type for explanatory patterns learners use in terms of both threshold concepts and key concepts. Therefore, more research is needed to elucidate potential effects of trait type on explanatory patterns, and whether trait type correlates with probabilistic reasoning about evolution by natural selection.

Moreover, the type of variation in the trait can be described as continuous or discontinuous (e.g., Mendelian inheritance: Mendel suggested discontinuous or discrete traits while many traits are continuous and are not controlled by single genes). In contrast, regarding trait loss, it is important to not only understand mutations, but also that a majority of mutations have a negative effect on a trait. In fact, it has been argued that trait loss is more common in evolution than trait gain. For example the occurrence of tetrapod limbs is thought to be a single event while there are numerous example of limbs loss in animal evolution (Johnson, Lahti, & Blumstein, 2012). Thus, it is interesting to note that the trait loss item (salamander item) has been confirmed in several studies to be challenging for students (Nehm & Ha, 2011), and several explanatory patterns were observed for this particular item: 1) energy saving as a reason for evolution is enough alone (teleological explanation), 2) compensatory traits evolve (salamanders that lost eyesight simultaneously gained better smell or hearing), 3) the removal of selection pressure or "need" for eyes (it is also possible that learners use need sometimes in a metaphoric sense) leads to loss of trait (sometimes mentioned to take place through accumulation of mutations). However, compensatory traits do evolve by means of an antagonistic selection. Thus, naïve ideas of compensatory traits do have some similarities to the outcome of an antagonistic selection. Additionally, trait loss can also be explained by relaxed selection: Traits that are energetically expensive to maintain tend be phased out more quickly. Therefore, rapid trait loss is more likely if relatively simple genetic changes are involved.

Nevertheless, our results suggest that item context is important for assessing learner's conceptual knowledge of natural selection, which is in line with previous research that found effects of item context such as trait gain/loss or familiar/unfamiliar taxa (Heredia et al., 2016; Nehm & Ha, 2011; Opfer et al., 2012). The effects of using micro-organisms such as bacteria have been poorly studied in evolution education, although our results indicate that randomness is expressed mostly in the micro-organism context. Moreover, the linkage between different organizational levels or spatial scales was more frequent in the micro-organism context. In unicellular micro-organisms such as bacteria, learners might be less tempted to ascribe evolutionary change to internal changes such as in organs. These changes might be conflated with intentional processes that are not as readily associated with the inner workings of cells. This suggest that micro-organisms could be a fruitful context to introduce in teaching evolution

On a broader scale, our results show that different items elicited different magnitudes of threshold concept such as randomness, probability, temporal and spatial scales. Thus, we suggest that different contexts might have different affordances for teaching and assessing evolution. The salamander item involving trait loss in animals seems to be a poor indicator of novice's conceptual knowledge. However, as a learner gradually acquires a more integrated view of evolution across all domains of biology, it would be meaningful to measure students' ability of abstraction and generalization regarding evolutionary knowledge by comparing answers to items with different features (i.e., familiar/unfamiliar taxa, trait gain/loss, and continuous/discontinuous traits. Learner's expertise should be mirrored in gradually decreasing item context effects. Nevertheless, in biology the devil is in the details because biological explanations tend to be elaborate and specific for each context. Thus, the underlying general principles such as natural selection might not be focused enough if learners are not encouraged to compare and contrast different examples of natural selection in order to appreciate their similarity. In fact, a study of evolution textbooks revealed that a majority of texts only encouraged recall of rote-memory facts, while learners are asked only to a low extent to apply their evolutionary knowledge to new contexts (Aleixandre, 1994).

7.4.1 Threshold Concepts

Randomness was mostly used in the context of mutations. The explanation for this could be related to the tendency to include genetic level explanations such as mutations and the fact the learners might have learned to associate randomness with mutations. This is supported by the fact that spatial scale linkages were most common in the bacteria item. Thus, the tendency to include mutations seems associated with mainly bacteria and hence also randomness. Of the answers to each item mentioning randomness, bacteria and cheetah had a similar percentage connected to mutations (71% and 69%) while randomness was less associated with mutations in the salamander item (39%). In very few instances, randomness was mentioned in other contexts: random death, random genetic drift. Being a causally central concept in evolution, it is a problematic fact that less than 2% of the respondents use randomness in all their answers. It is also troubling that almost none of the upper-secondary students in the sample used randomness in their evolutionary explanations. The reasons behind the low overall usage of randomness remains to be elucidated but several factors are possible. Previous research exemplifies that instructional material such as textbooks (Aleixandre, 1994) and online videos (Bohlin, Göransson, Höst, & Tibell, 2017) fail to adequately address randomness and probability in evolution. We are also aware of examples recently published textbooks for upper-secondary schools that fail to include randomness in origin of variation. Thus, there is a need for further studies on randomness in instructional material such as textbooks and also to what extent teachers address the concept in instruction. It is also known that students tend to reason in deterministic rather than probabilistic ways Differential survival potential as a probability (distribution?) in cheetah but not in bacteria item – Antibiotic resistance could be perceived as a qualitative trait requiring only one mutation rather than a quantitative trait.

> *"Some bacteria are resistant to the drug and thus the non-resistant die. When only the ones surviving antibiotics are left they can duplicate and then only antibiotic resistant bacteria are left."* (Swedish #1)

Quantitative traits such as running speed maybe easier to connect to probability reasoning. Small changes in running speed increase chance of catching prey. Everyday notion of luck in hunting underlying explanation. A majority of the students are likely familiar with cheetahs and can easily conceptualize how running speed is related to survival and reproductive success. As a trait, running speed is familiar from everyday life and we all have direct experience of variation in running speed in the human population. Thus, the students could easily transfer this to the cheetah context. The students probably have a non-deterministic model of hunting, i.e. that you don't always succeed in games such as run and catch (tag game).

> *"[…] a certain cheetah has once upon a time acquired a mutation that caused this individual run faster more easily. Therefore, it had more access to food and thus greater chances of survival and spreading its' genes […]" (Swedish #7)*

The bacteria context is in this regard more challenging for novices since they lack direct experience with how bacteria function. Some of the novices, primarily among the uppers secondary students used alternative explanations from their repertoire of familiar and superficially similar models, such as immunization:

> *"The bacterium has previously been exposed to these drugs and thus have developed resistance. It is exactly like when a human becomes immune to a disease by vaccination. It is a tricky question because bacteria are such small organisms that are not as advanced as a human" (Swedish #97)*

The metaphor *survival of the fittest* could be misleading for learners, especially from a probabilistic viewpoint. Several of the responses included deterministic formulations reminiscent of the survival of the only fit ones, rather than perceiving survival as a phenomenon with stochastic components. This is a threshold to understand how evolution by natural selection behaves in a population that traits might disappear although having positive fitness, especially in small populations.

While none of the items in the study specifically asked the students to address the role of time in evolution and we found no significant effect depending on the item context the detailed analysis revealed some differences between the items. Perhaps the most relevant aspect of time scales, generation time and number of generations differed between the bacteria item and the other two. The fact that generation time only was mentioned in connection with bacteria might be due to perhaps broader known fact that bacteria have short generation time compared to animals and that resistance is something that develops on shorter time scales than running speed and loss of sight. On the other hand, generations were more frequently mentioned in the cheetah and salamander items and this could be an effect of familiarity with animals reproduces in distinct generations. Understanding that evolutionary change only takes place over generations and that the number of generations is important for how fast evolution proceed is central. Generations as temporal scale – for transfer – how many use and discuss this. A small minority mentioned generations in their explanations. This could have several explanations, such as the unsuitability of the items to trigger temporal scale reasoning or that the learners do not consider time as an important factor in the explanations. In addition, some learners are probably aware of the importance of time for evolution but do for some reason not express this explicitly in their answers. Another key or threshold for a learner is to detect that an item should be explained by evolution is that it involves changes over the course of generations, thus overruling other explanations such as developmental or physiological responses to the environment (strategic knowledge). It is worth noting that only two of the items mentions change compared to ancestors, the cheetah item and the salamander item. One result that can be drawn from this is that assessments need additional types of items to correctly assesse learners' understanding of temporal scales in evolution. Moreover, more research is needed that focus on the question how temporal scales are expressed by the teacher and teaching materials such as textbooks and visualizations.

7.4.2 Limitations

Although the age group and education level was limited in our sample, we documented a broad range of different explanations from virtually none to very elaborate and correct explanations. Thus, we are confident that our data represents some of the explanation diversity likely to be found in broader samples. Written explanations are of course limited in information about learners' mental models of natural selection. Although learners cannot be expected to express their entire thinking in a written explanation, this limitation should apply equally to all of the different items, and thus, make them comparable. Given this, our findings of the item features effect on the expression of threshold concepts should have potential value for better understanding learners' difficulties of understanding and explaining evolution by natural selection. An issue with written answers to items that are isomorphic is the item order effect (Federer et al., 2015). However, the effect can be mitigated by items containing different surface features such as taxa and/or trait polarity. In our study, we used only three items varying with respect to item features. Additionally, we saw no correlation between item position and variables such as number of sentences or sentence length. Thus, the frequency of used concepts is not connected to item position. Indeed, some concepts decreased in frequency while others increased for the second or the third item. Therefore, we suggest that our data reflect effects of item features rather than item position. Nevertheless, to fully control for effects of item position, futures studies should use a counterbalanced design (Federer et al., 2015).

Another limitation of the study pertains to the method used to analyze concepts in isolation. Scientific explanation implements structures in which concepts are linked in an appropriate way. Thus, we plan to look into the explanatory structures in future studies.

7.5 Conclusions

Even though it is known from earlier studies that different features afford different explanations (Nehm & Ha, 2011), our results indicate that different features also elicits different frequencies of threshold concepts. Therefore, to measure threshold concept acquisition, the test items need to be considered regarding their item features since they could affect student's expression of threshold concepts. Our study confirms that trait loss seems to be challenging for students to explain, not only in terms of key concepts but also in term of threshold concepts. Similar patterns for threshold concepts usage in both Sweden and Germany may indicate that

Our results are not affected by different education systems but rather reflect some inherent cognitive tendency to employ concepts in specific ways depending on item features. Evolution education should encourage students to compare and contrast different examples of evolution across taxa and trait type. In addition, learners should be trained explicitly how to interpret and solve problems in biology within an evolutionary framework (establish cognitive strategies). Succeeding in this is dependent on metacognitive skills and learning students what constitutes a scientific explanation within the realm of evolutionary biology. Textbooks and teaching should also pay careful attention to probabilistic reasoning and randomness in evolution. The similarities (such as cells and DNA) among all living organisms should be reinforced in connection with evolution.

7.6 Acknowledgements

The study reported here was part of the Swedish-German cooperation project *"EvoVis: Challenging Threshold Concepts in Life Science – enhancing understanding of evolution by visualization"*, funded by the Swedish Research Council. We are very thankful to the students from Sweden and Germany for participating in this study.

7.7 References

Aleixandre, M. P. J. (1994). Teaching evolution and natural selection: a look at textbooks and teachers. *Journal of Research in Science Teaching, 31*, 519–535.

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

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

Bohlin, G., Göransson, A., Höst, G. E., & Tibell, L. A. E. (2017). A conceptual characterization of online videos explaining natural selection. *Science & Education, 26*, 975–999. doi:10.1007/s11191-017-9938-7

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

Cheek, K. A. (2012). Students' understanding of large numbers as a key factor in their understanding of geologic time. *International Journal of Science and Mathematics Education, 10*, 1047–1069. doi:10.1007/s10763-011-9312-1

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher, 35*, 125–129. doi:10.2307/4444260

Federer, M. R., Nehm, R. H., Opfer, J. E., & Pearl, D. (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,* 527–553. doi:10.1007/s11165-014-9435-9

Garvin-Doxas, K., & Klymkowsky, M. W. (2008). Understanding randomness and its impact on student learning: lessons learned from building the Biology Concept Inventory (BCI). *CBE-Life Sciences Education, 7*, 227–233. doi:10.1187/cbe.07-08-0063

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

Heredia, S. C., Furtak, E. M., & Morrison, D. (2016). Exploring the influence of plant and animal item contexts on student response patterns to natural selection multiple choice items. *Evolution: Education and Outreach, 9*, 1–10. doi:10.1186/s12052-016-0061-z

Holley, J. W., & Guilford, J. P. (1964). A note on the G index of agreement. *Educational and Psychological Measurement, 24*, 749–753.

