Evolution: Education and Outreach

RESEARCH ARTICLE

Open Access

Measuring biology trainee teachers' professional knowledge about evolution—introducing the Student Inventory

Julian Fischer[1]* , Thorben Jansen[2], Jens Möller[2] and Ute Harms[1]

Abstract

Background: To teach evolution efficiently teachers must be able to diagnose their students' ideas and understanding of the phylogeny of organisms. This encompasses different facets of content-specific professional knowledge, that is, knowledge about core ideas and theories, as well as knowledge about respective misconceptions. However, as findings from the field of psychology have shown, diagnostic activities comprise a further facet, namely, teachers' judgment accuracy. This refers to the question of whether achievement-irrelevant information about the student influences teachers' diagnoses. Against this background we conducted a study (1) to assess trainee teachers' abilities to diagnose (a) the scientific correctness of students' written answers, (b) students' misconceptions about evolution, and (2) to investigate the interplay of evolution specific and generic facets of professional knowledge during the diagnosis. For this purpose, we applied a digital instrument, the Student Inventory (SI). Using this instrument, the trainee teachers (N = 27) first diagnosed written answers (N = 6) from virtual students regarding their scientific correctness and regarding students' misconceptions about the natural selection of the peppered moth. Second, to test for judgment accuracy, the trainee teachers received—via the SI—achievement-irrelevant information about each virtual student, that is, the previous result of a multiple-choice questionnaire about evolution, before diagnosing the written answers.

Results: The trainee teachers were able to distinguish between scientifically correct (90.8%) and scientifically incorrect (91.7%) written answers. Trainee teachers faced problems when diagnosing specific misconceptions categories. Anthropomorphic misconceptions were diagnosed significantly more often (61.1%) than teleological misconceptions (27.8%). The achievement-irrelevant information influenced the trainee teachers' assessment of written answers (F [1,26] = 5.94, p < .022, η^2 = .186) as they scored the written answers higher if the performance in the questionnaire was good and vice versa.

Conclusion: The findings indicate that the diagnosis is easier or more difficult depending on the particular misconception category. However, the findings also reveal that, besides the evolution-specific facets of professional knowledge, generic facets interrelate with the quality of the diagnosis result. We conclude from these findings that an integration of evolution-specific and generic knowledge into the education of biology teachers is critical.

Introduction

The mission of evolution education is 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

*Correspondence: jfischer@leibniz-ipn.de
[1] Department of Biology Education, IPN — Leibniz Institute for Science and Mathematics Education, Olshausenstrasse 62, 24118 Kiel, Germany
Full list of author information is available at the end of the article

world. However, students throughout the educational hierarchy, the public, and even science teachers lack an understanding of the relevant principles and concepts of evolutionary change (e.g., Nadelson and Sinatra 2009; Nehm and Reilly 2007). Many of them also resist accepting the theory of evolution as the best scientific explanation for the similarities among organisms, for biological diversity, and for various features and processes in the living world (Berkman and Plutzer 2011). The factors influencing understanding are diverse and include influences of alternative conceptions (e.g., Kampourakis and Nehm 2014), mismatches of everyday language and scientific terminology (Rector et al. 2013), selection of model organisms and traits (i.e., animal vs. plant, trait gain vs. trait loss; Großschedl et al. 2018), students' thinking dispositions (Athanasiou and Papadopoulou 2012), feeling of certainty (Ha et al. 2012), and students' religious views (e.g., Allmon 2011; Athanasiou and Papadopoulou 2012). Thus, the correct teaching of the evolutionary theory by biology teachers is highly important for students, as it acts as a central link between different concepts and highlights the similarities in the complexity of biological concepts (Tibell and Harms 2017).

The teacher is one of the most important determinants of students' performance (e.g., Mahler et al. 2017). Teachers make a difference in the achievement of their students in science classrooms. Many studies have shown that teachers' professional knowledge is the key factor for teaching (e.g., Abell 2007; Kunter et al. 2013). Based on Shulman's (1987) taxonomy, professional knowledge in educational sciences can be differentiated into three main facets: content knowledge (CK), pedagogical content knowledge (PCK), and pedagogical knowledge (PK). In the context of evolution education, for example, CK is necessary to identify key ideas and principles about evolutionary mechanisms such as natural selection (e.g., Großschedl et al. 2015), PCK to diagnose prominent student misconceptions (e.g., Ziadie and Andrews 2018), and PK to make accurate judgments about the performance shown (e.g., Schrader 2006). Thus, these domains of professional knowledge and the corresponding facets play an important role in the diagnostic activities of teachers, for example, in the assessment of students' statements or written performance (e.g., Helmke et al. 2004). Without a comprehensive diagnosis of existing misconceptions, a subsequent individualized support of the students to enable them to succeed in achieving an elaborated understanding of evolution is not possible.

The aim of the present study was to investigate trainee teachers' biology-specific and generic facets of professional knowledge. The focus was on knowledge regarding evolutionary theory, specifically the process of natural selection. In an effort to gain insights into biology

teachers' diagnostic activities based on their professional knowledge, we further developed the digital survey instrument – the Student Inventory (SI)—of Kaiser et al. (2015). Furthermore, we gained first indications about the relationship between the declarative and procedural knowledge of the biology trainee teachers. In the following sections, we strongly focus on the teaching and learning of evolution, focusing on prominent student misconceptions related to natural selection as well as on the biology trainee teachers' professional knowledge.

Theoretical background
Teaching and learning evolution
Evolution is the central comprehensive explanatory framework not only in biology but also in all of the life sciences (e.g., German National Academy of Sciences Leopoldina 2017; National Research Council 2012). It has the power to explain the diversity of life and to foster the understanding of how and why populations change over time (e.g., Sickel and Friedrichsen 2013). Therefore, the theory of evolution is part of numerous science education curricula and standards in many countries (e.g., Germany: Standing Conference of the Ministers of Education and Cultural Affairs of the Federal States in the Federal Republic of Germany [KMK] 2005; USA: Next Generation Science Standards [NGSS] 2013). Evolutionary processes are the basis for all topics in the field of biology and form the foundation for a conceptual understanding of the life sciences (e.g., Anderson et al. 2002; Basel et al. 2014; Bishop and Anderson 1990; Furtak 2012; Opfer et al. 2012; Zabel and Gropengießer 2011). Despite the complexity of the theory of evolution, the core ideas of evolution can be summarized in a few sentences. Biological evolution can be defined at the lowest level (microevolution) as changes in the allele frequency of a population. These changes and the resulting genetic variation are caused by recombination and mutation. Variability is the basis of the core mechanism of evolution—natural selection, which describes the adaptation within species. Those individuals that are genetically better adapted to environmental conditions are more likely to survive as well as to have a higher reproduction rate. In the next generation, these beneficial traits occur more frequently and the population adapts to the environmental conditions (e.g., Andersson and Wallin 2006). Although there are differences in the number of relevant key concepts (e.g., Anderson et al. 2002; Moharreri et al. 2014; Nehm and Schonfeld 2008), the concepts of variation, inheritance, and selection provide a sufficient explanation of evolutionary change through the process of natural selection (Endler 1986; Mayr 1982).

However, the complexity of the evolutionary theory leads to great problems in teaching and learning.

Research has identified evolution as a very challenging component of the science curriculum and indicates that traditional teaching approaches are ineffective in transforming students' misconceptions into scientifically adequate ways of thinking (e.g., Basel et al. 2013; Bishop and Anderson 1990; Gregory 2009; Kampourakis and Zogza 2008; McVaugh et al. 2011; Opfer et al. 2012; Pazza et al. 2010). A major impediment to learning is the numerous misconceptions that exist with regard to the theory of evolution. Misconceptions are understood to be students' ideas and thoughts that are incompatible with scientific knowledge (Yip 1998). Students' misconceptions are dependent on their experiences, the language used in daily communication, the lack of CK of their teachers, and the textbooks they use (e.g., King 2010; Nehm et al. 2009). Misconceptions about natural phenomena arise at an early age and even before school education, when children explore their physical and social environments (e.g., Beardsley 2004; Bruckermann et al. 2020; Driver 1988; Evans 2000). These misconceptions are used to explain evolutionary mechanisms. Consequently, misconceptions not only are found at the level of young children but are also anchored in the minds of high school students (Demastes et al. 1995), biology majors (e.g., Dagher and BouJaoude 1997; Nehm and Reilly 2007; Nehm and Schonfeld 2008), medicine students (Brumby 1984), and science teachers (Nehm and Schonfeld 2007). The fact that the principles of evolutionary biology are widely misunderstood by students as well as by large parts of the public has motivated educators and researchers to focus on identifying evolutionary misconceptions and finding instructional strategies to overcome these (for an overview: on misconceptions, see Gregory (2009); on teaching strategies, see Ziadie and Andrews 2018; Harms and Reiss 2019).

Evolutionary misconceptions are manifold and refer to different evolutionary mechanisms. Evans (2000) observed that the process of the origin of species is seen as a spontaneous event regardless of the evolutionary processes involved. This is expressed in numerous creationist or religious ideas that explain that God or another creator is responsible for the origin of species (e.g., Basel et al. 2014; Berti et al., 2010; Billingsley et al. 2016; Großschedl et al. 2014; Rissler et al. 2014; Yasri and Mancy 2014). Here, evolutionary biological knowledge is particularly important, as international studies have shown that sophisticated knowledge about evolution can positively influence the acceptance of the theory of evolution (e.g., Barnes et al. 2017; Deniz et al. 2008; Fiedler et al. 2019; Ha and Baldwin 2015). Another set of problems results from the fact that many evolutionary concepts appear to be counterintuitive to students (e.g., Tibell and Harms 2017). In this context, misconceptions

can arise in the field of phylogeny, where the deep time of evolutionary processes is not understood correctly (e.g., van Dijk and Kattmann 2010), or the origin of related species leads to problems for students when the concept of the last common ancestor cannot be discerned (e.g., Baum et al. 2005; Catley et al. 2013; Gregory 2008; Phillips et al. 2012).