Johnson, N. A., Lahti, D. C., & Blumstein, D. T. (2012). Combating the assumption of evolutionary progress: Lessons from the decay and loss of traits. *Evolution: Education and Outreach, 5*, 1–128. doi:10.1007/s12052-011-0381-y

Kalinowski, S. T., Leonard, M. J., & Andrews, T. M. (2010). Nothing in evolution makes sense except in the light of DNA. *CBE-Life Sciences Education, 9*, 87–97.

KMK (Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany) (2010). Common structural guidelines of the Länder for the accreditation of Bachelor's and Master's study courses. Retrieved from http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2003/2003_10_10-Accredidation-Bachelors-Masters-study-courses.pdf (accessed 03 April 2018)

Krippendorff, K. (2013). *Content analysis : an introduction to its methodology.* London, United Kingdom: SAGE. Retrieved from https://login.e.bibl.liu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cat00115a&AN=lkp.639179&lang=sv&site=eds-live&scope=site (accessed 03 April, 2018)

Lederberg, J., & Lederberg, E. M. (1952). Replica plating and indirect selection of bacterial mutants. *Journal of Bacteriology, 63*, 399–406.

Lewontin, R. C. (1970). The units of selection. Annual Review of Ecology & Systematics, 1, 1–18.

Luria, S. E., & Delbrück, M. (1943). Mutations of bacteria from virus sensitivity to virus resistance. *Genetics, 28*, 491–511.

Mayr, E. (2001). *What evolution is*. New York, NY: Basic books.

Meyer, J. H., & Land, R. (2003). Threshold concepts and troublesome knowledge: linkages to ways of thinking and practising within the disciplines. In C. Rust (Ed.), *Improving Student Learning: Theory and Practice Ten Years On* (pp. 412–424). Oxford, United Kingdom: Oxford Centre for Staff and Learning Development (OCSLD).

Nadelson, L. S., & Southerland, S. A. (2009). Development and preliminary evaluation of the Measure of Understanding of Macroevolution: Introducing the MUM. *The Journal of Experimental Education, 78*, 151–190. doi:10.1080/00220970903292983

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

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

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

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

Novick, L. R., & Catley, K. M. (2012). Assessing students' understanding of macroevolution: Concerns regarding the validity of the MUM. *International*

Journal of Science Education, 34, 2679–2703. doi:10.1080/09500693.2012.727496

Opfer, J. E., Nehm, R. H., & Ha, M. (2012). Cognitive foundations for science assessment design: Knowing what students know about evolution. *Journal of Research in Science Teaching, 49*, 744–777. doi:10.1002/tea.21028

Perkins, D. N., & Grotzer, T. A. (2005). Dimensions of causal understanding: the role of complex causal models in students' understanding of science. *Studies in Science Education, 41*, 117–165. doi:10.1080/03057260508560216

Ross, P. M., Taylor, C. E., Hughes, C., Kofod, M., Whitaker, N., Lutze-Mann, L., & Tzioumis, V. (2010). Threshold concepts: Challenging the way we think, teach and learn in biology. In J. H. F. Meyer, R. Land, & C. Baillie (Eds.), *Threshold Concepts and Transformational Learning* (Vol. 1, pp. 165–177). Rotterdam, The Netherlands: SensPublishers.

Seel, N. M. (2012). Transfer of learning. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 3337–3341). Boston, MA: Springer. doi: 10.1007/978-1-4419-1428-6_166

Siegel, S., & Castellan, N. J., Jr. (1988). *Nonparametric statistics for the behavioral sciences.* New York, NY: McGraw-Hill.

Tibell, L. A., & Harms, U. (2017). Biological principles and threshold concepts for understanding natural selection. *Science & Education, 26*, 953–973. doi:10.1007/s11191-017-9935-x

Xu, S., & Lorber, M. F. (2014). Interrater agreement statistics with skewed data: Evaluation of alternatives to Cohen's kappa. *Journal of Consulting and Clinical Psychology, 82*, 1219–1227. doi:10.1037/a0037489

7.8 Supplemental Material

7.8.1 Complete List of Variables and their Reliabilities

Table 7.7 Complete list of variables and their reliabilities (Guilfors' G)

	Sweden		**Germany**	
Variables	*Agreement*	*Reliability (G)*	*Agreement*	*Reliability(G)*
Bacteria				
Randomness	100%	1,00	92%	0,84
Probability	100%	1,00	90%	0,8
Temporal scale	93%	0,86	88%	0,76
Spatial	98%	0,95	100%	1
Origin of variation	84%	0,68	94%	0,88
Individual variation	95%	0,91	86%	0,72
Differential survival	91%	0,82	86%	0,72
Selection pressure	95%	0,91	92%	0,84
Limited survival	91%	0,82	96%	0,92
Change in population	93%	0,86	92%	0,84
Inheritance of traits	89%	0,77	94%	0,88
Reproductive success	100%	1,00	94%	0,88
Cheetah				
Randomness	100%	1,00	100%	1
Probability	93%	0,86	94%	0,88
Temporal scale	93%	0,86	90%	0,8
Spatial	86%	0,73	82%	0,64
Origin of variation	98%	0,95	88%	0,76
Individual variation	86%	0,73	96%	0,92
Differential survival	86%	0,73	88%	0,76
Selection pressure	91%	0,82	86%	0,72
Limited survival	86%	0,73	96%	0,92
Change in population	98%	0,95	92%	0,84
Inheritance of traits	91%	0,82	92%	0,84
Reproductive success	89%	0,77	90%	0,8
Salamander				
Randomness	98%	0,95	98%	0,96
Probability	98%	0,95	94%	0,88
Temporal scale	98%	0,95	90%	0,8
Spatial	93%	0,86	86%	0,72
Origin of variation	91%	0,82	94%	0,88
Individual variation	89%	0,77	90%	0,8
Differential survival	89%	0,77	86%	0,72
Selection pressure	73%	0,45	100%	1
Limited survival	80%	0,59	90%	0,8
Change in population	100%	1,00	92%	0,84
Inheritance of traits	100%	1,00	100%	1
Reproductive success	100%	1,00	94%	0,88

8 SUMMARIES OF THE CONDUCTED STUDIES

8.1 Study 1 (Chapter 4): University Students' Conceptual Knowledge of Randomness and Probability in the Contexts of Evolution and Mathematics

The first study focused on the developmental process of the two test instruments "Randomness and Probability test in the context of Evolution" (RaProEvo, 21 items), and "Randomness and Probability test in the context of Mathematics" (RaProMath, 33 items) to explore the empirical structure of biology students' conceptual knowledge of randomness and probability, and to investigate its relationship to the conceptual knowledge of evolutionary theory. The developed test instruments (RaProEvo and RaProMath) were administered to 140 German university students (both biology majors and preservice biology teachers), and evidence regarding the test instruments' validity measures (expert ratings and criterion-related validity measures) was collected.

The results indicate that the two test instruments RaProEvo and RaProMath measure separate, but still related, competencies of university students. Furthermore, evidence of the instruments' reliability measures was promising, while experts and criterion-related indications confirmed their validity measures. Both the RaProEvo and RaProMath test also revealed strong internal validity measures.

Even though a detailed analysis implied a non-optimal difficulty of the RaProEvo test for this sample (many items clustered at the low end of the scale, and there was a lack of sufficiently difficult items to distinguish high performers), the Wright map provided indications of informative patterns regarding students' thinking. Most students could satisfactorily answer items concerning the *process of natural selection* (broad probabilistic aspects of the process), while only high performing students answered questions correctly that concentrate on the complex probabilistic backgrounds. Furthermore, items regarding the *origin of variation*, either with general focus or linked to specific sources of variation were distributed across the entire scale. In contrast, items concerning *random phenomena* seem quite challenging for the students in this sample because only 17% of the participants could explain that evolution through natural selection acts upon randomly generated variation correctly. In the mathematical context (RaProMath), most students could easily answer questions concerning *probability as a ratio*, while only high performers correctly answered items regarding the *probability of events*. In contrast, *single event* items were distributed across the scale, but in cases when students were asked about the (un)predictability of single events, they seemed to think about predictability in aggregated terms.

Regarding the relationship of students' conceptual knowledge of randomness and probability and their evolutionary knowledge, results indicate significant moderate correlations. Furthermore, biology students' RaProEvo test scores and evolutionary knowledge were all higher than those of the preservice biology teachers. In contrast, RaProMath test scores did not differ between biology students and preservice biology teachers.

8.2 Study 2 (Chapter 5): EvoSketch: Simple Simulations for Learning Random and Probabilistic Processes in Evolution, and Effects of Instructional Support on Learners' Conceptual Knowledge

The main aim of the second study was to test the effectiveness of EvoSketch simulations for teaching and learning the roles of randomness and probability in evolutionary contexts (i.e., mutation and selection, respectively). Nevertheless, a further aim was to identify the optimal kind of additional instructional support (if any) to use. In total, 267 German secondary school students from nine comprehensive schools (German *Gemeinschaftsschulen*) participated in the study. Students of each class were randomly assigned to four kinds of interventions: text-based, simulation-based (no additional support), simulation-based with interpretative support (worked example), and simulation-based with reflective support (reflective questions). Students' conceptual knowledge of randomness and probability in the context of evolution (RaProEvo) and their evolutionary knowledge (CINS) were captured two weeks before the intervention, directly after the intervention, and roughly eight weeks later. Additionally, time spent on the material, perceived cognitive load (PCL), and self-reported test taking efforts were examined.

Overall, mean posttest scores (RaProEvo and evolutionary knowledge) were lower, but mean follow-up test scores were higher than pretest scores. Further, RaProEvo test scores indicated that learners in the simulation intervention group (but not those in the simulation with additional support groups) acquired significantly more knowledge between the pre- and follow-up tests than text learners. Significant differences were also found between intervention groups in both PCL and time spent on the material. Learners in the simulation group with additional support worked significantly longer on their tasks than learners in the simulation group without additional support. Still, these groups did not differ in PCL. In contrast, students in the text group spent a moderate amount of time on their worksheets but reported a significantly higher PCL than students of the simulation groups without additional support and with reflective support. Further, students of all groups self-reportedly spent more effort in test-taking on the pretest than on the posttest (directly after the intervention), but no significant differences in these respects were detected among the groups.

8.3 Study 3 (Chapter 6): Is Statistical Reasoning Relevant for Evolution Education?

The third study aimed to explore the potential contributions of students' conceptual knowledge of randomness and probability and the relationship between evolutionary knowledge and acceptance of evolution. Further, the empirical structure of students' conceptual knowledge of randomness and probability was investigated with an international sample (similar to Study 1). A total of 538 American undergraduate students (both biology majors and non-science majors) enrolled in an introductory biology course participated in the study. At the end of the course, they were asked to complete an online diagnostic test consisting of instruments that assessed their conceptual knowledge of randomness and probability (RaProEvo and RaProMath), evolutionary knowledge (CANS), and acceptance of evolution (I-SEA).

The results indicate that this study could replicate the findings from the first study regarding the empirical structure of students' conceptual knowledge of randomness and probability, meaning that the RaProEvo and RaProMath tests measure two separate competencies of university students.

Detailed analyses implied a satisfactory difficulty of the RaProEvo and RaProMath test for this sample. Compared to the first study, the joint Wright map revealed comparable patterns regarding students' thinking. Most students could satisfactorily answer items concerning the *process of natural selection* (board probabilistic aspects of the process). Items regarding the *origin of variation* were distributed across the entire scale, although they seem to be a bit more difficult for this sample of students (four items were above the midpoint). Additionally, items concerning *random phenomena* seem quite challenging for the students in this sample. The patterns in the mathematical context (RaProMath) are also similar to the first study. Most students could answer questions concerning *probability as a ratio*, while only high performers answered items regarding the *probability of events* correctly. Again, *single event* items were distributed across the scale.

Correlation analyses revealed significant moderate to strong relationships between the RaProEvo test, evolutionary knowledge and acceptance of evolution. Similar patterns were also found for the RaProMath test, but with slightly lower associations. Multiple regression analyses indicated that the RaProEvo test and the RaProMath test accounted for 27.7% of the explained variance of evolutionary knowledge. Furthermore, the RaProEvo test and the RaProMath test explained 19.3% variance of the acceptance of evolution. Even after the inclusion of demographic variables and knowledge of evolution, the RaProEvo test and the RaProMath test serve as significant predictors of acceptance of evolution.

8.4 Study 4 (Chapter 7): Item Context Affects the Use of Threshold Concepts in Student Explanation of Evolution by Natural Selection

The aim of the fourth study was to describe how students apply key concepts and threshold concepts in their written explanations of evolutionary processes, and to characterize the relation between item surface features and the expression of threshold concepts. A total of 247 university students from Sweden and Germany participated in this study and were introduced to respond to three open response items focusing on evolutionary processes (antibiotic resistant bacteria, fast running cheetah, and blind cave salamanders).

The findings indicate that students' written explanations of the three evolutionary processes differ regarding key concepts and threshold concepts. Overall, the use of key concepts (mentioned at least once in the three items) was generally higher than the use of threshold concepts. Regarding threshold concepts, students most often expressed spatial scales but less often used randomness or probability in their written answers. Moreover, the use of key concepts and threshold concepts across the three items decreased, meaning that students most often used a key or threshold concept at least in one item but seldom across the three items. Only the consistent use of the key concepts *individual variation*, *selection pressure*, and *limited survival* was relatively high with 20% to 35%. Nevertheless, there was no consistent pattern regarding the use of either key or threshold concepts across the three items. Rather, each concept varied in frequency depending on the item. Still, lexical statistics showed that students' answers did not differ in terms of word number, sentences or sentences length, and thus, could not be an artefact of answers' length. Analyses regarding the usage of concepts (key or threshold) revealed significant differences among the items for all concepts except *change in population* (key concept) and *temporal scale* (threshold concept). Interestingly even after separation into the two nationalities (Sweden and Germany), similar patterns emerged.

A more fine-grained analysis of the expressions of threshold concepts indicated that *randomness* was most frequently used in the bacteria item, and often in relation to mutations or the occurrence of novel traits. In contrast, *probability* was most frequently mentioned in the cheetah item, particularly as a description of the *survival probability*. *Temporal scales* were used across all three items, although the expression was most often as *events occur over time*. Specific mentioning of time was mostly indicated by *years* throughout the three items, and *days* in the bacteria item. In contrast, *hours*, *minutes* or *seconds* were totally absent. Again, *spatial scale* was most often mentioned in the bacteria item with higher tendencies of lower organization level indications (submicro or micro level). At least, *spatial scale level linkages* revealed that the bacteria and cheetah item showed somewhat similar patterns of links with submicro, macro, and supermacro levels. Within the salamander item a different pattern emerged with most links made between the macro and supermacro level and very few links to the (sub-)micro level.

9 DISCUSSION AND PERSPECTIVES

The following sections include an overall discussion of the results (Section 9.1) and a summary of the limitations (Section 9.2). Furthermore, implications for teaching and learning evolution (Section 9.3) and implications for future research (Section 9.4) are provided.

9.1 Overall Discussion

The four studies conducted within the framework of this dissertation contribute to the extension of research by exhibiting the relevance of threshold concepts such as randomness and probability for a more in-depth conceptual knowledge of evolutionary theory. After the development of instruments to measure students' conceptual knowledge of randomness and probability (in an evolutionary and a mathematical context), evidence of the relationships between this knowledge, evolutionary knowledge, and acceptance of evolution was collected. Furthermore, information was gathered on how to foster students' conceptual knowledge of randomness and probability by using visualizations, and how item context influences students' use of threshold concepts.

The results of the respective studies are discussed in detail in the respective research article (Chapter 4) and manuscripts (Chapter 5–7). Therefore, this overall discussion should give a more comprehensive and integrated insight by focusing on three key aspects that are relevant for this dissertation: the selection of instruments to assess evolutionary knowledge (Section 9.1.1), the unexpected low posttest scores in Study 2 (Section 9.1.2), and the item effects for the use of the threshold concepts randomness and probability in written evolutionary explanations (Section 9.1.3).

9.1.1 Selection of Instruments to Assess Evolutionary Knowledge

In the conducted studies, evolutionary knowledge was measured using different test instruments with varying answer formats. Whereas open response items were used in Study 1 and Study 4, a multiple-choice approach was used in Study 2 and Study 3. The decisions were based on the actual intention as well as on the design of the conducted studies, which will be explained in detail in the following sections.

Intention-based decisions. The open-ended approach was used for Study 1 and Study 4 to measure evolutionary knowledge and the use of threshold concepts in students' written evolutionary explanations. Since no test instrument was developed before the RaProEvo, an alternative measurement was needed to gather insights into students' use of threshold concepts in an evolutionary context. Using a test instrument with open response items has the advantage to display learners' spontaneous associations of threshold concepts while explaining evolutionary processes.

Design-based decisions. The instrument which was used to measure evolutionary knowledge was exchanged in Study 2 and Study 3. These studies were planned as a repeated measures design study with four intervention groups (Study 2) and as a cross-sectional study with a sample size of more than 500 students. Thus, these studies were planned with a larger sample size than we had in Study 1 and Study 4. Furthermore, in Study 2 student data was measured on three different days. The Conceptual Inventory of Natural Selection (CINS) and the Conceptual Assessment of Natural Selection (CANS) were developed to be used in large classes and to obtain valid and reliable inferences about students' conceptions (D. L. Anderson et al., 2002; Kalinowski et al., 2016). Moreover, the CINS test was already used in repeated measure design studies with two occasion times (e.g., Andrews, Kalinowski, & Leonard, 2011; Lui & Slotta, 2014; Speth et al., 2009). Still, the test was originally designed for undergraduate students, while our sample contained tenth grade school students.

Besides, the literature reveals that both formats generate valid and reliable inferences using different measures of both knowledge and alternative conceptions of natural selection (Nehm & Schonfeld, 2008). Moreover, results of correlation analyses presented at the NARST 2018 indicate that students' conceptual knowledge of randomness and probability (RaProEvo and RaProMath) is significantly related to both open ended and closed responses test instruments (Fiedler, Nehm, Sbeglia, & Harms, 2018). Additionally, the conducted Study 1 and Study 3 displayed significant positive correlations between RaProEvo scores and evolutionary knowledge scores (both open ended and multiple-choice approaches).

9.1.2 Reasons for the Unexpected low Posttest Scores

Results of Study 2 revealed that mean posttest scores (evolutionary knowledge and RaProEvo) were lower, but mean follow-up test scores were higher than pretest scores. Possible reasons for these unexpected patterns are discussed in detail in the related manuscript (Chapter 5) and will be discussed below by focusing on students' test-taking effort and intervention timeframe.

Test-taking effort. Any test scores obtained by takers are likely to be dependent on the effort they expend while taking it (Wise & Kong, 2005). Results of Study 2 revealed that learners of all four intervention groups self-reportedly spent more effort in test-taking on the pretest than on the posttest (directly after the intervention), but no significant differences in these respects were detected among the intervention groups. This could explain why mean posttest scores were lower but mean follow-up test scores were higher than pretest scores. Moreover, students received a rather high amount of new information, which could have deterred learners with low interest and consequently reduced motivation to answer the test items correctly (Amabile, Hill, Hennessey, & Tighe, 1994; Pintrich & Schrauben, 1992). Moreover, this might have introduced a substantial random behavioral response factor in the posttest results (e.g., Meijer, 2003).

Intervention timeframe. German biology teachers have approximately 90 minutes per week together with their tenth grade students. Thus, they have a rather short time period to teach relevant biology topics to their students. Therefore, the timeframe for the intervention of Study 2 was designed for a school lesson of approximately 45 minutes. This might have been too short for the learning of abstract concepts such as randomness and probability and to successfully integrate this knowledge into a broader understanding of evolution through natural selection. In fact, most intervention studies focusing on learning evolution reveal modest to high learning gains with an instruction length of several hours, weeks or months (for an overview see Beardsley, Bloom, & Wise, 2012). Thus, enlarging the intervention timeframe to incorporate interventions on several days or weeks might also increase students' understanding of randomness and probability as well as evolutionary knowledge.

9.1.3 Item Effects for the Use of Randomness and Probability

The results of Study 4 indicated that the explanations for the three evolutionary problems (antibiotic resistant bacteria, fast running cheetah, and blind cave salamanders) differ for both key concepts and threshold concepts. This is in line with previous research that found effects of item context such as trait gain/loss or familiar/unfamiliar taxa for students' use of key concepts (Heredia, Furtak, & Morrison, 2016; Nehm & Ha, 2011; Opfer, Nehm, & Ha, 2012). The results of Study 4 extend this research, since item context also elicited different magnitudes of threshold concept such as randomness, probability, temporal scales, and spatial scales.

Randomness. The term randomness was most frequently expressed in combination with mutations. This could either be explained by the tendency to include genetic level explanations or the fact, that learners have learned to associate randomness with mutations, or both. In fact, spatial scale linkages were most common in the bacteria item indicating a higher tendency to include mutations in the written explanations. Still, the use of randomness in the cheetah item was similar, while only one-third of the learners mentioned randomness in the salamander item. Moreover, only in very few instances randomness was mentioned in other contexts such as *random death* or *random genetic drift*. Particularly problematic is that less than 2% used the concept of randomness throughout all three explanations, although randomness is a crucial key concept in evolution. The reasons behind the low usage across the items remain to be elucidated, but several factors are possible. For instance, previous research determined that instructional material such as textbooks or online videos failed to adequately address randomness and probability in evolution (e.g., Aleixandre, 1994; Bohlin et al., 2017). Moreover, students often tend to reason in deterministic rather than probabilistic ways (e.g., Gregory, 2009; Mead and Scott, 2010).

Probability. Differential survival potential as category for probability expressions, was most frequent in the cheetah and salamander item but not in the bacteria item. Maybe students perceive antibiotic resistance as a qualitative trait requiring only one mutation rather than a quantitative trait (such as running speed). Additionally, quantitative traits might be easier for learners to connect to probability reasoning due to their everyday experiences. Thus, for instance, small changes in running speed increases the chance of catching prey, which is probably connected to some type of luck. Moreover, students are likely familiar with cheetahs and can conceptualize how running speed is related to survival and reproductive success. In fact, learners have direct experiences of the variation in running speed in the human population (e.g. sport games), and thus, students could easily transfer their everyday knowledge to the cheetah context. Similar everyday connections are possible in the salamander item (i.e., blind humans). In contrast, the bacteria context is maybe more challenging for novices since they lack direct experiences with how bacteria function. In addition, the metaphor *survival of the fittest* could be misleading for learners, especially from a probabilistic viewpoint, since it indicates a deterministic worldview (Gregory, 2009). Indeed, several of the responses included deterministic formulations regarding the survival of *only the fit ones*, rather than perceiving survival as a phenomenon with stochastic components.

Overall, different contexts might not only trigger different threshold concepts, but also need different affordances for teaching evolution and assessing students' evolutionary knowledge. Thus, the salamander item (addressing trait loss in animals) seems to be a poor indicator for novices' conceptual knowledge of evolutionary processes (it elicits fewer key concepts and threshold concepts). However, as a learner gradually acquires a more integrated view of evolution across all domains of biology, it would be meaningful to measure students' ability of abstraction and generalization of evolutionary knowledge by comparing answers to items with different contexts (e.g., familiar/unfamiliar taxa, trait gain/loss, and continuous/discontinuous traits).