Natural selection is the key mechanism of evolutionary change that leads to features adapting to new environmental conditions. The concept of natural selection is widely accepted by biologists today and can be briefly summarized: Species are adapted to their environment because individuals with the most suitable traits for that environment have a higher probability of survival and pass these traits on to their offspring. Over time, this leads to changes in the frequency of the hereditary traits of populations (Mayr 1982). This simple explanation of natural selection suggests that it would be easy to communicate, but years of research have shown that it is one of the most difficult topics to teach in biology (e.g., Bishop and Anderson 1990; Nehm and Reilly 2007). In connection with the process of natural selection, numerous misconceptions that inhibit students' understanding are described in the literature (e.g., Bishop and Anderson 1990; Ferrari and Chi 1998; Gregory 2009; Nehm et al. 2009; Shtulman 2006). These include, in particular, anthropomorphic and teleological misconceptions. Problems arise when students transfer human thinking, including emotion, motivation, and reasoning, to non-human organisms such as animals or plants. These anthropomorphic beliefs are based on the conception that the change of a trait is the result of an intentional and purposeful action performed by the individual to cope with new environmental conditions (e.g., Byrne et al. 2009; Demastes et al. 1995; Gregory 2009; Kallery and Psillos 2004; Sinatra et al. 2008; Tamir and Zohar 1991). An example of an anthropomorphic explanation is that the eagle's good eyes have developed because the eagles thought that good eyes would help them to spot the mouse from a far distance (Neubrand 2017). Here, the development of the eyes is directed solely by the individual, who judges the characteristic to be beneficial. This trait development is a singular event and does not refer to any evolutionary mechanisms. A further conceptual bias related to anthropomorphism is teleology, in which the environment itself causes traits to change over time. Teleological misconceptions always follow a "start-finish scheme" with an unchangeable final result (Stover and Mabry 2007). Here, the development of a trait is target-oriented and purposeful (e.g., Alters and Nelson 2002; Andrews et al. 2011; Beardsley 2004; Bishop and Anderson 1990; Kampourakis and Zogza 2008; Nehm et al. 2009; Nehm and Reilly 2007; Nehm and Schonfeld

Fischer *et al. Evo Edu Outreach*   (2021) 14:4

Page 4 of 16

2008; Settlage 1994; Sinatra et al. 2008). To explain the good eyes of the eagle, students could say that they have evolved in order to give the eagle an advantage when hunting (Neubrand 2017). However, the dynamics of the adaptation process of living organisms are far more complex, and underlying concepts of evolution such as the influence of randomness and probability are completely ignored in these misconceptions (e.g., Fiedler et al. 2017, 2018; Garvin-Doxas and Klymkowsky 2008).

The use of the goal- and purpose-oriented explanation of evolution is a natural human tendency and is intuitively based on humans' own personal experiences in goal- and problem-oriented thinking (Gregory 2009). This process is reinforced by the fact that pupils are often asked in science classes to explain natural phenomena causally (Olander 2012). Additionally, this tendency is supported by everyday language but also by scientific language (e.g., Alters and Nelson 2002; Nehm et al. 2010). Terms such as selection and adaptation suggest that these are directed processes that can in fact be viewed as beneficial under the current environment (e.g., Baalmann et al. 2004; Gregory 2009). However, the goal or purpose is not the determining factor for the development of a trait; instead, evolutionary biological mechanisms such as variability, selection, and inheritance are the determining factors (e.g., Godfrey-Smith 2007; Tibell and Harms 2017).

Frequently, there are problems in distinguishing between the individual and population level in evolutionary processes. The process of natural selection is based on individual traits and their interrelation with the environment. Finally, it is genetic variability that causes the differences in the phenotype. Individuals in a population therefore exhibit morphological, physiological, and behavioral differences, which can manifest themselves through generations in a population (Andersson and Wallin 2006). If the precondition that adaptation takes place on an individual level is ignored, an essentialist view can result. The essentialist misconception is characterized by the assumption that members can be assigned to a category that has an underlying "true nature" that is permanent and heritable. This true nature gives these members their basic identity (Evans 2000; Shtulman 2006). Here, differences between the evolutionary processes of populations (i.e., between-category differences) are overestimated, whereas variability at the individual level (within-category differences) is underestimated, which poses a threat to the understanding of the evolutionary theory (Opfer et al. 2012).

Learning difficulties regarding the topic of evolution have been shown in many studies (e.g., Nehm and Reilly 2007; Nehm and Schonfeld 2007; Wandersee et al. 1995) and this finding has spurred researchers to focus on identifying and addressing common misconceptions (Anderson et al. 2002). Thus, the overarching goal of biology teaching is to support students to acquire conceptually and biologically correct knowledge about evolution and to prevent misconceptions (Gregory 2009). Here, several facets of the trainee teachers' professional knowledge, which forms the basis for diagnostic activities, are necessary to evaluate scientific correctness and are essential to identify students' difficulties in understanding evolution (i.e., misconceptions) and to make adequate interventions.

Biology teachers' professional knowledge

Every day, biology teachers are confronted with diagnostic activities in the classroom. These include, for example, assessing the correctness of student answers during lessons or evaluating written performance, as shown in exams (Förtsch et al. 2018). Within the domain of professional knowledge, CK, PCK, and PK are relevant in the diagnosis of student performance (e.g., Brunner et al. 2011; Helmke et al. 2004; Kunter et al. 2013). Both CK and PCK are primarily described as content-specific facets, while PK can be considered as content-independent (e.g., Förtsch et al. 2018).

CK in general addresses the knowledge about facts and terms as well as conceptual understanding (Shulman 1986). In the context of evolution, CK primarily comprises the knowledge about the key ideas and principles of evolution. In addition, biology-specific CK includes the knowledge to determine validity within the domain (i.e., knowledge of research methods) and the knowledge about the nature of science (e.g., Großschedl et al. 2015). Several studies have shown that elaborated CK is essential for effective teaching (e.g., Baumert et al. 2010; Friedrichsen et al. 2009), but CK alone is not sufficient to enable teachers to perform diagnostic activities that lead to adaptive teaching and interventions in learning (e.g., Abell 2007; Baumert et al. 2010; Förtsch et al. 2018). CK is an important prerequisite for the development of PCK, which was defined by Shulman (1987) as a synthesis of content and pedagogy, and goes beyond subject matter knowledge. This knowledge domain is required to make the subject matter understandable. Shulman's model (1987) describes at least two facets of PCK, the knowledge about students' conceptions and preconceptions, and the knowledge about strategies to overcome them. Numerous research groups agreed with this initial conceptualization and defined the knowledge about students' understanding and the knowledge about instructional strategies for teaching as the most important facets of PCK (e.g., Förtsch et al. 2018; Grossman 1990; Hill et al. 2008; Lee and Luft 2008; Mahler et al. 2017; Park and Oliver 2008). Knowledge about student misconceptions

includes knowledge about the context in which student misconceptions occur, the context-specific categories of misconceptions, and the extent to which these misconceptions can impede the learning of scientific concepts. By anticipating these misconceptions, a teacher can plan questions to reveal this thinking and to teach in such a way that will help students to develop scientifically adequate ideas about natural selection (Ziadie and Andrews 2018). The knowledge about instructional strategies comprises knowledge on how to integrate the representation of subject matter and how to address specific learning difficulties (Großschedl et al. 2015; Hill et al. 2008; Lee and Luft 2008). Additionally, other facets of PCK have been introduced in the past, such as knowledge of the curriculum (e.g., Tamir 1988; Ziadie and Andrews 2018), knowledge of assessment methods (e.g., Hashweh 2005; Magnusson et al. 1999; Ziadie and Andrews 2018), knowledge about models (Tepner et al. 2012), or knowledge of teaching resources (Lee and Luft 2008). To diagnose whether a student has already developed a scientific concept in evolution, the teacher needs knowledge about the key ideas and principles of evolution as a facet of CK. If the student holds a misconception, the first facet of PCK, the knowledge about student understanding, is relevant. The teacher must assess the quality of the student's understanding, that is, which type of misconception is present (Förtsch et al. 2018).

In comparison to CK and PCK, facets of PK transcend the content-related areas and focus on knowledge about learning strategies, knowledge about effective classroom management, and knowledge about judgment accuracy (e.g., Brunner et al. 2011; Kunter et al. 2013). The latter refers to the ability to assess individuals appropriately (Schrader 2006). Previous research on diagnostic competence (here: knowledge about judgment accuracy; facet of PK) of teachers has shown that teachers' judgments are influenced by judgment errors, such as the halo effect, and affect the judgment accuracy (Vögelin et al. 2019). A halo effect occurs when one feature affects the judgment of another independent feature. The halo effect refers to the tendency to form an overall impression based on a prominent, dominant feature, which prevents the teacher from distinguishing between different features of performance assessment (e.g., Borman 1975; Murphy and Reynolds 1988).

Regardless of this differentiation between CK, PCK, and PK, the dichotomous classification of teacher professional knowledge into declarative ("knowledge that") and procedural ("knowledge how") knowledge has been established based on psychological approaches (e.g., Fenstermacher 1994; König et al. 2014). Declarative professional knowledge comprises factual knowledge that is accessible (explicit) to consciousness and is acquired primarily in academic discourse (Baumert and Kunter 2013). Procedural professional knowledge comprises action-oriented knowledge, which is often implicit and therefore difficult to verbalize. Through systematic practice and contextualization, procedural knowledge may develop from declarative knowledge (e.g., Schneider and Stern 2010). However, the two types of knowledge can also be unconnected or even contradictory (Shulman 1986). This reveals a problem in teacher education, namely, that the declarative knowledge acquired in teacher education often remains tacit and is not transformed into procedural knowledge (e.g., Blömeke et al. 2008; Renkl 1996). Accordingly, declarative knowledge cannot be retrieved and applied in real classroom situations. Instead, unexamined beliefs based on personal experience as a student, trainee, or teacher often determine practical action.