9.2 Limitations of the Conducted Studies

The presented research article (Chapter 4) and manuscripts (Chapter 5–7) of the conducted studies list different limitations and concerns. This section gives an overview of the most crucial limitations relevant for this dissertation.

Measuring conceptual knowledge of randomness and probability. The conducted studies assess conceptual knowledge of randomness and probability in the context of evolution by focusing on the process of natural selection and the random and probabilistic processes inside this framework, while excluding other specific random processes of evolution such as genetic drift. This decision was based on the literature that indicates that students often tend to misunderstand the process of natural selection itself and the importance of random processes (see Section 2.1 and Section 2.2). In contrast, the mathematical test instrument focused on random and probabilistic topics that could also be transferred into an evolutionary context such as single events or probability as a ratio. Moreover, the RaProMath also implements faulty statistical reasoning known from the literature (e.g., Garfield, 2003; Kahneman, Slovic, & Tversky, 1982). Still, the developed instruments were neither intended to be summative evaluation tools nor to assess every aspect of randomness and probability exhaustively.

Cross-sectional design. Cross-sectional designs do not allow to make causal conclusions (Döring & Bortz, 2016). However, due to the limited time frame of this dissertation project, the presented research was mainly done with cross-sectional designs. A repeated measures design was only used in Study 2. Thus, this concern is primarily related to Study 1, 3 and 4. All these studies solely enable insights into a specific moment and how the variables are connected to each other. But based on the results of the studies, a statement about the variables' order in terms of influence cannot be given.

Limitations regarding time and EvoSketch simulations. A particular limitation of Study 2 is that the developed visualization is an abstract version of the random and probabilistic process of evolution by natural selection. While simulations often implement a diversity of environments and influencing factors, the program EvoSketch was intended to be non-context-specific. In turn, this might have resulted in difficulties in understanding the simulations and thus influenced the learning with the simulations. Moreover, the simulations could have been too complex for learners with minor spatial ability (Wang et al., 2017).

Generality of the findings. The generality of the findings obtained in this dissertation has to be replicated by other studies on national and international level. Although Study 3 could replicate the initial findings of Study 1, more studies with diverse samples of different universities or nationalities are still required to extend the insights into the effectiveness of the test instruments and the relationships to related variables such as knowledge and acceptance of evolution. Additionally, the participating students of the second study were limited to one German federal state (Schleswig-Holstein) and one form of educational program (comprehensive schools). In contrast, the results of Study 4 could indicate similar patterns across two nations (Sweden and Germany). Nevertheless, comparison groups from other nations are also needed to obtain a generality of item effects on the use of threshold concepts.

9.3 Implications for Teaching and Learning Evolution

This dissertation encompasses explorative studies to provide first insights into students' conceptual knowledge of randomness and probability and the relationship to the knowledge of evolutionary theory. The following general recommendations can be drawn from the conducted research and earlier works:

In line with the science standards and other statements (e.g., Leopoldina, 2016; NGSS, 2013; NRC, 2012), the overall recommendation is that evolution should be taught as a unifying theme throughout the biological education. In fact, former research reveals that students are capable of understanding evolutionary concepts when introduced to the topics in early stages of their education (e.g., Horwitz, McIntyre, Lord, O'Dwyer, & Staudt, 2013).

In addition, teachers should focus on underlying abstract concepts (threshold concepts) such as randomness and probability as well as spatial and temporal scales, since these concepts are essential for students to critically address processes and findings associated with their environment and everyday life. Combining the principles of evolution and their connected underlying threshold concepts can provide students with a deeper understanding of evolutionary processes and foster their awareness regarding the importance of these concepts (Tibell & Harms, 2017). Furthermore, this might also lead to higher acceptance of evolutionary theory.

Based on students' everyday experiences, they develop and maintain a variety of alternative conceptions about evolutionary theory, randomness, and probability (see Study 1, 3 and 4), that can be identified by using the RaProEvo test instrument. These alternative conceptions should then be picked up and be openly addressed by the teacher or instructor to cause cognitive conflicts; so conceptual change is likely to occur (see Section 2.1.1). Particular visualizations such as EvoSketch simulations can foster students understanding of the importance and nature of random processes in the context of evolution (see Study 2).

At least, whenever teachers are using the term randomness or random in their instruction, they should define the meaning of this term in the actual context. Research has shown that some problems regarding the understanding of evolution are based on linguistic features (Nehm et al., 2010; Pramling, 2008; Rector et al., 2013). So, instead of skipping the word (as advised from Ben-Ari, 2004), teachers are better urged to be explicit about how the term is used in the given context (Mead & Scott, 2010).

9.4 Implications for Future Research

In the first study, two test instruments were developed to assess students' conceptual knowledge of randomness and probability in the context of evolution and mathematics. Although results indicate that the instruments generate reliable and valid inferences, the tests still require further verification of students' thinking regarding the particular items used in these test instruments. Future studies should focus on interview data or think-aloud protocols with both secondary school students and university students.

Regarding the results of the second study, there are many opportunities for future experimental research. At first, the presented study was only conducted with secondary school students of comprehensive schools. However, results may differ when focusing on students for grammar schools (German *Gymnasien*). Additionally, the instructional support only focused on two specific kinds of additional instructional support. Particularly the effects of feedback were not investigated in this study but might increase students' understanding of the simulated abstract concepts. Furthermore, other factors related to the working with EvoSketch should be checked for influencing characteristics like abstract reasoning ability. Research indicates that spatial ability could influence learning outcomes (K.-E. Chang, Chen, Lin, & Sung, 2008). In addition, eye-tracking or think-aloud approaches may help capturing cognitive processes that are involved in learning with EvoSketch.

Regarding the promising results of Study 3, future studies should also examine the learning gains concerning conceptual knowledge of randomness and probability in introductory courses focusing on (a) statistics for biology students, and (b) evolutionary theory, in order to better understand how improving the conceptual knowledge of randomness and probability might improve the knowledge of evolutionary theory (and acceptance). Furthermore, an opportunity could be to extend the research regarding students' use of threshold concepts in their open response answers compared to other nations such as the United States of America, which might increase the generality of the findings in Study 4.

REFERENCES

Note: The references of the research article (Study 1) and the manuscripts (Study 2 to Study 4) are listed at the end of Chapter 4, 5, 6, and 7, respectively.

Adadan, E., Irving, K. E., & Trundle, K. C. (2009). Impacts of multi-representational instruction on high school students' conceptual understandings of the particulate nature of matter. *International Journal of Science Education, 31*, 1743–1775. doi:0.1080/09500690802178628

Ainsworth, S., & VanLabeke, N. (2004). Multiple forms of dynamic representation. *Learning and Instruction, 14*, 241–255. doi:10.1016/j.learninstruc.2004.06.002

Akyol, G., Tekkaya, C., Sungur, S., & Traynor, A. (2012). Modeling the interrelationships among pre-service science teachers' understanding and acceptance of evolution, their views on nature of science and self-efficacy beliefs regarding teaching evolution. *Journal of Science Teacher Education, 23*, 937–957. doi:10.1007/s10972-012-9296-x

Aleixandre, M. P. J. (1994). Teaching evolution and natural selection: a look at textbooks and teachers. *Journal of Research in Science Teaching, 31*, 519–535. doi: 10.1002/tea.3660310507

American Association for the Advancement of Science (2006). *Evolution on the front line: An abbreviated guide for teaching evolution (Project 2061).* Received from http://www.project2061.org/publications/guides/evolution.pdf (accessed 03 April 2018)

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

Anderson, J. R. (2007). *Kognitive Psychologie [Cognitive psychology]*. Heidelberg, Germany: Spektrum-Verlag.

Andrews, T. M., Kalinowski, S. T., & Leonard, M. J. (2011). "Are humans evolving?" A classroom discussion to change student misconceptions regarding natural selection. *Evolution: Education and Outreach, 4*, 456–466. doi:10.1007/s12052-011-0343-4

Andrews, T. M., Price, R., Mead, L., McElhinny, T., Thanukos, A., Perez, K., . . . Lemons, P. (2012). Biology undergraduates' misconceptions about genetic drift. *CBE-Life Sciences Education, 11*, 248–259. doi:10.1187/cbe.11-12-0107

Athanasiou, K., & Papadopoulou, P. (2012). Conceptual ecology of the evolution acceptance among Greek education students: Knowledge, religious practices and social influences. *International Journal of Science Education, 34*, 903–924. doi:10.1080/09500693.2011.586072

Baalmann, W., Frerichs, V., Weitzel, H., Gropengießer, H., & Kattmann, U. (2004). Schülervorstellungen zu Prozessen der Anpassung–Ergebnisse einer Interviewstudie im Rahmen der Didaktischen Rekonstruktion. [Students conceptions of adaptation processes - Results of an interview study within the framework of didactic reconstruction]. *Zeitschrift für Didaktik der Naturwissenschaften, 10*, 7–28.

Barradell, S. (2013). The identification of threshold concepts: a review of theoretical complexities and methodological challenges. *Higher Education, 65*, 265–276. doi:10.1007/s10734-012-9542-3

Basey, J. M., Maines, A. P., Francis, C. D., Melbourne, B., Wise, S. B., Safran, R. J., & Johnson, P. T. (2014). Impact of pre-lab learning activities, a post-lab written report, and content reduction on evolution-based learning in an undergraduate plant biodiversity lab. *Evolution: Education and Outreach, 7*, 1–9. doi:10.1186/s12052-014-0010-7

Batanero, C., Green, D. R., & Serrano, L. R. (1998). Randomness, its meanings and educational implications. *International Journal of Mathematical Education in Science and Technology, 29*, 113–123. doi:10.1080/0020739980290111

Batzli, J. M., Knight, J. K., Hartley, L. M., Maskiewicz, A. C., & Desy, E. A. (2016). Crossing the threshold: Bringing biological variation to the foreground. *CBE-Life Sciences Education, 15*, es9. doi:10.1187/cbe.15-10-0221

Beardsley, P. M., Bloom, M. V., & Wise, S. B. (2012). Challenges and opportunities for teaching and designing effective K-12 evolution curricula. In K. S. Rosengren, S. K. Brem, E. M. Evans, & G. M. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 287–310). New York, NY: Oxford University Press.

Beggrow, E. P., & Nehm, R. H. (2012). Students' mental models of evolutionary causation: natural selection and genetic drift. *Evolution: Education and Outreach, 5*, 429–444. doi:10.1007/s12052-012-0432-z

Ben-Ari, M. (2004). On random numbers and design. *Science & Education, 13*, 235–241. doi:10.1023/B:SCED.0000025576.56763.fe

Bennett, D. J. (1998). *Randomness*. Cambrigde, United Kingdom: Harvard University Press.

Berkman, M. B., & Plutzer, E. (2011). Defeating creationism in the courtroom, but not in the classroom. *Science, 331*, 404–405. doi:10.1126/science.1198902

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

Bohlin, G., Göransson, A., Höst, G. E., & Tibell, L. A. (2017). A conceptual characterization of online videos explaining natural selection. *Science & Education, 26*, 975–999. doi:10.1007/s11191-017-9938-7

Booth, J. (2006). On the mastery of philosophical concepts: socratic discourse and the unexpected 'affect'. In J. H. Meyer & R. Land (Eds.), *Overcoming barriers to* *student understanding: Threshold concepts and troublesome knowledge* (pp. 173–181). Abington, United Kingdom: Routledge.