Research questions

The overall goal of this study, in addition to investigating the content-related facets of professional knowledge on evolution (i.e., CK and PCK), was to capture generic knowledge (i.e., PK) in order to gain deeper insights into the complex diagnostic activities of biology trainee teachers. Therefore, the selected facets of professional knowledge were operationalized with virtual student exams and transferred into a digital instrument—the Student Inventory (SI). The SI allowed us to experimentally vary different information within a virtual student exam and, based on this, to analyze trainee teachers' diagnostic activities in different facets of the trainee teachers' generic as well as biology-specific professional knowledge. Another strength of the SI is that it ensured that each trainee teacher received the accurate variation in a standardized way, resulting in high implementation fidelity. We integrated virtual student exams on evolution into the SI, which each presented a multiple-choice performance and a written answer. In order to assess the students' written answers, biology trainee teachers had to apply their knowledge about the core ideas and principles of evolution (facet of CK) and their knowledge about student understanding (facet of PCK) to assess scientific quality as well as potentially existing misconceptions. To arrive at adequate diagnoses within the virtual student exams, that is, to include only relevant information in the assessment, trainee teachers needed their knowledge about judgment accuracy (facet of PK). The facets of professional knowledge needed to assess the virtual student exams were conceptualized as procedural knowledge, because the trainee teachers had to apply their knowledge in a specific action-related situation, that is, during their assessment of the virtual student exams (Förtsch et al. 2018; Kaiser et al. 2015). Additionally, we examined

whether declarative knowledge, assessed in a short questionnaire about knowledge of evolution, influenced the procedural knowledge surveyed in the SI (see Fig. 1).

One facet of the CK knowledge domain is the knowledge about the core ideas and principles of evolution (Großschedl et al. 2015). This facet of CK had to be applied by the trainee teachers in order to assess the scientific quality of the students' written answers, which meant differentiating between scientifically correct and scientifically incorrect explanations.

> *RQ 1: To what extent are trainee teachers able to distinguish scientifically correct from scientifically incorrect students' written answers on the evolutionary process of natural selection in the SI? (i.e., knowledge about the core ideas and principles of evolution – facet of CK; procedural knowledge).*
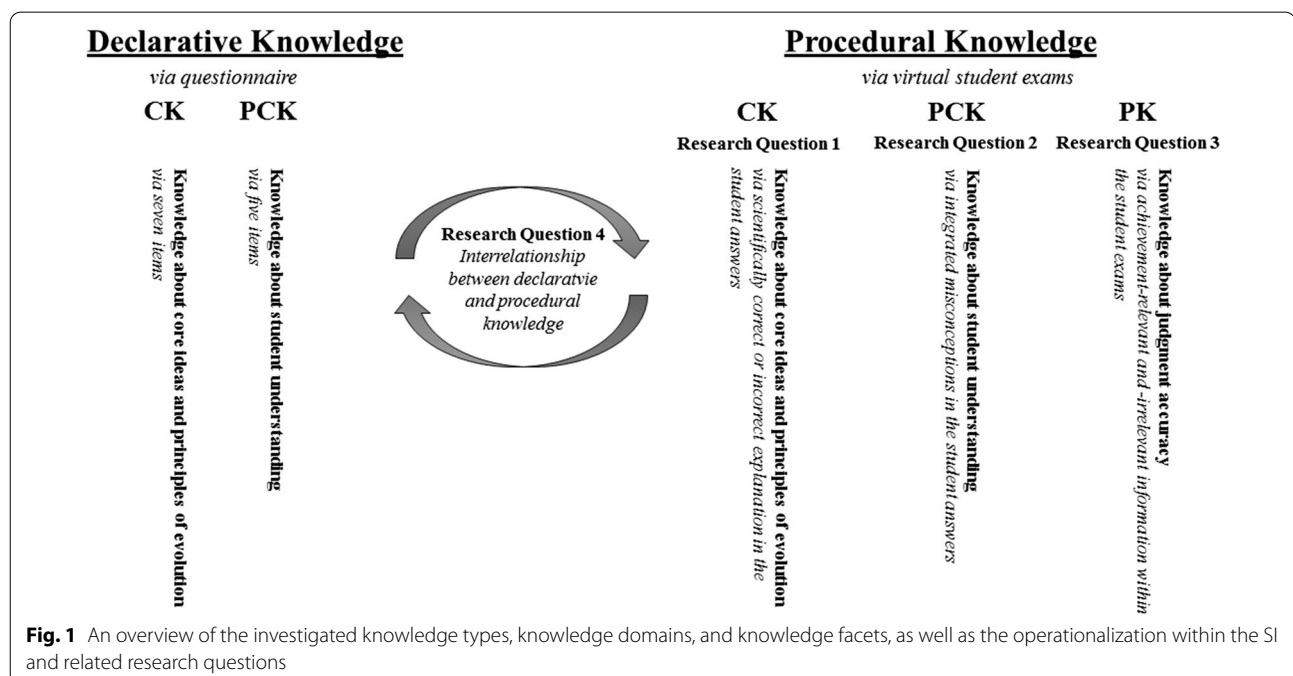
Knowledge about student understanding is, according to many studies in science education, a central facet of PCK (e.g., Förtsch et al. 2018; Hill et al. 2008; Lee and Luft 2008; Mahler et al. 2017; Park and Oliver 2008). This facet of PCK enables the trainee teachers to identify specific misconceptions about evolution in the virtual students' written answers.

> *RQ 2: To what extent are trainee teachers able to diagnose misconception categories (i.e., anthropomorphic or teleological) in the students' written answers on the evolutionary process of natural selection in the SI? (i.e., knowledge about student understanding – facet of PCK; procedural knowledge).*

Educational psychological research has described numerous judgment errors made by teachers that contribute significantly to the distortion of judgment accuracy. The knowledge about judgment accuracy (facet of PK) is therefore necessary in order for teachers to make accurate judgments (e.g., Jansen et al. 2019, 2021; Kaiser et al. 2015; Schrader 2006; Vögelin et al. 2019). Studies have already shown that teachers have problems in assessing relevant performance without including previously shown performance, which should actually be assessed independently. (Malouff and Thorsteinsson 2016; Oudman et al. 2018). Thus, the quality of an answer in a previous task within an exam can have an impact on the assessment of the quality of a subsequent answer. However, only achievement-relevant information should be considered, that is, in the present study, performance in a student's written answer on the natural selection of the peppered moth. Any influence of a previous performance in a multiple-choice test on evolution can be seen as causing bias as the previous performance is achievement-irrelevant information.

> *RQ 3: To what extent does achievement-irrelevant information (previous performance) influence the diagnosis of subsequent performance (achievement-relevant information) and lead to judgment errors among trainee teachers? (i.e., knowledge about judgment accuracy – facet of PK; procedural knowledge).*

In psychological approaches, two types of knowledge have been identified, which differ in their applicability

Fig. 1 An overview of the investigated knowledge types, knowledge domains, and knowledge facets, as well as the operationalization within the SI and related research questions

to different situations (e.g., Fenstermacher 1994; König et al. 2014). Knowledge that is necessary for answering questions in a questionnaire is operationalized as factual knowledge and is thus assigned to declarative knowledge (Baumert and Kunter 2013). However, if knowledge is involved in a specific context of action, which is the case when assessing students' written answers in the SI, it is operationalized as procedural knowledge (Schneider and Stern 2010).

> **RQ 4:** *Which first indications of the interrelationship can be observed between the trainee teachers' declarative knowledge (i.e., facets of professional knowledge, which are surveyed in a questionnaire on evolution) and their procedural knowledge (i.e., facets of professional knowledge surveyed in the SI)?*

Method
Sample
The SI was completed by 27 in-service trainee teachers ($N=27$; 22% male). In Germany, the teacher education program is divided into university education (i.e., bachelor's and master's degree; first state exam) and in-service training (second state exam). The university teacher education encompasses three and a half to 5 years and focuses on the development of CK, PCK, and PK. Within this time, there are short practical phases in schools, which last between 2 and 5 months (KMK 2014). In our study, the trainee teachers had already attended lectures and courses that explicitly teach CK, that is, knowledge about the core ideas and principles of evolution, PCK, that is, knowledge about student understanding, and PK, that is, knowledge about judgment accuracy. The in-service training takes 18 months to 2 years and encompasses the teaching of a regular school class, which is guided by mentors (Neumann et al. 2017). All of the in-service trainee teachers in our sample aspired to a teaching qualification for academic track secondary schools (i.e., Gymnasium) in the federal state of Schleswig–Holstein (Germany) and studied biology as one of their teaching subjects. On average, the trainee teachers were 29.3 ($SD=4.8$) years old.