Braun, E., Gusy, B., Leidner, B., & Hannover, B. (2008). Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). [The Berlin evaluation instrument for selfapproximated, student competence]. *Diagnostica, 54*, 30–42. doi:10.1026/0012-1924.54.1.30

Brumby, M. (1979). Problems in learning the concept of natural selection. *Journal of Biological Education, 13*, 119–122. doi:10.1080/00219266.1979.9654240

Büchter, A., Hußmann, S., Leuders, T., & Prediger, S. (2005). Den Zufall im Griff? – Stochastische Vorstellungen fördern. *Praxis der Mathematik in der Schule, 47*, 1–7.

Buiatti, M., & Longo, G. (2013). Randomness and multilevel interactions in biology. *Theory in Biosciences, 132*, 139–158. doi:10.1007/s12064-013-0179-2

Butcher, K. R. (2014). The Multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 174–205). New York, NY: Cambridge University Press.

Cai, L., Friedman, N., & Xie, X. S. (2006). Stochastic protein expression in individual cells at the single molecule level. *Nature, 440*, 358–362. doi:10.1038/nature04599

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293–332. doi:10.1207/s1532690xci0804_2

Chang, H.-Y. (2017). How to augment the learning impact of computer simulations? The designs and effects of interactivity and scaffolding. *Interactive Learning Environments, 25*, 1083–1097. doi:10.1080/10494820.2016.1250222

Chang, H.-Y., & Linn, M. C. (2013). Scaffolding learning from molecular visualizations. *Journal of Research in Science Teaching, 50*, 858–886. doi:10.1002/tea.21089

Chang, K.-E., Chen, Y.-L., Lin, H.-Y., & Sung, Y.-T. (2008). Effects of learning support in simulation-based physics learning. *Computers & Education, 51*, 1486–1498. doi:10.1016/j.compedu.2008.01.007

Cheek, K. A. (2013). Exploring the relationship between students' understanding of conventional time and deep (geologic) time. *International Journal of Science Education, 35*, 1925–1945. doi:10.1080/09500693.2011.587032

Chinn, C. A., & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching, 35*, 623–654. doi:10.1002/(Sici)1098-2736(199808)35:6<623::Aid-Tea3>3.3.Co;2-K

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*, 149–210. doi:10.1007/BF01320076

Cohen, L. J. (1995). *Essay on belief and acceptance*. Oxford, United Kingdom: Oxford University Press.

Coley, J. D., & Tanner, K. D. (2012). Common origins of diverse misconceptions: cognitive principles and the development of biology thinking. *CBE-Life Sciences Education, 11*, 209–215. doi:10.1187/cbe.12-06-0074

Coley, J. D., & Tanner, K. D. (2015). Relations between intuitive biological thinking and biological misconceptions in biology majors and nonmajors. *CBE-Life Sciences Education, 14*, ar8. doi:10.1187/cbe.14-06-0094

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684. doi:10.1016/S0022-5371(72)80001-X

Davies, P. (2006). Threshold concepts: How can we recognize them? In J. H. F. Meyer & R. Land (Eds.), *Overcoming barriers to student understanding: Threshold concepts and troublesome knowledge* (pp. 70–84). Abington, United Kingdom: Routledge.

Davies, P., & Mangan, J. (2007). Threshold concepts and the integration of understanding in economics. *Studies in Higher Education, 32*, 711–726. doi:10.1080/03075070701685148

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science, 38*, 105–134. doi:10.1007/s11251-009-9110-0

de Jong, T., Martin, E., Zamarro, J.-M., Esquembre, F., Swaak, J., & van Joolingen, W. R. (1999). The integration of computer simulation and learning support: An example from the physics domain of collisions. *Journal of Research in Science Teaching, 36*, 597–615. doi:10.1002/(sici)1098-2736(199905)36:5<597::aid-tea6>3.0.co;2-6

de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*, 179–201. doi:10.3102/00346543068002179

Deadman, J., & Kelly, P. (1978). What do secondary school boys understand about evolution and heredity before they are taught the topics? *Journal of Biological Education, 12*, 7–15. doi:10.1080/00219266.1978.9654169

Demastes, S. S., Settlage, J., & Good, R. (1995). Students' conceptions of natural selection and its role in evolution: Cases of replication and comparison. *Journal of Research in Science Teaching, 32*, 535–550. doi:10.1002/tea.3660320509

Deniz, H., Donnelly, L. A., & Yilmaz, I. (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*, 420–443. doi:10.1002/tea.20223

Dole, J. A., & Niederhauser, D. S. (1990). Students' level of commitment to their naive conceptions and their conceptual change learning from texts. *National Reading Conference Yearbook, 39*, 303–310.

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in the Sozial- und Humanwissenschaften [Research methods and evaluation in the Social and Human Sciences]*. Berlin Heidelberg, Germany: Springer-Verlag.

Dunk, R. D., Petto, A. J., Wiles, J. R., & Campbell, B. C. (2017). A multifactorial analysis of acceptance of evolution. *Evolution: Education and Outreach, 10*, 1–8. doi:10.1186/s12052-017-0068-0

Eble, G. J. (1999). On the dual nature of chance in evolutionary biology and paleobiology. *Paleobiology, 25*, 75–87. doi:10.1666/0094-8373(1999)025<0075:OTDNOC>2.3.CO;2

Eckhardt, M., Urhahne, D., Conrad, O., & Harms, U. (2013). How effective is instructional support for learning with computer simulations? *Instructional Science, 41*, 105–124. doi:10.1007/s11251-012-9220-y

Eterovic, A., & Santos, C. M. D. (2013). Teaching the role of mutation in evolution by means of a board game. *Evolution: Education and Outreach, 6*, 1–10. doi:10.1186/1936-6434-6-22

Euler, M., Schüttler, T., & Hausamann, D. (2015). Schülerlabore: Lernen durch Forschen und Entwickeln [Student laboratories: Learning through research and development]. In E. Kircher, R. Girwidz, & P. Häußler (Eds.), *Physikdidaktik* (pp. 759–782). Berlin, Heidelberg, Germany: Springer Spektrum.

Evans, E. M. (2000). The emergence of beliefs about the origins of species in school-age children. *Merrill-Palmer Quarterly, 46*, 221–254.

Evans, E. M. (2001). Cognitive and contextual factors in the emergence of diverse belief systems: Creation versus evolution. *Cognitive Psychology, 42*, 217–266. doi:10.1006/cogp.2001.0749

Evans, E. M., Rosengren, K. S., Lane, J. D., & Price, K. L. (2012). Encountering counterintuitive ideas: Constructing a developmental learning progression for evolution understanding. In K. S. Rosengren, S. K. Brem, E. M. Evans, & G. M. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 174–199). New York, NY: Oxford University Press.

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review, 104*, 301–318. doi:10.1037/0033-295x.104.2.301

Feller, W. (1968). *An introduction to probability theory and its applications* (Vol. 1). London, United Kingdom: John Wiley & Sons.

Fenner, A. (2013). *Schülervorstellungen zur Evolutionstheorie: Konzeption und Evaluation von Unterricht zur Anpassung durch Selektion [Students conceptions about the theory of evolution: Design and evaluation of a teaching sequence about adaption through selection]* (doctoral dissertation). Justus-Liebig-Universität, Gießen, Germany.

Ferrari, M., & Chi, M. T. (1998). The nature of naive explanations of natural selection. *International Journal of Science Education, 20*, 1231–1256. doi:10.1080/0950069980201005

Ferreira, A., Lemmer, M., & Gunstone, R. (2017). Alternative conceptions: turning adversity into advantage. *Research in Science Education*, 1–22. doi:10.1007/s11165-017-9638-y

Fiedler, D., Nehm, R. H., Sbeglia, G. C., & Harms, U. (2018). The Role of Statistical Thinking in Learning, Understanding, and Accepting Evolution. *Paper presented at the National Association for Research in Science Teaching (NARST) conference, Atlanta, United States of America, March 10–March 13.*

Fosnot, C. T., & Perry, R. S. (2005). Constructivism: A psychological theory of learning. In C. T. Fosnot (Ed.), *Constructivism: Theory, perspectives, and practice* (2nd ed., pp. 8–33). New York, NY: Teachers College Press.

Franco, C., & Colinvaux, D. (2000). Grasping mental models. In J. K. Gilbert & C. J. Boulter (Eds.), *Developing models in science education* (pp. 93–118). Boston, MA: Springer.

Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal, 2*, 22–38.

Garvin-Doxas, K., & Klymkowsky, M. W. (2008). Understanding randomness and its impact on student learning: lessons learned from building the Biology Concept Inventory (BCI). *CBE-Life Sciences Education, 7*, 227–233. doi:10.1187/cbe.07-08-0063

Gibson, J. P., & Hoefnagels, M. H. (2015). Correlations between tree thinking and acceptance of evolution in introductory biology students. *Evolution: Education and Outreach, 8*, 1–17. doi:10.1186/s12052-015-0042-7

Gilbert, J. K. (2005). Visualization: A metacognitive skill in science and science education. In J. K. Gilbert (Ed.), *Visualization in science education* (pp. 9–27). Dordrecht, The Netherlands: Springer.

Gobert, J. D. (2000). A typology of causal models for plate tectonics: Inferential power and barriers to understanding. *International Journal of Science Education, 22*, 937–977. doi:10.1080/095006900416857

Gobert, J. D. (2005). Leveraging technology and cognitive tehory on visualization to promote students' science. In J. K. Gilbert (Ed.), *Visualization in science education* (pp. 73–90). Dordrecht, The Netherlands: Springer.

Godfrey-Smith, P. (2007). Conditions for evolution by natural selection. *The Journal of Philosophy, 104*, 489–516.

Graf, D., & Hamdorf, E. (2011). Evolution: Verbreitete Fehlvorstellungen zu einem zentralen Thema [Evolution: Common misconceptions about a central topic]. In D. Dreesmann, D. Graf, & K. Witte (Eds.), *Evolutionsbiologie*. Heidelberg, Germany: Spektrum Akademischer Verlag.

Graf, D., & Soran, H. (2011). Einstellung und Wissen von Lehramtsstudierenden zur Evolution – ein Vergleich zwischen Deutschland und der Türkei [Preservice teachers' attitude and knowledge of evolution - a comparison between Germany and Turkey]. In D. Graf (Ed.), *Evolutionstheorie - Akzeptanz und Vermittlung im europäischen Vergleich* (pp. 141–161). Berlin, Heidelberg: Springer.

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

Großschedl, J., Konnemann, C., & Basel, N. (2014). Pre-service biology teachers' acceptance of evolutionary theory and their preference for its teaching. *Evolution: Education and Outreach, 7*, 1–16. doi:10.1186/s12052-014-0018-z

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

Heams, T. (2014). Randomness in biology. *Mathematical Structures in Computer Science, 24*, e240308. doi:10.1017/S096012951200076X

Heddy, B. C., & Sinatra, G. M. (2013). Transforming misconceptions: Using transformative experience to promote positive affect and conceptual change in students learning about biological evolution. *Science Education, 97*, 723–744. doi:10.1002/sce.21072

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1084–1102.

Hegarty, M. (2004). Dynamic visualizations and learning: getting to the difficult questions. *Learning and Instruction, 14*, 343–351. doi:10.1016/j.learninstruc.2004.06.007

Heredia, S. C., Furtak, E. M., & Morrison, D. (2016). Exploring the influence of plant and animal item contexts on student response patterns to natural selection multiple choice items. *Evolution: Education and Outreach, 9*, 10. doi:10.1186/s12052-016-0061-z

Hester, S., Buxner, S., Elfring, L., & Nagy, L. (2014). Integrating quantitative thinking into an introductory biology course improves students' mathematical reasoning in biological contexts. *CBE-Life Sciences Education, 13*, 54–64. doi:10.1187/cbe.13-07-0129

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction, 17*, 722–738. doi:10.1016/j.learninstruc.2007.09.013

Horwitz, P., McIntyre, C. A., Lord, T. L., O'Dwyer, L. M., & Staudt, C. (2013). Teaching 'evolution readiness' to fourth graders. *Evolution: Education and Outreach, 6*, 1–12. doi:10.1186/1936-6434-6-21

Johannsen, M., & Krüger, D. (2005). Schülervorstellungen zur Evolution–eine quantitative Studie. [Students conceptions on evolution - a qualitative study]. *Berichte des Instituts für Didaktik der Biologie der Westfälischen Wilhelms-Universität Münster (IDB), 14*, 23–48.