The Student Inventory (SI)
The SI is a digital instrument that can be used with a web browser and allows a split-screen on the PC monitor. This provides a multitasking function, where, for example, virtual student exams can be read on one side and an evaluation of the student exams can simultaneously be made on the other side. Hereby, the SI differs substantially from a regular questionnaire. The SI was developed by Kaiser et al. (2015) and was initially used exclusively

for research in the field of educational psychology to measure the judgment accuracy of pre-service, trainee, or in-service teachers with regard to the assessment of student exams or separate tasks in mathematics. Therefore, the SI systematically combines achievement-relevant information with achievement-irrelevant information within student exams and investigates how these types of information influence judgment accuracy. For example, if student performance is to be accurately assessed, oral and written performances shown in the lesson may be considered relevant to achievement. Other information should not be included in an accurate performance judgment and is therefore described as being irrelevant to achievement. Many studies have already examined the influence of student characteristics that are not relevant to judgment on teachers' judgments of student performance (e.g., Schrader and Helmke 1990; Ritts et al. 1992; Ready and Wright 2011). Kaiser et al. (2015) used the SI to investigate the judgment accuracy of trainee teachers; the teachers gave the virtual students simple mathematics exercises (addition, subtraction, multiplication, and division) and received either correct or incorrect answers (information relevant to achievement). Additionally, the trainee teachers received achievement-irrelevant information about each virtual student (grade in a German test, intelligence, self-concept, family background, and gender). Kaiser et al. (2015) were able to show that the achievement-irrelevant information German grade, intelligence, and gender (female) had a positive effect on the assessment of mathematics performance and thus biased the accuracy of judgments. Based on the same theoretical framework, a study by Jansen et al. (2019, 2021) that used the SI with pre-service teachers investigated, among other things, the extent to which students' gender or an immigrant background (achievement-irrelevant information) influenced the assessment of students' English essays. Experimental variation in achievement-relevant (essay quality) and achievement-irrelevant (gender, immigrant background) information showed no effect on the assessment accuracy of the English essays. In both studies, it was the teachers' task to diagnose and evaluate the achievement-relevant information. The comparison of the real performance of the students with the performance diagnosed by the teachers provided information on whether an accurate judgment was made or a bias in the judgment occurred due to the achievement-irrelevant information (Kaiser et al. 2017). In the above-mentioned studies, the construct to be investigated was the knowledge about judgment accuracy, whereby initial aspects of content (i.e., professional judgment of English essays) were also considered (Vögelin et al. 2018).

We used the SI on the diagnostic competence of preservice, trainee, and in-service teachers, which was

developed by Kaiser et al. (2015) and used to investigate educational psychological research questions (Jansen et al. 2019, 2021). We adapted it for our purposes to measure the generic knowledge about judgment accuracy (i.e., diagnosis of achievement-relevant information; facet of PK) and, simultaneously, biology-specific facets of professional knowledge such as the knowledge about the core ideas and principles of evolution (i.e., diagnosis of scientific correctness; facet of CK) and the knowledge about student understanding (i.e., diagnosis of specific misconception categories; facet of PCK). Similar to the psychological studies, we also experimentally varied achievement-relevant with achievement-irrelevant information within virtual student exams in the SI. As achievement-relevant information, students' written answers on the natural selection of the peppered moth were integrated. The previous performance in a multiple-choice test on evolution was presented as achievement-irrelevant information. The trainee teachers' task was to assess the written answers of the virtual students without being influenced by the multiple-choice test performance previously achieved by the students.

As achievement-relevant information within the virtual student exams, six students' written answers were produced, which either were scientifically correct or expressed a specific misconception (i.e., anthropomorphic or teleological). The students' written answers are available in English as Additional file 1. The misconceptions articulated in the students' written answers were based on explanations from real students (Baalmann et al. 2004) and were modified for the SI. For reasons of homogeneity, all students' written answers were based on the Toulmin Argument Pattern (Toulmin 2003) and were tailored to the same length (115 words). The qualities of the written answers (i.e., anthropomorphic, teleological, or scientifically correct) were evaluated by three independent experts and resulted in an inter-rater agreement of 94%. According to AERA et al. (2014), this result can be interpreted as an indicator of content validity and gives an indication of the fit between the test items (i.e., the students' written answers) and the theoretical construct (i.e., the knowledge about the core ideas and principles of evolution: facet of CK; knowledge about student understanding: facet of PCK). Thus, the students' written

answers clearly expressed a specific category of misconception or a scientifically correct explanation, and we were able to assume that the students' written answers could be applied as a measure of the trainee teachers' diagnostic knowledge about the core ideas and principles of evolution (facet of CK) and knowledge about student understanding on evolution (facet of PCK). To investigate a possible judgment bias (knowledge about judgment accuracy, facet of PK; see above), students' performance on a multiple-choice test on evolution (bad or good performance) was integrated into the virtual student exams as achievement-irrelevant information (see Table 1). This performance had already been completed by the students and only needed to be noticed by the trainee teachers. The multiple-choice test on evolution was a separate task in the virtual student exams, so student performance shown in it should actually not affect the scoring of the further task (i.e., students' written answers on the natural selection of the peppered moth). The diagnostic activities on the facets of knowledge about the core ideas and principles of evolution (facet of CK), knowledge about student understanding of evolution (facet of PCK), and knowledge about judgment accuracy (facet of PK) were related to procedural knowledge according to psychological approaches (e.g., Fenstermacher 1994), because diagnostic knowledge had to be applied in an explicit situation (i.e., assessing students' exams; Förtsch et al. 2018; Kaiser et al. 2015).

Assessment of declarative and procedural knowledge with the SI

In order to get first indications about the interrelationship between the declarative and procedural knowledge of trainee teachers about the facets of professional knowledge on evolution, we additionally integrated a short questionnaire with evolution-specific questions into the SI. This questionnaire on evolution consisted of 12 items, of which seven items were assigned to the CK domain and five items to the PCK domain. The items were taken from questionnaires previously used in other studies (KiL: Kleickmann et al. 2014; ProwiE: Großschedl et al. 2015). Two translated items are available in Additional file 2. In the domain of CK, questions were asked about speciation, adaptation, and different evolutionary

Table 1 Scheme of the variation of qualities of students' written answers and of the students' performance in the multiple-choice test in the SI

Performance in the multiple-choice test	Quality of written answers		
	Teleological	Anthropomorphic	Scientifically correct
Multiple-Choice: Good Multiple-Choice: Bad	2 exams per variation (i.e., a total of 12 examinations)		

Fischer *et al. Evo Edu Outreach*   (2021) 14:4

Page 9 of 16

theories (Darwin, Lamarck). The PCK items focused on the reasons for misconceptions among students and the diagnosis of specific categories of misconceptions (e.g., anthropomorphic and teleological misconceptions). The knowledge required to answer the questions was classified as declarative knowledge because it is considered to be part of expert knowledge, which is explicit and learnable in academic discourse (Baumert and Kunter 2013). The psychometric characteristics of the questionnaire on evolution (CK, PCK) were satisfactory (Cronbach's $\alpha = 0.63$). The aim was not to analyze the separate knowledge facets of CK and PCK, but to reveal the declarative knowledge of the trainee teachers. Accordingly, we considered the entire scale of items and we operationalized the results as the declarative professional knowledge of trainee teachers on evolution integrating the respective CK and the PCK. The teachers' diagnoses of the students' written answers in the SI were operationalized as the procedural knowledge of the trainee teachers because, here, knowledge had to be applied in an explicit action-oriented situation (i.e., assessing students' written answers; Förtsch et al. 2018; Kaiser et al. 2015; Schneider and Stern 2010).

Procedure

The trainee teachers needed approximately 60 min to complete the SI. At the beginning of the survey, the trainee teachers received short instructions on the SI. Each trainee teacher received six randomly selected virtual student exams. Each exam included the student's previous performance in the multiple-choice test about evolution (achievement-irrelevant information) and a written answer by the student on the natural selection of the industrial melanism of the peppered moth (achievement-relevant information), which included a misconception or a scientifically correct way of thinking. For an overview of a virtual student exam in the SI, see Additional file 3. The performance in the previous multiple-choice test had already been assessed (good or bad multiple-choice performance) and was given to the trainee teachers without further information. Here, the trainee teachers only had to add up the points in the multiple-choice test to receive the final result for each student. If the students achieved 12 out of 20 points, this indicated a bad multiple-choice performance, whereas 19 out of 20 points indicated a good multiple-choice performance. The main task of the trainee teachers was to evaluate the students' written answers, whereby two written answers were scientifically correct and four written answers contained a misconception (i.e., two anthropomorphic and two teleological). A further task of the trainee teachers was to make sure that the students' written answers were assessed independent of the previous

performance (i.e., multiple-choice test). The evaluation of the students' written answers included a scoring (i.e., between 0 and 20 points) and the diagnosis of the quality out of a list including five options (qualities used: anthropomorphic, teleological, scientifically correct; distractors: essential, religious). After the six exams of the students had been assessed by the trainee teachers, the trainee teachers completed a questionnaire in which their knowledge about the core ideas and principles of evolution (facet of CK; seven items) and their knowledge about student understanding (facet of PCK; five items) with regard to evolution was measured. Finally, the demographic information of the trainee teachers (e.g., age, gender, course of study, teaching experience) was recorded before the study was completed.

Analyses

Within the student exams, the achievement-relevant information (i.e., quality of students' written answers) was randomly combined with the achievement-irrelevant information (i.e., previous performance in the multiple-choice test), so that each quality type of written answer was combined with each performance in the multiple-choice test. The design was a fully crossed 2×3 design. This resulted in two independent variables (IV): (1) the quality (i.e., anthropomorphic, teleological, or scientifically correct) and (2) performance in the multiple-choice test (i.e., good or bad performance). The dependent variables (DV) were: (1) the trainee teachers' content-specific knowledge in the different facets of professional knowledge (i.e., diagnosis of the scientific correctness, facet of CK; diagnosis of specific misconceptions, facet of PCK) and (2) the trainee teachers' content-independent facet of professional knowledge (i.e., assessment of achievement-relevant information, facet of PK). A total of 162 student exams were included in the analysis (i.e., 27 trainee teachers who each had to assess six student exams). Relative frequencies in the diagnosis of scientific correctness and of the misconception category provided information about the biology-specific facets of professional knowledge within the CK (refers to RQ 1) and PCK domains (refers to RQ 2). The analysis of variance provided evidence of whether students' previous performance in a multiple-choice test on evolution (achievement-irrelevant information) influenced the assessment of the students' written answers (achievement-relevant information) within the virtual student exams (refers to RQ 3). Correlation analyses were used to obtain first indications about the interrelationship between the declarative knowledge (the trainee teachers' performance in the questionnaire) and the procedural knowledge (the trainee teachers' diagnosis of the students' written answers) of trainee teachers (refers to RQ 4).

Results

Research question 1: the CK of trainee teachers

The trainee teachers were able to distinguish between scientifically correct (90.8% diagnosis rate) and scientifically incorrect (91.7% diagnosis rate) students' written answers. Accordingly, the analyses of variance revealed a significant main effect of the *quality of students' written answers* ($F[2,25] = 78.65$, $p < 0.001$, $\eta^2 = 0.863$; big effect), which means that scientifically correct written answers were scored higher than written answers with a misconception.

Research question 2: the PCK of trainee teachers

Within the scientifically incorrect students' written answers, 61.1% of the anthropomorphic misconceptions and 27.8% of the teleological misconceptions were correctly diagnosed. Overall, the trainee teachers correctly diagnosed 44.4% of the misconceptions into the respective misconception category. The results thus indicate a significant difference between the diagnosis of scientifically correct written answers and written answers that expressed a specific misconception category ($x^2[1] = 43.29$, $p < 0.001$, $\phi = 0.517$).

Research question 3: the PK of trainee teachers

The study indicates a significant main effect of the previous performance in the multiple-choice test ($F[1,26] = 5.94$, $p < 0.022$, $\eta^2 = 0.186$; small effect; see Fig. 2) on the judgment of the students' written answers. This means that scientifically correct students' written answers that were combined with a good multiple-choice performance received a higher score ($M = 18.63$, $SD = 1.80$, Min $= 15.00$, Max $= 20.00$) than students' written answers of the same quality that were combined with a bad multiple-choice performance ($M = 18.00$, $SD = 2.40$, Min $= 12.00$, Max $= 20.00$). Similar results were obtained for students' written answers with teleological misconceptions (good multiple-choice performance: $M = 9.78$, $SD = 4.73$, Min $= 0.00$, Max $= 19.00$; bad multiple-choice performance: $M = 7.81$, $SD = 3.25$, Min $= 2.00$, Max $= 14.00$). Students' written answers with anthropomorphic misconceptions received, on average, the same scores, regardless of the multiple-choice performance (good multiple-choice performance: $M = 7.19$, $SD = 4.28$, Min $= 2.00$, Max $= 16.00$; bad multiple-choice performance: $M = 7.19$, $SD = 3.94$, Min $= 0.00$, Max $= 16.00$). No interaction effect was found between previous performance in the multiple-choice test and quality of the written answers ($F[2,25] = 0.234$, $p < 0.05$, *ns*).

Research question 4: the PK of trainee teachers

The correlation between the trainee teachers' performance in the questionnaire and the diagnosis given in the CK domain (diagnosis of scientifically correct and scientifically incorrect students' written answers) revealed a strong effect ($r = 0.631$, $p < 0.001$). A moderate effect ($r = 0.499$, $p < 0.008$) was observed between the performance in the questionnaire and the diagnosis of the specific misconception category (PCK).
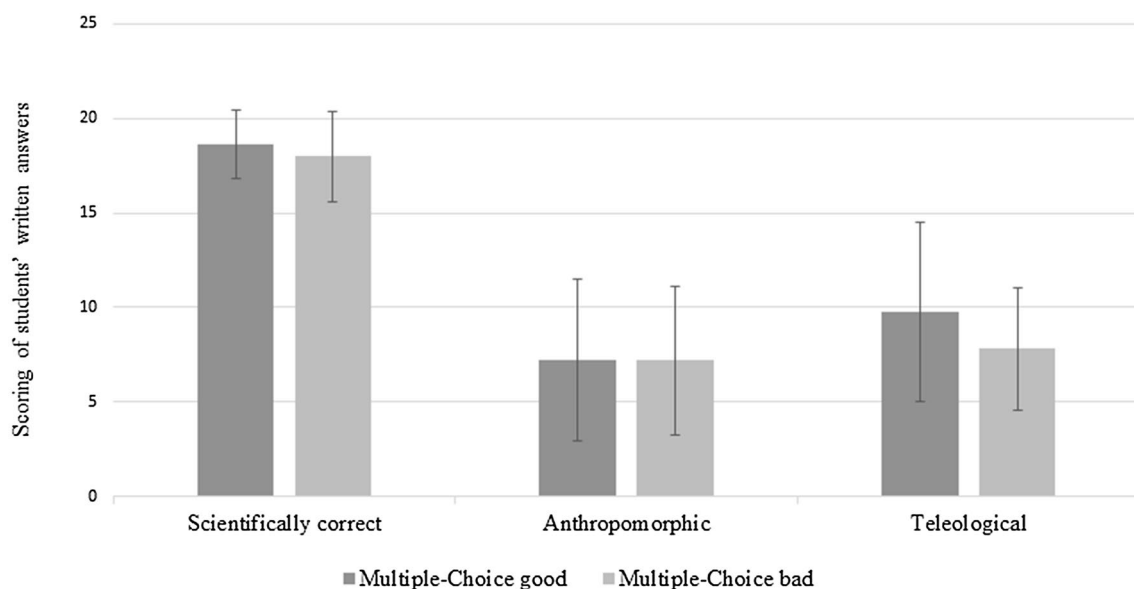

Fig. 2 Main effect of the previous performance in the multiple-choice test

Discussion

The present study examined different facets of the professional knowledge of trainee teachers, specifically their knowledge of the core ideas and principles of evolution (facet of CK), their knowledge about student understanding (facet of PCK), and their knowledge about judgment accuracy (facet of PK). The trainee teachers used a digital instrument—the SI—to assess virtual student exams on evolution in biology. We used students' biology exams to combine achievement-relevant information (i.e., students' written answers with different quality levels: scientifically correct, anthropomorphic, teleological) with achievement-irrelevant information (previous performance in a multiple-choice test on evolution: good or bad performance). This experimental setting allowed us to investigate the trainee teachers' diagnosis in the facet of CK (i.e., diagnosis of scientific correctness), the facet of PCK (i.e., diagnosis of a specific misconception), and the facet of PK (i.e., diagnosis of achievement-relevant information). Based on a questionnaire, we were able to obtain first indications about the interrelationship between declarative and procedural knowledge.

The results show that trainee teachers were able to distinguish between scientifically correct and scientifically incorrect written answers in student exams. The diagnosis rate of both quality levels was over 90%. The assessment accuracy of the scientific quality level provided an indication of the trainee teachers' CK, which includes the knowledge facet of diagnosing the core ideas and principles of evolution (Großschedl et al. 2015). Research has shown that CK alone is not sufficient to ensure adaptive learning and the learning success of students (e.g., Abell 2007; Baumert et al. 2010; Förtsch et al. 2018). Diagnosing specific misconceptions about evolution requires knowledge about student understanding (facet of PCK) from trainee teachers and can be helpful in providing insights into the origin of unexamined and scientifically incorrect student answers. Trainee teachers diagnosed 44.4% of the specific misconception categories (i.e., anthropomorphic or teleological) in the student exams. Overall, this low diagnosis rate of misconceptions reveals a lack of the important facet of PCK, the knowledge about student understanding, among trainee teachers. These findings are in line with previous research that even college students have problems in understanding evolutionary concepts (Alters and Nelson 2002). The diagnosis of teleological misconceptions (diagnosis rate: 27.8%) was significantly less frequent than that of anthropomorphic misconceptions (diagnosis rate: 61.1%). Moreover, when a student's written answer with a teleological misconception was diagnosed, trainee teachers rated this student's written answer better (i.e., a higher overall score) than a student's written answer with an anthropomorphic misconception. A study by Zohar et al. (1998) revealed that most high school students reason about biological phenomena with a mixture of teleological and causal reasoning. This reasoning is based on the tendency of humans to explain natural phenomena causally (e.g., Olander 2012). Consequently, processes that are explained on the basis of a goal- and purpose-oriented explanation are more likely to be accepted and may also be diagnosed less frequently as being scientifically incorrect by trainee teachers (e.g., Gregory 2009; Gresch and Martens 2019). The acceptance of teleological misconceptions manifests itself not only in the use of everyday language but also in unconscious actions in the classroom. This leads to teleology being an obstacle to understanding and explaining evolutionary processes (e.g., Evans et al. 2012; Kampourakis and Zogza 2008; Kelemen 2012; Sinatra et al. 2008). In contrast, trainee teachers scored students' written answers with anthropomorphic misconceptions significantly lower than students' written answers with teleological misconceptions. These findings indicate that trainee teachers consider anthropomorphic misconceptions to be more scientifically incorrect. The results of our study also confirm previous research that revealed that teachers often use teleological and anthropomorphic misconceptions to explain evolutionary processes (e.g., Kallery and Psillos 2004). The enormity of the challenge facing biologists and educators to diagnose the widespread misconception of natural selection is matched only by the importance of this task (Gregory 2009).

Independent of the articulated misconceptions, students' previous performance (i.e., performance in the multiple-choice test) influenced teachers' assessments of the students' written answers and revealed a lack of judgment accuracy (i.e., a halo effect). Thus, the students' written answers were rated higher if the corresponding performance in the previous multiple-choice test was good, indicating a judgment error (based on their PK; e.g., Südkamp et al. 2012). Achievement-irrelevant information was not directly related to performance but was often taken into account by teachers when assessing student performance, as shown in some previous studies (e.g., Ready and Wright 2011; Ritts et al. 1992; Schrader and Helmke 2001). This judgment error has also been reported in other studies that used the SI with preservice, trainee, and in-service teachers in English and mathematics (Jansen et al. 2019, 2021; Kaiser et al. 2015). To the best of our knowledge, our results are the first to provide an indication that the halo effect also occurs in trainee teachers of biology. These results are particularly relevant because they show that, even after completing university education in biology, trainee teachers with bachelor's or master's degrees are influenced by

Fischer *et al. Evo Edu Outreach*   (2021) 14:4

Page 12 of 16

achievement-irrelevant information that distorts their judgments.