Jungck, J. R. (1997). Ten equations that changed biology: mathematics in problem-solving biology curricula. *BioScience, 23*, 11–36.

Kærn, M., Elston, T. C., Blake, W. J., & Collins, J. J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics, 6*, 451–464. doi:10.1038/nrg1615

Kahneman, D. (2012). *Thinking, fast and slow*. London, United Kingdom: Penguine Books.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, United Kingdom: Cambridge University Press.

Kalinowski, S. T., Leonard, M. J., & Taper, M. L. (2016). Development and validation of the Conceptual Assessment of Natural Selection (CANS). *CBE-Life Sciences Education, 15*, ar64.

Kampourakis, K., & Zogza, V. (2007). Students' preconceptions about evolution: How accurate is the characterization as "Lamarckian" when considering the history of evolutionary thought? *Science & Education, 16*, 393–422. doi:10.1007/s11191-006-9019-9

Kampourakis, K., & Zogza, V. (2008). Students' intuitive explanations of the causes of homologies and adaptations. *Science & Education, 17*, 27–47. doi:10.1007/s11191-007-9075-9

Kampourakis, K., & Zogza, V. (2009). Preliminary evolutionary explanations: A basic framework for conceptual change and explanatory coherence in evolution. *Science & Education, 18*, 1313–1340. doi:10.1007/s11191-008-9171-5

Kaplan, J. J., Rogness, N. T., & Fisher, D. G. (2014). Explotting lexical ambiguity to help students understand the meaning of random. *Statistics Education Research Journal, 13*, 9–24.

Karunaratne, P. S. M., Breyer, Y. A., & Wood, L. N. (2016). Transforming the economics curriculum by integrating threshold concepts. *Education+Training, 58*, 492–509. doi:10.1108/ET-02-2016-0041

Kattmann, U. (2015). *Schüler besser verstehen: Alltagsvorstellungen im Biologieunterricht [Understanding school students: Everyday images in biology lessons]*. Hallbergmoos, Germany: Aulis.

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

Kelemen, D. (2012). Teleological minds: How natural intuitions about agency and purpose influence learning about evolution. In K. S. Rosengren & E. M. Evans (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution*: Oxford University Press.

Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology, 25*, 111–146. doi:10.1006/cogp.1993.1003

Kombartzky, U., Ploetzner, R., Schlag, S., & Metz, B. (2010). Developing and evaluating a strategy for learning from animations. *Learning and Instruction, 20*, 424–433. doi:10.1016/j.learninstruc.2009.05.002

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Kuzmak, S. D., & Gelman, R. (1986). Young children's understanding of random phenomena. *Child Development, 57*, 559–566.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago, IL: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Land, R. (2011). There could be trouble ahead: Using threshold concepts as a tool of analysis. *International Journal for Academic Development, 16*, 175–178. doi:10.1080/1360144X.2011.568747

Large, A. (1996). Computer animation in an instructional environment. *Library & Information Science Research, 18*, 3–23. doi:10.1016/S0740-8188(96)90028-6

Lazonder, A. W., Hagemans, M. G., & de Jong, T. (2010). Offering and discovering domain information in simulation-based inquiry learning. *Learning and Instruction, 20*, 511–520. doi:10.1016/j.learninstruc.2009.08.001

Lee, S. W.-Y., & Tsai, C.-C. (2013). Technology-supported learning in secondary and undergraduate biological education: observations from literature review. *Journal of Science Education and Technology, 22*, 226–233. doi:10.1007/s10956-012-9388-6

Lenormand, T., Roze, D., & Rousset, F. (2009). Stochasticity in evolution. *Trends in Ecology & Evolution, 24*, 157–165. doi:10.1016/j.tree.2008.09.014

Lenski, R. E., & Mittler, J. E. (1993). The directed mutation controversy and neo-Darwinism. *Science, 259*, 188–194. doi:10.1126/science.7678468

Leonard, M. J., Kalinowski, S. T., & Andrews, T. C. (2014). Misconceptions yesterday, today, and tomorrow. *CBE-Life Sciences Education, 13*, 179–186. doi:10.1187/cbe.13-12-0244

Li, W.-H. (1997). *Molecular evolution.* Sunderland, MA: Sinauer Associates.

Lord, T., & Marino, S. (1993). How university students view the theory of evolution. *Journal of College Science Teaching, 22*, 353–357.

Lowe, R. K. (2003). Animation and learning: selective processing of information in dynamic graphics. *Learning and Instruction, 13*, 157–176. doi:10.1016/S0959-4752(02)00018-X

Lui, M., & Slotta, J. D. (2014). Immersive simulations for smart classrooms: exploring evolutionary concepts in secondary science. *Technology, Pedagogy and Education, 23*, 57–80. doi:10.1080/1475939x.2013.838452

Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences, 104*, 8597–8604. doi:10.1073/pnas.0702207104

Marbach-Ad, G., Rotbain, Y., & Stavy, R. (2008). Using computer animation and illustration activities to improve high school students' achievement in molecular genetics. *Journal of Research in Science Teaching, 45*, 273–292. doi:10.1002/tea.20222

Masel, J. (2012). Rethinking Hardy–Weinberg and genetic drift in undergraduate biology. *BioEssays, 34*, 701–710. doi:10.1002/bies.201100178

Maskiewicz, A. C., & Lineback, J. E. (2013). Misconceptions are “so yesterday!”. *CBE-Life Sciences Education, 12*, 352–356. doi:10.1187/cbe.13-01-0014

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*, 1–19. doi:10.1207/s15326985ep3201_1

Mayer, R. E. (2011). Instruction based on visualizations. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (second ed., pp. 427–445). New York, NY: Routledge Chapman & Hall.

Mayer, R. E. (2014). Cognitive theory of muldtimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 43–71). New York, NY: Cambridge University Press.

Mayr, E. (2001). *What evolution is*. New York, NY: Basic books.

Mead, L. S., & Scott, E. C. (2010). Problem concepts in evolution part II: Cause and chance. *Evolution: Education and Outreach, 3*, 261–264. doi:10.1007/s12052-010-0231-3

Meyer, J. H., & Land, R. (2003). Threshold concepts and troublesome knowledge: linkages to ways of thinking and practising within the disciplines. In C. Rust (Ed.), *Improving Student Learning: Theory and Practice Ten Years On* (pp. 412–424). Oxford, United Kingdom: Oxford Centre for Staff and Learning Development (OCSLD).

Meyer, J. H., & Land, R. (2006). Threshold concepts and troublesome knowledge: an introduction. In J. H. Meyer & R. Land (Eds.), *Overcoming barriers to student understanding: Threshold concepts and troublesome knowledge* (pp. 3–18). Abington, United Kingdom: Routledge.

Meyer, J. H., Land, R., & Baillie, C. (2010). *Threshold concepts and transformational learning*. Rotterdam, The Netherlands: Sense Publishers.

Ministerium für Schule und Berufsbildung des Landes Schleswig-Holstein (2016). *Fachanforderungen Biologie. Allgemein bildende Schulen, Sekundarstufe I – Gymnasium, Sekundarstufe II [Subject requirements biology. Grammar schools, lower and upper secondary school]*. Retreived from http://lehrplan.lernnetz.de/index.php?wahl=199 (accessed 03 April 2018)

Miller, J. D., Scott, E. C., & Okamoto, S. (2006). Public acceptance of evolution. *Science, 313*, 765-766. doi:10.1126/science.1126746

Millstein, R. L. (2000). Chance and macroevolution. *Philosophy of Science, 67*, 603–624. doi:10.1086/392857

Mix, L. J., & Masel, J. (2014). Chance, purpose, and progress in evolution and christianity. *Evolution, 68*, 2441–2451. doi:10.1111/evo.12434

Moreno, R., & Valdez, A. (2005). Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback. *Educational Technology Research and Development, 53*, 35–45. doi:10.1007/BF02504796

Nadelson, L. S., & Sinatra, G. M. (2009). Educational professionals' knowledge and acceptance of evolution. *Evolutionary Psychology, 7*, 490–516. doi:10.1177/147470490900700401

Nadelson, L. S., & Southerland, S. (2012). A more fine-grained measure of students' acceptance of evolution: development of the Inventory of Student Evolution Acceptance—I-SEA. *International Journal of Science Education, 34*, 1637–1666. doi:10.1080/09500693.2012.702235

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.* Washington, DC: The National Academies Press.

Nationale Akademie der Wissenschaften Leopoldina (2017): *Evolutionsbiologische Bildung in Schule und Universität.* [Teaching evolutionary biology at schools and universities] Deutsche Akademie der Naturforscher Leopoldina e.V. Nationale Akademie der Wissenschaften, Halle (Saale). Retreived from https://www.leopoldina.org/uploads/tx_leopublication/2017_Stellungnahme_Evolutionsbiologie.pdf (accessed 03 April 2018)

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

Nehm, R. H., Rector, M. A., & Ha, M. (2010). "Force-Talk" in evolutionary explanation: Metaphors and misconceptions. *Evolution: Education and Outreach, 3*, 605–613. doi:10.1007/s12052-010-0282-5

Nehm, R. H., & Reilly, L. (2007). Biology majors' knowledge and misconceptions of natural selection. *BioScience, 57*, 263–272. doi:10.1641/b570311

Nehm, R. H., & Ridgway, J. (2011). What do experts and novices "see" in evolutionary problems? *Evolution: Education and Outreach, 4*, 666–679. doi:10.1007/s12052-011-0369-7

Nehm, R. H., & Schonfeld, I. S. (2007). Does increasing biology teacher knowledge of evolution and the nature of science lead to greater preference for the teaching of evolution in schools? *Journal of Science Teacher Education, 18*, 699–723. doi:10.1007/s10972-007-9062-7

Neubrand, C. (2017). *Evolution unterrichten : Untersuchung von Aspekten erfolgreichen Lehrens und Lernens mit Fokus auf Beispielaufgaben und vorwissensangepasste Impulse [Teaching evolution: Investigate aspects of successful teaching and learning with focus on worked examples and knowledge adapted prompts]* (doctoral dissertation). Kiel University, Kiel, Germany.

Neubrand, C., Borzikowsky, C., & Harms, U. (2016). Adaptive prompts for learning evolution with worked examples - Highlighting the students between the "Novices" and the "Experts" in a classroom. *International Journal of Environmental and Science Education, 11*, 6774–6795.

NGSS Lead States (2013). *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press.

Niedersächsisches Kultusministerium (2009): *Kerncurriculum für das Unterrichtsfach Biologie für die gymnasiale Oberstufe am Gymnasium und an der Gesamtschule, für das Berufliche Gymnasium, das Abendgymnasium und das Kolleg [Curriculum for the subject biology for upper secondary level at the Gymnasium and at the Gesamtschule, for the Berufs Gymnasium, the Abendgymnasium and the Kolleg]*. Retreived from http://db2.nibis.de/1db/cuvo/datei/kc_biologie_go_i_2009.pdf (accessed 03 April 2018)

Niedersächsisches Kultusministerium (2015): *Kerncurriculum für die Naturwissenschaften in den Schuljahrgängen 5–10 des Gymnasiums [Curriculum for the natural sciences in grades 5-10 of the Gymnasium]*. Retreived from http://www.nibis.de/uploads/1gohrgs/kc_2015/NW_Gym_SI_KC_Druck.pdf (accessed 03 April 2018)

Niedersächsisches Kultusministerium (2017): *Kerncurriculum für das Unterrichtsfach Biologie für die gymnasiale Oberstufe am Gymnasium und an der Gesamtschule, für das Berufliche Gymnasium, das Abendgymnasium und das Kolleg [Curriculum for the subject biology for upper secondary level at the Gymnasium and at the Gesamtschule, for the Berufs Gymnasium, the Abendgymnasium and the Kolleg]*. Retreived from http://db2.nibis.de/1db/cuvo/datei/bi_go_kc_druck_2017.pdf (accessed 03 April 2018)

OECD, Organisation for Economic Co-operation and Development (2010). *PISA computer-based assessment of student skills in science.* Paris, France: OECD.