The present study provides first indications of the extent to which declarative knowledge (i.e., performance in a questionnaire on evolution) and procedural knowledge (i.e., diagnosis of scientific correctness or of specific misconceptions in students' written answers on natural selection) are interrelated among trainee teachers. Our results indicate that knowledge on evolution that is gained in the academic career can be transferred into specific action-oriented situations (e.g., assessing students' written answers on natural selection) and can generate procedural knowledge (Blömeke et al. 2010). Especially in the context of the theory of evolution, it is important to focus on and promote the proceduralization of declarative knowledge in university teacher education. In the future, further digital systems could offer new opportunities to investigate and close the existing theory–practice gap (e.g., Grossman and McDonald 2008). This may help to prepare pre-service teachers for the complexity of future classroom situations and to foster procedural knowledge on evolution.

In summary, natural selection is a key mechanism of modern evolutionary theory, which – in turn – is the connecting theme of all biology topics. Without a sophisticated understanding of this process and its consequences, it is simply impossible to even remotely understand the diversity of life. Thus, professional knowledge on evolution that is conveyed in university education must be focused in order to build an adequate knowledge base among pre-service teachers and, ultimately, to support students' learning of the concept of evolution.

Limitations

The SI we used presented students' written answers that dealt with the natural selection of the peppered moth and that contained various misconceptions and scientifically correct ways of thinking. Research shows that, depending on the biological organism (bacteria, plant, animal, or human), different misconceptions have a different probability of occurrence. As we used only texts on a zoological organism in the SI, we were not able to capture these dependencies.

Numerous misconceptions that can hinder the teaching and learning of evolution have been described in scientific research (e.g., Bishop and Anderson 1990; Gregory 2009; Settlage 1994; Nehm and Schonfeld 2008). In our study we applied just two common misconceptions (i.e., anthropomorphic and teleological misconceptions) that teachers often use to explain the process of natural selection.

The SI was processed by trainee teachers who had completed their university education, which mainly focuses on the development of declarative CK, PCK, and PK, with the exception of short phases of practice in which procedural knowledge can be applied. Independent of the short practical training in the first phase of the teacher education program, the in-service teacher training program gives the trainee teachers the opportunity to gain in-depth teaching experience for the first time. Accordingly, it can be assumed that trainee teachers had not yet had much time to apply and train the declarative knowledge domains of PCK and PK in action-related situations (Blömeke et al. 2008). A sample of experienced teachers who have several years of practical experience could therefore lead to different results, which could indicate, for example, higher knowledge of student understanding (facet of PCK; Clermont et al. 1994; Grossmann 1990; Lederman et al. 1994; Schmelzing et al. 2013; van Driel et al. 2002) and generally more accurate judgments (facet of PK; Blömeke et al. 2015; Edelenbos and Kubanek-German 2004; Jansen et al. 2019, 2021).

Implications for further research and practical implications

The professional knowledge and the related diagnostic activities of pre-service and trainee teachers remain key aspects in biology education research. Education in schools as well as the university education of future teachers must place great importance on the theory of evolution in order to increase the awareness of the numerous misconceptions, as our results indicate that teachers face problems in dealing with their own misconceptions, which are actually similar to those of high school students (Demastes et al. 1995).

Misconceptions have been intensively researched (Gregory 2009). However, few studies have conceptualized which follow-up actions (e.g., cognitive conflict) are required to confront misconceptions (e.g., Demastes et al. 1995). Further research should focus on what action should be taken after the diagnosis of misconceptions. Here, the second facet of PCK, which contains knowledge about instructional strategies, is relevant. These instructional strategies can, depending on the diagnosed misconception, initiate and support the learning of the scientific concept (e.g., Ziadie and Andrews 2018). To counter students' misconceptions about evolution, the conceptual change theory, for example, is a suitable approach (Demastes et al. 1995). Here, conceptual change is described as the learning process from pre-instructional conceptions to the acquisition of scientific concepts (Heitz et al. 2010).

Initial analyses indicate a correlation between the declarative and procedural knowledge of trainee

Fischer *et al. Evo Edu Outreach*   (2021) 14:4

Page 13 of 16

teachers. However, the extent to which declarative knowledge (i.e., performance in a questionnaire) is related to procedural knowledge (i.e., action-oriented knowledge) remains unclear. Therefore, further analyses should be conducted with more extensive questionnaires (i.e., with additional items) in order to gain a deeper insight into the interrelationships between the different types of knowledge. This could be achieved by using digital instruments, such as simulated classroom environments. The results obtained from such analyses might help to reveal deficits in university teacher education and to indicate which specific contexts need to be targeted to bring declarative knowledge into action-oriented situations and to thereby generate procedural knowledge.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12052-021-00144-0.

Additional file 1. Students' written answers about the natural selection of the peppered moth.

Additional file 2. Translated items from the questionnaire to measure declarative knowledge.

Additional file 3. Screenshots of the Student Inventory.

Author details
[1] Department of Biology Education, IPN — Leibniz Institute for Science and Mathematics Education, Olshausenstrasse 62, 24118 Kiel, Germany. [2] IPL — Institute of Psychology of Learning and Instruction (IPL) at Kiel University, Olshausenstrasse 75, 24118 Kiel, Germany.

References
Abell SK. Research on science teacher knowledge. In: Abell SK, editor. Handbook of research on science education. NJ: Lawerence Erlbaum Associates; 2007. p. 1105–49.

AERA, APA, NCME. Standards for educational and psychological testing. Washington: American Educational Research Association. 2014.

Allmon WD. Why don't people think evolution is true? Implications for teaching, in and out of the classroom. Evol Educ Outreach. 2011;4(4):648–65.

Alters BJ, Nelson CE. Perspective: Teaching evolution in higher education. Evol Int J Org Evol. 2002;56(10):1891–901.

Anderson DL, Fisher KM, Norman GJ. Development and evaluation of the conceptual inventory of natural selection. J Res Sci Teach. 2002;39(10):952–78.

Andersson B, Wallin A. On developing content-oriented theories taking biological evolution as an example. Int J Sci Educ. 2006;28(6):673–95.

Andrews TM, Kalinowski ST, Leonard MJ. "Are humans evolving?" A classroom discussion to change student misconceptions regarding natural selection. Evol Educ Outreach. 2011;4(3):456–66.

Athanasiou K, Papadopoulou P. Conceptual ecology of the evolution acceptance among Greek education students: knowledge, religious practices and social influences. Int J Sci Educ. 2012;34(6):903–24.

Baalmann W, Frerichs V, Weitzel H, Gropengießer H, Kattmann U. Schülervorstellungen zu Prozessen der Anpassung-Ergebnisse einer Interviewstudie im Rahmen der Didaktischen Rekonstruktion (Students' conceptions of processes of adaptation results of an interview study in the context of didactic reconstruction). ZfDN. 2004;10(1):7–28.

Barnes ME, Evans EM, Hazel A, Brownell SE, Nesse RM. Teleological reasoning, not acceptance of evolution, impacts students' ability to learn selection. Evol Edu Outreach. 2017;10(7):1–12.

Basel N, Harms U, Prechtl H. Analysis of students' arguments on evolutionary theory. J Biol Educ. 2013;47(4):192–9.

Basel N, Harms U, Prechtl H, Weiss T, Rothgangel M. Students' arguments on the science and religion issue: the example of evolutionary theory and genesis. J Biol Educ. 2014;48(4):179–87.

Baum DA, DeWitt-Smith S, Donovan SSS. The tree-thinking challenge. Sci. 2005;310(5750):979–80.

Baumert J, Kunter M. The COACTIV model of teachers' professional competence. In: Cognitive activation in the mathematics classroom and professional competence of teachers. Boston: Springer; 2013. p. 25–48

Baumert J, Kunter M, Blu W, Brunner M, Voss T, Jordan A, Klusmann U, Kauss S, Neubrand M, Tsai YM. Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. Amer Educ Res J. 2010;47:133–80.

Beardsley PM. Middle school student learning in evolution: Are current standards achievable? Amer Biol Tea. 2004;66(9):604–13.

Berkman MB, Plutzer E. Defeating creationism in the courtroom, but not in the classroom. Sci. 2011;331(6016):404–5.

Berti AE, Toneatti L, Rosati V. Children's conceptions about the origin of species: a study of Italian children's conceptions with and without instruction. J Lear Sci. 2010;19(4):506–38.

Billingsley B, Brock R, Taber KS, Riga F. How students view the boundaries between their science and religious education concerning the origins of life and the universe. Sci Educ. 2016;100(3):459–82.

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

Blömeke S, Kaiser G. TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich. Professional competence and learning opportunities of future primary school teachers in international comparison). Münster: Waxmann; 2010.

Blömeke S, Felbrich A, Müller C, Kaiser G, Lehmann R. Effectiveness of teacher education. State of research, measurement issues and consequences for future studies. Int J Math Educ. 2008;40(5):719–34.

Blömeke S, Gustafsson JE, Shavelson RJ. Beyond dichotomies. Zeitschrift für Psychologie. 2015;223:3–13.

Borman WC. Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. J App Psy. 1975;60(5):556–60.

Bruckermann T, Fiedler D, Harms U. Identifying precursory concepts in evolution during early childhood–a systematic literature review. Stud Sci Educ. 2020;2020:1–43.

Brumby MN. Misconceptions about the concept of natural selection by medical biology students. Sci Educ. 1984;68(4):493–503.

Brunner M, Anders Y, Hachfeld A, Krauss S. Diagnostische Fähigkeiten von Mathematiklehrkräften (Diagnostic abilities of mathematics teachers). In: Kunter M, Baumert J, Blum W, Klusmann U, Krauss S, Neubrand M, editors. Professionelle Kompetenz von Lehrkräften - Ergebnisse des Forschungsprogramms COACTIV (Professional competence of teachers - results of the COACTIV research program). Münster: Waxmann; 2011. p. 215–34.

Byrne J, Grace M, Hanley P. Children's anthropomorphic and anthropocentric ideas about micro-organisms: Educational research. J Biol Educ. 2009;44(1):37–43.

Catley KM, Phillips BC, Novick LR. Snakes and eels and dogs! Oh, my! Evaluating high school students' tree-thinking skills: An entry point to understanding evolution. Res Sci Educ. 2013;43:2327–48.

Clermont CP, Borko H, Krajcik JS. Comparative study of the pedagogical content knowledge of experienced and novice chemical demonstrators. J Res Sci Teach. 1994;31:419–41.

Dagher ZR, BouJaoude S. Scientific views and religious beliefs of college students: the case of biological evolution. J Res Sci Teach. 1997;34(5):429–45.

Demastes SS, Settlage J, Good R. Students' conceptions of natural selection and its role in evolution: cases of replication and comparison. J Res Sci Teach. 1995;32(5):535–50.

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

Driver R. Changing conceptions. J Res Educ. 1988;161–98.

Edelenbos P, Kubanek-German A. Teachers assessment: the concept of diagnostic competence. Language Testing. 2004;21(3):259–83.

Endler JA. Natural selection in the wild Monographs in population biology, vol. 21. Princeton: Princeton University Press; 1986.

Evans EM. The emergence of beliefs about the origins of species in school-age children. Merr-Palm Quart. 2000;46(2):221–54.

Evans EM, Rosengren K, Lane JD, Price KLS. Encountering counterintuitive ideas: constructing a developmental learning progression for evolution understanding. In: Rosengren K, Brem SK, Evans EM, Sinatra GM, editors. Evolution challenges: Integrating research and practice in teaching and learning about evolution. New York: Oxford University Press; 2012. p. 174–99.

Fenstermacher GD. The knower and the known: the nature of knowledge in research on teaching. Rev Res Educ. 1994;20(1):3–56.

Ferrari M, Chi MT. The nature of naïve explanations of natural selection. Int J Sci Educ. 1998;20(10):1231–56.

Fiedler D, Tröbst S, Harms U. University students' conceptual knowledge of randomness and probability in the contexts of evolution and mathematics. CBE Life Sci Educ. 2017;16(2):1–16.

Fiedler D, Tröbst S, Großschedl J, Harms U. EvoSketch: Simple simulations for learning random and probabilistic processes in evolution, and effects of instructional support on learners' conceptual knowledge. Evol Educ Outreach. 2018;11(15):1–17.

Fiedler D, Sbeglia GC, Nehm RH, Harms U. How strongly does statistical reasoning influence knowledge and acceptance of evolution? J Res Sci Teach. 2019;56(9):1183–206.

Förtsch C, Sommerhoff D, Fischer F, Fischer M, Girwidz R, Obersteiner A, Reiss K, Stürmer K, Siebeck M, Schmidmaier R, Seidel T, Ufer S, Wecker C, Neuhaus BJ. Systematizing professional knowledge of medical doctors and teachers: development of an interdisciplinary framework in the context of diagnostic competences. Educ Sci. 2018;8(4):207–25.

Friedrichsen PJ, Abdell S, Pareja E, Brown P, Lankford D, Volkmann M. Does teaching experience matter? Examining biology teachers' prior knowledge for teaching in an alternative certification program. J Res Sci Teach. 2009;46(4):357–83.

Furtak EM. Linking a learning progression for natural selection to teachers' enactment of formative assessment. J Res Sci Teach. 2012;49(9):1181–210.

Garvin-Doxas K, Klymkowsky MW. Understanding randomness and its impact on student learning: lessons learned from building the Biology Concept Inventory (BCI). CBE Life Sci Educ. 2008;7(2):222–37.

German National Academy of Sciences Leopoldina. Evolutionsbiologische Bildung in Schule und Universität (Evolutionary biology education in schools and universities). Germany: Halle; 2017.

Godfrey-Smith P. Conditions for evolution by natural selection. J Philos. 2007;104(10):489–516.

Gregory TR. Understanding evolutionary trees. Evol Educ Outreach. 2008;1(2):121–37.

Gregory TR. Understanding Natural selection. Essential concepts and common misconceptions. Evol Educ Outreach. 2009;2(2):156–75.

Gresch H, Martens M. Teleology as a tacit dimension of teaching and learning evolution: a sociological approach to classroom interaction in science education. J Res Sci Teach. 2019;56(3):243–69.

Grossman PL, McDonald M. Back to the future: directions for research in teaching and teacher education. Amer Educ Res J. 2008;45(1):184–205.

Grossmann PL. The making of a teacher: Teacher knowledge and teacher education. New York: Teacher College Press; 1990.

Großschedl J, Konnemann C, Basel N. Pre-service biology teachers' acceptance of evolutionary theory and their preference for its teaching. Evol Educ Outreach. 2014;7(18):1–16.

Großschedl J, Harms U, Kleickmann T, Glowinski I. Preservice biology teachers' professional knowledge: structure and learning opportunities. J Sci Teach Educ. 2015;26(3):291–318.

Großschedl J, Seredszus F, Harms U. Angehende Biologielehrkräfte: evolutionsbezogenes Wissen und Akzeptanz der Evolutionstheorie (Prospective biology teachers: evolution-related knowledge and acceptance of the theory of evolution). ZfDN. 2018;24(1):51–70.

Ha M, Baldwin BC. Nehm RH. The long-term impacts of short-term professional development: Science teachers and evolution. Evol Edu Outreach; 2015. p. 1–23.

Ha M, Haury DL, Nehm RH. Feeling of certainty: Uncovering a missing link between knowledge and acceptance of evolution. J Res Sci Teach. 2012;49(1):95–121.

Harms U, Reiss MJ. Evolution Education Re-considered: Understanding What Works. Basel: Springer; 2019.

Hashweh MZ. Teacher pedagogical constructions: A reconfiguration of pedagogical content knowledge. Teachers Teaching. 2005;11:273–92.

Heitz JG, Cheetham JA, Capes EM, Jeanne RL. Interactive evolution modules promote conceptual change. Evol Educ Outreach. 2010;3(3):436–42.

Helmke A, Hosenfeld I, Schrader FW. Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften (Comparative studies as an instrument for improving the diagnostic competence of teachers). In: Arnold R, Griese C, editors. Schulleitung und Schulentwicklung. Hohengehren: Schneider-Verlag; 2004. p. 119–44.

Hill HC, Ball DL, Schilling SG. Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. J Res Math Educ. 2008;39:372–400.

Jansen T, Vögelin C, Machts N, Keller S, Möller J. Das Schülerinventar ASSET zur Beurteilung von Schülerarbeiten im Fach Englisch: Drei experimentelle Studien zu Effekten der Textqualität und der Schülernamen (The student inventory ASSET for evaluating student work in English: Three experimental studies on effects of text quality and student names). Psychologie in Erziehung und Unterricht. 2019;66(4):303–15.

Jansen T, Vögelin C, Machts N, Keller S, Köller O, Möller J. Judgment accuracy in experienced versus student teachers: Assessing essays in English as a foreign language. Teach Teacher Educ. 2021;97:1–11.

Kaiser J, Möller J, Helm F, Kunter M. Das Schülerinventar: Welche Schülermerkmale die Leistungsurteile von Lehrkräften beeinflussen (The Student Inventory: What Student Characteristics Influence Teacher Performance Judgments). Zeitschrift für Erziehungswissenschaft. 2015;18(2):279–302.

Kaiser J, Südkamp A, Möller J. The effects of student characteristics on teachers' judgment accuracy: Disentangling ethnicity, minority status, and achievement. J Educ Psychol. 2017;109(6):871–88.

Kallery M, Psillos D. Anthropomorphism and animism in early years science: Why teachers use them, how they conceptualise them and what are their views on their use. Res Sci Educ. 2014;34(3):291–311.

Kampourakis K, Nehm RH. History and philosophy of science and the teaching of evolution: Students' conceptions and explanations. International Handbook of Research in History, Philosophy and Science Teaching. Dordrecht: Springer; 2014. p. 377–99.

Kampourakis K, Zogza V. Students' intuitive explanations of the causes of homologies and adaptations. Sci Educ. 2008;17(1):27–47.

Kelemen D. Teleological minds: How natural intuitions about agency and purpose influence learning about evolution. In: Rosengren K, Evans EM, editors. Evolution challenges: Integrating research and practice in teaching and learning about evolution. England: Oxford University Press; 2012. p. 66–92.

King CJH. An analysis of misconceptions in science textbooks: Earth science in England and Wales. Intl J Sci Educ. 2010;32(5):565–601.

Kleickmann T, Großschedl J, Harms U, Heinze A, Herzog S, Hohenstein F, Köller O, Kröger J, Lindmeier A, Loch C, Mahler D, Möller J, Neumann K, Parchmann I, Steffensky M, Taskin V, Zimmermann F. Professionswissen von Lehramtsstudierenden der mathematisch-naturwissenschaftlichen Fächer – Testentwicklung im Rahmen des Projekts KiL (Professional knowledge of pre-service teachers in mathematics and science - Test development within the framework of the KiL project). Unterrichtswissenschaft. 2014;42:280–8.

König J, Blömeke S, Klein P, Suhl U, Busse A, Kaiser G. Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. Teach Teacher Educ. 2014;38:76–88.

Kunter M, Klusmann U, Baumert J, Richter D, Voss T, Hachfeld A. Professional competence of teachers: effects on instructional quality and student development. J Educ Psychol. 2013;105(3):805–20.

Lead States NGSS, [NGSS], . Next generation science standards: for states, by states. Washington, DC: The National Academies Press; 2013.

Lederman NG, Gess-Newsome J, Latz MS. The natur and development of pre-service science teachers' conceptions of subject matter and pedagogy. J Res Sci Teach. 1994;31:129–46.