Opfer, J. E., Nehm, R. H., & Ha, M. (2012). Cognitive foundations for science assessment design: Knowing what students know about evolution. *Journal of Research in Science Teaching, 49*, 744–777. doi:10.1002/tea.21028

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (second ed., pp. 27–43). New York, NY: Cambridge University Press.

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York, NY: Oxford University Press.

Palmer, D. H. (1999). Exploring the link between students' scientific and nonscientific conceptions. *Science Education, 83*, 639–653. doi:10.1002/(Sici)1098-237x(199911)83:6<639::Aid-Sce1>3.0.Co;2-O

Park, E. J., & Light, G. (2009). Identifying atomic structure as a threshold concept: Student mental models and troublesomeness. *International Journal of Science Education, 31*, 233–258. doi:10.1080/09500690701675880

Peker, D., Comert, G. G., & Kence, A. (2010). Three decades of anti-evolution campaign and its results: Turkish undergraduates’ acceptance and understanding of the biological evolution theory. *Science & Education, 19*, 739–755. doi:10.1007/s11191-009-9199-1

Perkins, D. (2006). Constructivism and troublesome knowledge. In J. H. Meyer & R. Land (Eds.), *Overcoming barriers to student understanding: Threshold concepts and troublesome knowledge* (pp. 33–47). Abington, United Kingdom: Routledge.

Pfeiffer, V. D., Scheiter, K., & Gemballa, S. (2012). Comparing and combining traditional teaching approaches and the use of video clips for learning how to identify species in an aquarium. *Journal of Biological Education, 46*, 140–148. doi:10.1080/00219266.2011.584892

Piaget, J. (1929). *The child's conception of the world*. New York: Harcourt Brace.

Piaget, J. (1930). *The chield's conception of physical causality*. London, United Kingdom: Kegan Paul.

Piaget, J. (1974). *Understanding causality*. New York: W. W. Norton.

Ploetzner, R., & Lowe, R. (2004). Dynamic visualisations and learning. *Learning and Instruction, 14*, 235–240. doi:10.1016/j.learninstruc.2004.06.001

Ploetzner, R., & Lowe, R. (2012). A systematic characterisation of expository animations. *Computers in Human Behavior, 28*, 781–794. doi:10.1016/j.chb.2011.12.001

Ploetzner, R., Lowe, R., & Schlag, S. (2013). A systematic characterization of cognitive techniques for learning from textual and pictorial representations. *Journal of Education and Learning, 2*, 78–95. doi:10.5539/jel.v2n2p78

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*, 211–227. doi:10.1002/sce.3730660207

Pramling, N. (2008). The role of metaphor in Darwin and the implications for teaching evolution. *Science Education, 93*, 535–547. doi:10.1002/sce.20319

Price, R. M., Pope, D. S., Abraham, J. K., Maruca, S., & Meir, E. (2016). Observing populations and testing predictions about genetic drift in a computer simulation improves college students' conceptual understanding. *Evolution: Education and Outreach, 9*, 1–14. doi:10.1186/s12052-016-0059-6

Rapp, D. N. (2005). Mental models: Theoretical issues for visualizations in science education. In J. K. Gilbert (Ed.), *Visualization in science education* (pp. 43–60). Dordrecht, The Netherlands: Springer.

Rapp, D. N., & Kurby, C. A. (2008). The 'ins' and 'outs' of learning: Internal representations and external visualizations. In J. K. Gilbert (Ed.), *Visualization: Theory and practice in science education* (pp. 29–52). Dordrecht, The Netherlands: Springer.

Raup, D. M., Gould, S. J., Schopf, T. J., & Simberloff, D. S. (1973). Stochastic models of phylogeny and the evolution of diversity. *The Journal of Geology, 81*, 525–542. doi:10.1086/627905

Rector, M. A., Nehm, R. H., & Pearl, D. (2013). Learning the language of evolution: Lexical ambiguity and word meaning in student explanations. *Research in Science Education, 43*, 1107–1133. doi:10.1007/s11165-012-9296-z

Reid, D. J., Zhang, J., & Chen, Q. (2003). Supporting scientific discovery learning in a simulation environment. *Journal of Computer Assisted Learning, 19*, 9–20. doi:10.1046/j.0266-4909.2003.00002.x

Renkl, A. (2005). The worked-out examples principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. Cambridge, United Kingdom: Cambridge University Press.

Robson, R. L., & Burns, S. (2011). Gain in student understanding of the role of random variation in evolution following teaching intervention based on Luria-Delbruck experiment. *Journal of Microbiology & Biology Education: JMBE, 12*, 3–7. doi:10.1128/jmbe.v12i1.272

Ross, P. M., Taylor, C. E., Hughes, C., Kofod, M., Whitaker, N., Lutze-Mann, L., & Tzioumis, V. (2010). Threshold concepts: Challenging the way we think, teach and learn in biology. In J. H. F. Meyer, R. Land, & C. Baillie (Eds.), *Threshold Concepts and Transformational Learning* (Vol. 1, pp. 165–177). Rotterdam, The Netherlands: SensPublishers.

Ross, P. M., Taylor, C. E., Hughes, C., Whitaker, N., Lutze-Mann, L., Kofod, M., & Tzioumis, V. (2010). Threshold concepts in learning biology and evolution. *Biology International, 47*, 47–52.

Rutten, N., van Joolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education, 58*, 136–153. doi:10.1016/j.compedu.2011.07.017

Ryoo, K., & Linn, M. C. (2012). Can dynamic visualizations improve middle school students' understanding of energy in photosynthesis? *Journal of Research in Science Teaching, 49*, 218–243. doi:10.1002/tea.21003

Sanders, M., & Ngxola, N. (2009). Addressing teachers' concerns about teaching evolution. *Journal of Biological Education, 43*, 121–128. doi:10.1080/00219266.2009.9656166

Scharfenberg, F.-J. (2005). *Experimenteller Biologieunterricht zu Aspekten der Gentechnik im Lernort Labor: empirische Untersuchung zu Akzeptanz, Wissenserwerb und Interesse [Experimental biology sessions on aspects of genetic engineering in the learning laboratory: Empirical study on acceptance, knowledge acquisition and interest]* (doctoral dissertation). University of Bayreuth, Bayreuth, Germany.

Schnotz, W. (2002). Commentary: Towards an integrated view of learning from text and visual displays. *Educational Psychology Review, 14*, 101–120. doi:10.1023/A:1013136727916

Schnotz, W. (2014). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (second ed., pp. 72–103). New York, NY: Cambridge University Press.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction, 13*, 141–156. doi:10.1016/S0959-4752(02)00017-8

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss (Jahrgangsstufe 10) (Beschluss der Kultusministerkonferenz vom 16.12.2004) [National educational standards for the subject of biology concerning the 10th grade]*. Munich, Germany: Luchterhand.

Shtulman, A. (2006). Qualitative differences between naïve and scientific theories of evolution. *Cognitive Psychology, 52*, 170–194. doi:10.1016/j.cogpsych.2005.10.001

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition, 124*, 209–215. doi:10.1016/j.cognition.2012.04.005

Simulation (2018). In Merriam-Webster's Collegiate Dictionary. Retrieved from https://www.merriam-webster.com/dictionary/simulation (accessed 03 April 2018)

Sinatra, G. M., Brem, S. K., & Evans, E. M. (2008). Changing minds? Implications of conceptual change for teaching and learning about biological evolution. *Evolution: Education and Outreach, 1*, 189–195. doi:10.1007/s12052-008-0037-8

Sinatra, G. M., Southerland, S. A., McConaughy, F., & Demastes, J. W. (2003). Intentions and beliefs in students' understanding and acceptance of biological evolution. *Journal of Research in Science Teaching, 40*, 510–528. doi:10.1002/tea.10087

Smith, M. U., & Siegel, H. (2004). Knowing, believing, and understanding: What goals for science education? *Science & Education, 13*, 553–582. doi:10.1023/B:SCED.0000042848.14208.bf

Smith, M. U., & Siegel, H. (2016). On the relationship between belief and acceptance of evolution as goals of evolution education. *Science & Education, 25*, 473–496. doi:10.1007/s11191-016-9836-4

Sniegowski, P. D., Gerrish, P. J., Johnson, T., & Shaver, A. (2000). The evolution of mutation rates: separating causes from consequences. *BioEssays, 22*, 1057–1066. doi:10.1002/1521-1878(200012)22:12<1057::AID-BIES3>3.0.CO;2-W

Soderberg, P., & Price, F. (2003). An examination of problem-based teaching and learning in population genetics and evolution using EVOLVE, a computer simulation. *International Journal of Science Education, 25*, 35–55. doi:10.1080/09500690110095285

Southerland, S. A., Abrams, E., Cummins, C. L., & Anzelmo, J. (2001). Understanding students' explanations of biological phenomena: Conceptual frameworks or p-prims? *Science Education, 85*, 328–348. doi:10.1002/sce.1013

Spanjers, I. A., Wouters, P., van Gog, T., & van Merriënboer, J. J. (2011). An expertise reversal effect of segmentation in learning from animated worked-out examples. *Computers in Human Behavior, 27*, 46–52. doi:10.1016/j.chb.2010.05.011

Speth, E. B., Long, T. M., Pennock, R. T., & Ebert-May, D. (2009). Using Avida-ED for teaching and learning about evolution in undergraduate introductory biology courses. *Evolution: Education and Outreach, 2*, 415–428. doi:10.1007/s12052-009-0154-z

Spindler, L., & Doherty, J. (2009). Assessment of the teaching of evolution by natural selection through a hands-on simulation. *Teaching Issues and Experiments in Ecology, 6*, 1–20.

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. A. Duschl & R. J. Hamilton (Eds.), *Philosophy of science, cognitive psychology, and educational theory and practice*. Albany, NY: State University of New York Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285. doi:10.1207/s15516709cog1202_4

Taylor, C. (2006). Threshold concepts in biology: do they fit the definition? In J. H. Meyer & R. Land (Eds.), *Overcoming barriers to student understanding: Threshold concepts and troublesome knowledge* (pp. 87–99). Abington, United Kingdom: Routledge.

Tibell, L. A. E., & Harms, U. (2017). Biological principles and threshold concepts for understanding natural selection. *Science & Education, 26*, 953–973. doi:10.1007/s11191-017-9935-x

Tibell, L. A. E., & Rundgren, C.-J. (2010). Educational challenges of molecular life science: characteristics and implications for education and research. *CBE-Life Sciences Education, 9*, 25–33. doi:10.1187/cbe.08-09-0055

To, C., Tenenbaum, H. R., & Hogh, H. (2017). Secondary school students' reasoning about evolution. *Journal of Research in Science Teaching, 54*, 247–273. doi:10.1002/tea.21347

Toulmin, S. (1972). *Human understanding*. Princeton, NJ: Princeton University Press.

Trani, R. (2004). I won't teach evolution; it's against my religion. And now for the rest of the story…. *The American Biology Teacher, 66*, 419–427. doi:10.1662/0002-7685(2004)066[0419:Iwtiam]2.0.Co;2

Trey, L., & Khan, S. (2008). How science students can learn about unobservable phenomena using computer-based analogies. *Computers & Education, 51*, 519–529. doi:10.1016/j.compedu.2007.05.019

Tversky, B. (2005). Prolegomenon to Scientifi Visualizations. In J. K. Gilbert (Ed.), *Visualizations in science education* (pp. 29–42). Dordrecht, The Netherlands: Springer.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate? *International Journal of Human-Computer Studies, 57*, 247–262. doi:10.1006/ijhc.2002.1017

Urhahne, D. (2002). *Motivation und Verstehen [Motivation and understanding]*. Münster, Germany: Waxmann Verlag.