Lee E, Luft JA. Experienced secondary science teachers' representation of pedagogical content knowledge. Int J Sci Educ. 2008;30:1343–63.

Magnusson S, Krajcik J, Borko H. Nature, sources, and development of pedagogical content knowledge for science teaching. In: Examining pedagogical content knowledge. Dodrecht: Springer; 1999. p. 95–132.

Mahler D, Großschedl J, Harms U. Using doubly latent multilevel analysis to elucidate relationships between science teachers' professional knowledge and students' performance. Int J Sci Educ. 2017;39(2):213–37.

Malouff JM, Thorsteinsson EB. Bias in grading: A meta-analysis of experimental research findings. Australian J Educ. 2016;60(3):245–56.

Mayr E. The growth of biological thought: Diversity, evolution and inheritance. Cambridge: Harvard University Press; 1982.

McVaugh NK, Birchfield J, Lucero MM, Petrosino AJ. Evolution education: Seeing the forest for the trees and focusing our efforts on the teaching of evolution. Evol Educ Outreach. 2011;4(2):286–92.

Moharreri K, Ha M, Nehm R. EvoGrader: an online formative assessment tool for automatically evaluating written evolutionary explanations. Evol Edu Outreach. 2014;7(1):15.

Murphy KR, Reynolds DH. Does true halo affect observed halo? J Appl Psychol. 1988;73(2):235–8.

Nadelson LS, Sinatra GM. Educational professionals' knowledge and acceptance of evolution. Evol Psychol. 2009;7(4):490–516.

National Research Council. Thinking evolutionarily: Evolution education across the life sciences: Summary of a convocation. Washington, DC: The National Academies Press; 2012.

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

Nehm RH, Schonfeld IS. Does increasing biology teacher knowledge of evolution and the nature of science lead to greater preference for the teaching of evolution in schools? J Sci Teach Educ. 2007;18(5):699–723.

Nehm RH, Schonfeld IS. Measuring knowledge of natural selection: A comparison of the CINS, an open-response instrument, and an oral interview. J Res Sci Teach. 2008;45(10):1131–60.

Nehm RH, Poole TM, Lyford ME, Hoskins SG, Carruth L, Ewers BE, Colberg PJ. Does the segregation of evolution in biology textbooks and introductory courses reinforce students' faulty mental models of biology and evolution? Evol Educ Outreach. 2009;2(3):527–32.

Nehm RH, Rector MA, Ha M. "Force-Talk" in evolutionary explanation: Metaphors and misconceptions. Evol Educ Outreach. 2010;3:605–13.

Neubrand C. Untersuchung von Aspekten erfolgreichen Lehrens und Lernens mit Fokus auf Beispielaufgaben und vorwissensangepasste Impulse (Investigation of aspects of successful teaching and learning with focus on example tasks and pre-knowledge-adapted impulses) [Unpublished doctoral dissertation]. Kiel: Leibniz Institute for Science and Mathematics Education; 2017.

Neumann K, Härtig H, Harms U, Parchmann I. Science Teacher Preparation in Germany: An International Comparison of What Works Best. In: Pederson J, Isozaki T, Hirano T, editors. Science Teacher Education: A Multi Country Comparison. Charlotte, NC: Information Age Publishing Inc; 2017. p. 29–52.

Olander C. Teaching biological evolution-internal and external evaluation of learning outcomes. Nord Stud Sci Educ. 2012;5(2):171–84.

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

Oudman S, van de Pol J, Bakker A, Moerbeek M, van Gog T. Effects of different cue types on the accuracy of primary school teachers' judgments of students' mathematical understanding. Teaching Teacher Educ. 2018;76:214–26.

Park S, Oliver JS. Revisiting the conceptualization of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. Res Sci Educ. 2008;38:261–84.

Pazza R, Penteado PR, Kavalco KF. Misconceptions about evolution in Brazilian freshmen students. Evol Educ Outreach. 2010;3(1):107–13.

Phillips BC, Novick LR, Catley KM, Funk DJ. Teaching tree thinking to college students: It's not as easy as you think. Evol Educ Outreach. 2012;5(4):595–602.

Ready DD, Wright DL. Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. Ameri Educ Res J. 2011;48(2):335–60.

Rector MA, Nehm RH, Pearl D. Learning the language of evolution: lexical ambiguity and word meaning in student explanations. Res Sci Educ. 2013;43(3):1107–33.

Renkl A. Träges Wissen: Wenn Erlerntes nicht genutzt wird (Tacit knowledge: If learned knowledge is not used). Psychologische Rundschau. 1996;47:78–92.

Rissler LJ, Duncan SI, Caruso NM. The relative importance of religion and education on university students' views of evolution in the Deep South and state science standards across the United States. Evol Educ Outreach. 2014;7(1):7–24.

Ritts V, Patterson ML, Tubbs ME. Expectations, impressions, and judgments of physically attractive students: A review. Rev Educ Res. 1992;62(4):413–26.

Schmelzing S, van Driel JH, Jüttner M, Brandenbusch S, Sandmann A, Neuhaus BJ. Development, evaluation, and validation of a paper-and-pencil test for measuring two components of biology teachers' pedagogical content knowledge concerning the 'cardiovascular system.' Int J Sci Math Edu. 2013;11:1369–90.

Schneider M, Stern E. The developmental relations between conceptual and procedural knowledge: A multimethod approach. Develop Psychol. 2010;46(1):178–92.

Schrader FW. Diagnostische Kompetenz von Eltern und Lehrern (Diagnostic competence of parents and teachers). In: Rost DH, editor. Handwörterbuch Pädagogische Psychologie. Weinheim: Beltz; 2006. p. 95–100.

Schrader FW, Helmke A. Alltägliche Leistungsbeurteilung durch Lehrer (Day-to-day performance evaluation by teachers). Leistungsmessungen in Schulen. 2001;2:45–58.

Settlage J. Conceptions of natural selection: A snapshot of the sense-making process. J Res Sci Teach. 1994;31(5):449–57.

Shtulman A. Qualitative differences between naïve and scientific theories of evolution. Cog Psychol. 2006;52(2):170–94.

Shulman LS. Those who understand: Knowledge growth in teaching. Educ Res. 1986;15(2):4–14.

Shulman LS. Knowledge and teaching: Foundations of the new reform. Harvard Educ Rev. 1987;57(1):1–22.

Sickel AJ, Friedrichsen P. Examining the evolution education literature with a focus on teachers: major findings, goals for teacher preparation, and directions for future research. Evol Educ Outreach. 2013;6(1):6–23.

Sinatra GM, Brem SK, Evans EM. Changing minds? Implications of conceptual change for teaching and learning about biological evolution. Evol Educ Outreach. 2008;1(2):189–95.

Standing Conference of the Ministers of Education and Cultural Affairs of the Federal States in the Federal Republic of Germany [KMK]. Bildungs-standards im Fach Biologie für den Mittleren Schulabschluss (Beschluss der Kultusministerkonferenz vom 16.12.2004) (Science standards for middle school biology). Munich: Luchterhand; 2005.

Standing Conference of the Ministers of Education and Cultural Affairs of the Federal States in the Federal Republic of Germany [KMK]. Standards für die Lehrerbildung Bildungswissenschaften Munich: Luchterhand; 2014.

Stover SK, Mabry ML. Influences of Teleological and Lamarckian Thinking on Student Understanding of Natural Selection. Bioscience. 2007;33(1):11–8.

Südkamp A, Kaiser J, Möller J. Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. J Educ Psychol. 2012;104(3):743–62.

Tamir P. Subject matter and related pedagogical knowledge in teacher education. Teaching and Teacher Education. 1988;4(2):99–110.

Tamir P, Zohar A. Anthropomorphism and teleology in reasoning about biological phenomena. Sci Educ. 1991;75(1):57–67.

Tepner O, Borowski A, Fischer HE, Jüttner M, Kirschner S, Leutner D. Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften (Model for the development of test items for the assessment of professional knowledge of science teachers). Zeitschrift für Didaktik der Naturwissenschaft. 2012;18:7–28.

Tibell LA, Harms U. Biological principles and threshold concepts for understanding natural selection. Sci Educ. 2017;26(7):953–73.

Toulmin SE. The uses of argument. Cambridge. UK: Cambridge University Press; 2003.

van Dijk EM, Kattmann U. Evolution im Unterricht: Eine Studie über fachdidaktisches Wissen von Lehrerinnen und Lehrern (Evolution in teaching: a study of the didactic knowledge of teachers). ZfDN. 2010;16:7–21.

van Driel JH, De Jong O, Verloop N. The development of preservice chemistry teachers' pedagogical content knowledge. Sci Edu. 2002;86:572–90.

Vögelin C, Jansen T, Keller S, Machts N, Möller J. The influence of lexical features on teacher judgments of ESL argumentative essays. Assess Writing. 2019;39:50–63.

Wandersee JH, Good RG, Demastes SS. Forschung zum Unterricht über Evolution: Eine Bestandsaufnahme (Research on the teaching of evolution: an inventory). ZfDN. 1995;1(1):43–54.

Yasri P, Mancy R. Understanding student approaches to learning evolution in the context of their perceptions of the relationship between science and religion. Int J Sci Educ. 2014;36(1):24–45.

Yip DY. Identification of misconceptions in novice biology teachers and remedial strategies for improving biology learning. Int J Sci Educ. 1998;20(4):461–77.

Zabel J, Gropengiesser H. Learning progress in evolution theory: climbing a ladder or roaming a landscape? J Biol Educ. 2011;45(3):143–9.

Ziadie MA, Andrews TC. Moving evolution education forward: a systematic analysis of literature to identify gaps in collective knowledge for teaching. CBE Life Sci Educ. 2018;17(1):1–10.

Zohar A, Ginossar S. Lifting the taboo regarding teleology and anthropomorphism in biology education—heretical suggestions. Sci Educ. 1998;82(6):679–97.

Publisher's Note