Urhahne, D., & Harms, U. (2006). Instruktionale Unterstützung beim Lernen mit Computersimulationen. [Instructional support for learning with computer simulations]. *Unterrichtswissenschaft, 34*, 358–377.

van Berkum, J. J. A., & de Jong, T. (1991). Instructional environments for simulations. *Education and Computing, 6*, 305–358. doi:10.1016/0167-9287(91)80006-J

Vazquez, B. (2017). A state-by-state comparison of middle school science standards on evolution in the United States. *Evolution: Education and Outreach, 10*, 1–19. doi:10.1186/s12052-017-0066-2

Veenman, M. V., & Elshout, J. J. (1995). Differential effects of instructional support on learning in simultation environments. *Instructional Science, 22*, 363–383. doi:10.1007/BF00891961

Vosniadou, S. (2007). Conceptual change and education. *Human Development, 50*, 47–54. doi:10.1159/000097684)

Vreman-de Olde, C., & de Jong, T. (2006). Scaffolding learners in designing investigation assignments for a computer simulation. *Journal of Computer Assisted Learning, 22*, 63–73. doi:10.1111/j.1365-2729.2006.00160.x

Wagner, A. (2012). The role of randomness in Darwinian evolution. *Philosophy of Science, 79*, 95–119. doi:10.1086/663239

Wang, J.-Y., Wu, H.-K., & Hsu, Y.-S. (2017). Using mobile applications for learning: Effects of simulation design, visual-motor integration, and spatial ability on high school students' conceptual understanding. *Computers in Human Behavior, 66*, 103–113. doi:10.1016/j.chb.2016.09.032

Weitzel, H., & Gropengiesser, H. (2009). Vorstellungsentwicklung zur stammesgeschichtlichen Anpassung: Wie man Lernhindernisse verstehen und förderliche Lernangebote machen kann. [Development of students' conceptions of biological adaptation: Understanding students' learning difficulties and addressing them with appropriate learning activities]. *Zeitschrift für Didaktik der Naturwissenschaften, 15*, 287–305.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*, 3–118. doi:10.1207/s1532690xci1601_2

Wockenfuß, V., & Raatz, U. (2006). Über den Zusammenhang zwischen Testleistung und Klassenstufe bei muttersprachlichen C-Tests. [About the relationship between test performance and grade level in native language C-tests]. *Der C-Test: Theorie, Empirie, Anwendungen/The C-Test: Theory, Empirical Research, Applications*, 211–242.

Yaman, M., Nerdel, C., & Bayrhuber, H. (2008). The effects of instructional support and learner interests when learning using computer simulations. *Computers & Education, 51*, 1784–1794. doi:10.1016/j.compedu.2008.05.009

Yang, E.-M., Andre, T., Greenbowe, T. J., & Tibell, L. A. (2003). Spatial ability and the impact of visualization/animation on learning electrochemistry. *International Journal of Science Education, 25*, 329–349. doi:g/10.1080/09500690210126784

Yarden, H., & Yarden, A. (2010). Learning using dynamic and static visualizations: students' comprehension, prior knowledge and conceptual status of a biotechnological method. *Research in Science Education, 40*, 375–402. doi:10.1007/s11165-009-9126-0

Yates, T. B., & Marek, E. A. (2014). Teachers teaching misconceptions: a study of factors contributing to high school biology students' acquisition of biological evolution-related misconceptions. *Evolution: Education and Outreach, 7*, 1–18. doi:10.1186/s12052-014-0007-2

Yates, T. B., & Marek, E. A. (2015). A study identifying biological evolution-related misconceptions held by prebiology high school students. *Creative Education, 6*, 811–834. doi:10.4236/ce.2015.68085

Zabel, J., & Gropengiesser, H. (2011). Learning progress in evolution theory: climbing a ladder or roaming a landscape? *Journal of Biological Education, 45*, 143–149. doi:10.1080/00219266.2011.586714

Zander, C., Boustedt, J., Eckerdal, A., McCartney, R., Moström, J.-E., Ratcliffe, M., & Sanders, K. (2008). Threshold concepts in computer science: a multi-national empirical investigation. In R. Land, J. H. F. Meyer, & J. Smith (Eds.), *Threshold concepts within the disciplines* (pp. 102–118). Rotterdam, The Netherlands: Sense.

Zhang, J., Chen, Q., Sun, Y., & Reid, D. J. (2004). Triple scheme of learning support design for scientific discovery learning based on computer simulation: experimental research. *Journal of Computer Assisted Learning, 20*, 269–282. doi:10.1111/j.1365-2729.2004.00062.x

DECLARATION

I hereby declare that the work presented in this dissertation — apart from the advice given by my supervisor — is my own in both format and content. This is my first dissertation and the work has never been used in any other dissertation attempts. The dissertation complies with the standards for good scientific practice proposed by the German Research Foundation (DFG). As indicated at their respective beginnings, one research article (Chapter 4) and three manuscripts (Chapter 5–7) presented in this dissertation have been published, submitted, or are intended for submission at scientific journals.

Kiel, April 4th, 2018 ____

Daniela Fiedler

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my supervisor Ute Harms for her constant support and demand that have significantly contributed to my professional and personal development. I am very thankful for her willingness to foster my research stay at Stony Brook University in the United States.

I would also like to thank Ross Nehm for his enormous hospitality, advice, and great discussions on statistical reasoning in evolution education during my time at both Stony Brook University and NARST 2018. Special thanks to Gena Sbeglia for inviting me to the Malaysian dinner in Atlanta. It was a delicious experience!

My grateful thanks are extended to my mentors Jörg Großschedl and Steffen Tröbst for their supporting help concerning data analyses. Warm thanks are also dedicated to my current and former colleagues with whom I shared great times at the IPN. Thank you, Julia Schwanewedel, Birgit Heyduck, Julia Arnold, Till Bruckermann, Yasmin Appelhans, Sebastian Opitz, Ulrike Wernecke, Charlotte Neubrand, Deidre Bauer, Frauke Voitle, Hanno Michel, Julian Fischer, Dirk Mittenzwei, Katharina Zimmermann, and Ulrike Gessner-Thiel.

Special thanks also to Daniela Mahler, Devasmita Chakraverty, Jasmin Colakoglu, Johanna Hansen, Carola Garrecht, and Lara Magnus for sharing both offices inside and wonderful times outside the IPN. I also want to thank Marc Rodemer, Denise Buchmann, and Charlotte Hölterhoff for assisting in data collection and entry.

Grateful thanks are also extended to my colleagues of the *EvoVis*-project in Sweden for their warm welcome and great discussions about threshold concepts. Thank you to all of you with whom I have had the pleasure to work with during my dissertation project!

My most profound gratitude is dedicated to my best friend Sarah-Janina for writing magical stories around loving characters that drew me into fantastic worlds. I would not be the person I am today without her wonderful friendship. Thank you for being there!

Finally, I want to thank my family for their ongoing support and encouragement. Special thanks to my dad for living two months in Kiel to take care of my two cats Freya and Skadi, who teach me now and then how to relax by settling down on my work for a short catnap. You make my day!

CURRICULUM VITAE

Personal Data

Name:	Daniela Fiedler
Date of Birth:	August 3rd, 1987
Place of Birth:	Offenbach am Main, Germany
Nationality:	German

Education

Since 08/2014 — Christian-Albrechts University at Kiel, Germany
- PhD Student in Science Education
- Supervisor: Professor Dr Ute Harms (Biology Education)

10/2011–12/2013 — Justus-Liebig-University, Giessen, Germany
- Master of Science (grade 1.2)
- Focus on Animal Ecology and Conservation
- Master's thesis: *Effects of management type and landscape diversity on web building spiders and their prey*

10/2008–10/2011 — Justus-Liebig-University, Giessen, Germany
- Bachelor of Science (grade 2.3)
- Focus on Ecology, Biology Education, and Conservation
- Bachelor's thesis: *Investigation of visitors' behavior in the 'Katzendschungel' of the Zoological Garden Frankfurt am Main*

Professional Experience

Since 08/2014 — Leibniz Institute for Science and Mathematics Education (IPN), Kiel, Germany
- Research Assistant at the Department of Biology Education

02/2014–05/2014 — Justus-Liebig-University, Giessen, Germany
- Student Assistant at the Department of Animal Ecology & Systematics

09/2013 Justus-Liebig-University, Giessen, Germany
- Student Assistant at the Department of Animal Ecology & Systematics

05/2012–07/2012 Justus-Liebig-University, Giessen, Germany
- Student Assistant at the Department of Biology Education

International Experience

04/2017–06/2017 Stony Brook University, New York, USA
- Visiting Scholar
- Hosted at the Department of Ecology & Evolution by Professor Dr Ross H Nehm

12/2007–07/2008 Work-and-Travel in Australia

Achievements

2017 Travel award for the research stay at Stony Brook University (USA), ESERA Early Career Research Travel Award 2017, 1.000€

2018 Travel grant for the 2018 NARST Annual International Conference in Atlanta (USA), German Academic Exchange Service (DAAD), 1.684€

Publications

Peer review – published

Karp, D. S., Chaplin-Kramer, R., Meehan, T. D., Martin, E. A., …, **Fiedler, D.,** … (2018). Crop pests and predators exhibit inconsistent responses to surrounding landscape composition. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 115*, E7863–E7870. doi:10.1073/pnas.1800042115

Fiedler, D., Tröbst, S., & Harms, U. (2017). University students' conceptual knowledge of randomness and probability in the context of evolution and mathematics. *CBE-Life Sciences Education, 16*, ar38. doi: 10.1187/cbe.16-07-0230

Mader, V., Diehl, E., **Fiedler, D.**, Thorn, S., Wolters, V., & Birkhofer, K. (2017). Trade-offs in arthropod conservation between productive and non-productive agrienvironmental schemes along a landscape complexity gradient. *Insect Conservation and Diversity, 10*, 236–247. doi: 10.1111/icad.12220

Fiedler, D. & Harms, U. (2016). Die Bedeutung eines Begriffs von Zufall und Wahrscheinlichkeit für das Evolutionsverständnis – Pilotstudie zur Entwicklung eines Testinstruments [The relevance of the concepts randomness and probability for the understanding of evolution – a pilot study for the development of a test instrument]. In U. Gebhardt & M. Hammann (Eds.), *Lehr- und Lernforschung in der Biologiedidaktik* (Bd. 7, pp. 95–109). Innsbruck, Austria: Studienverlag.

Mader, V., Birkhofer, K., **Fiedler, D.**, Thorn, S. Wolters, V., & Diehl, E. (2016). Land use at different spatial scales alters the functional role of web-building spiders in arthropod food-webs. *Agriculture, Ecosystems and Environment, 219*, 152–162. doi: 10.1016/j.agee.2015.12.017

Peer review – accepted

Harms, U. & **Fiedler, D.** (2018). Training the understanding of randomness to support evolution learning. In: U. Harms & M. Reiss (Eds.), *Evolution Education Re-considered – understanding what works*. Springer

Fiedler, D., Sbeglia, G., Nehm, R. H., & Harms, U. (accepted with major revisions, 07/2018). How strongly does statistical reasoning influence knowledge and acceptance of evolution? *Journal of Research in Science Teaching (JRST).*

Intended for peer review publication

Fiedler, D., Tröbst, S., Großschedl, J. & Harms, U. (minor revisions, 08/2018). EvoSketch: Simple simulations for learning random and probabilistic processes in evolution, and effects of instructional support on learners' conceptual knowledge. *Evolution: Education and Outreach.*

Göransson, A., Orraryd, D., **Fiedler, D.**, & Tibell, L. A. E. (in preparation). Item context affects the use of threshold concepts in student explanation of evolution by natural selection.