Received: 2 November 2021 | Revised: 14 April 2022 | Accepted: 19 June 2022
DOI: 10.1111/jcal.12718

ARTICLE

Journal of Computer Assisted Learning WILEY

The Simulated Classroom Biology—A simulated classroom environment for capturing the action-oriented professional knowledge of pre-service teachers about evolution

Julian Fischer[1] | Nils Machts[2] | Till Bruckermann[1,3] | Jens Möller[2] | Ute Harms[1]

[1]Department of Biology Education, IPN – Leibniz Institute for Science and Mathematics Education, Kiel, Germany
[2]IPL – Institute of Psychology of Learning and Instruction (IPL) at Kiel University, Kiel, Germany
[3]Institute of Education at Leibniz University Hannover, Hannover, Germany

Correspondence
Julian Fischer, Department of Biology Education, IPN – Leibniz Institute for Science and Mathematics Education, Olshausenstrasse 62, 24118 Kiel, Germany.
Email: jfischer@leibniz-ipn.de

Abstract

Background: The professional knowledge of pre-service teachers is highly important for effective and successful teaching. In recent years, many research groups have been engaged in developing simulated classroom environments to capture especially the pedagogical knowledge (PK) of pre-service teachers, neglecting the content-related facets of professional knowledge such as pedagogical content knowledge (PCK).

Objectives: In the present study, we describe the development of a simulated classroom environment—the Simulated Classroom Biology (SCR^{Bio})—and provide evidence regarding its validity to assess pre-service biology teachers' action-oriented PCK in the area of evolution.

Methods: This study examined the evidence supporting the validity of using the SCR^{Bio} to investigate action-oriented PCK of pre-service biology teachers. The (1) evidence based on test content (expert ratings) and the (2) evidence based on relation to other variables (known-groups comparison) was obtained. We tested the SCR^{Bio} with $N = 76$ German pre-service biology teachers.

Results and Conclusions: Our results show the successfully operationalized PCK in the SCR^{Bio} through explicit allocation of specific misconceptions to each virtual student's answer and the valid measurement of pre-service biology teachers' action-oriented PCK. This results in a validated simulated classroom environment for pre-service but also in-service teachers. In the future, the SCR^{Bio} will be developed from an assessment instrument to a training tool to simulate explicit teaching situations. This allows to complement the predominantly theoretical components of university-based teacher education with practice-based simulated classroom environments.

KEYWORDS
evolution, professional knowledge, simulation, teacher education

1 | INTRODUCTION

Teachers strongly influence student achievement. Many scholars have shown that professional knowledge, the core of professional competence, is the key element for effective and successful teaching (e.g., Abell, 2007; Hattie, 2009). Based on Shulman (1987), researchers generally agree that content knowledge (CK), pedagogical content knowledge (PCK) and pedagogical knowledge (PK) are essential domains of professional knowledge which influence teaching and learning (e.g., Abell, 2007; Baumert & Kunter, 2011). Within the conceptualization of professional knowledge in science education, Shulman (1986) stressed the relevance of PCK for creating learning opportunities and initiating student learning processes. Consequently, there has been a lot of research on what PCK is (e.g., Magnusson et al., 1999; Park & Oliver, 2008), how it differs from other knowledge domains (e.g., Großschedl et al., 2014; Großschedl et al., 2015; Mahler et al., 2017), and how it can be measured and supported (e.g., Henze & van Driel, 2015). Regarding PCK, knowledge about student understanding is essential for the diagnosis of students' misconceptions in the science teaching process and necessary for future students' learning success (Förtsch et al., 2018). Recently, a new conceptualization of PCK (Refined Consensus Model of PCK) was introduced by distinguishing different realms of PCK, of which the enacted PCK (ePCK) represents the specific knowledge and skills of teachers in particular teaching situations (Carlson & Daehler, 2019). Thus, the ePCK can be defined and described as a PCK close to action—the action-oriented PCK.

Previous studies on PCK lacked a focus on its application in teaching situations in the sense of capturing ePCK. The methodology of the most quantitative studies which record PCK through questionnaires are not suitable to measure ePCK (e.g., Großschedl et al., 2019; Kleickmann et al., 2013). Qualitative approaches (for an overview see Depaepe et al., 2015), such as interviews (e.g., Jüttner & Neuhaus, 2013; Rozenszajn & Yarden, 2014), video and audio recordings (e.g., Park & Oliver, 2008) could address this problem and provide a deeper insight into ePCK. However, these methodologies often have small sample sizes and are very time-consuming.

Simulated classroom environments might be a solution for this problem, as they can capture PCK in action-oriented settings and can efficiently collect large samples simultaneously (e.g., Dalgarno et al., 2016). At the same time, this is a central challenge of research to develop and provide action-oriented settings in which ePCK of pre-service teachers can effectively be captured. Furthermore, the user is confronted with a digital ecology that can provide a highly engaging and realistic interactive science learning experience (e.g., Fiedler et al., 2002; Plass et al., 2009). To date, most simulated classroom environments are developed to measure the teachers' PK in action (e.g., The Simulated Classroom: Fiedler et al., 2002; Fiedler et al., 2007; Kaiser et al., 2013; Kaiser et al., 2017; Südkamp et al., 2008), for example classroom management and diagnostic competence, but they neglect content-related domains (i.e., PCK, CK) of pre-service teachers' professional knowledge. Respective studies show that simulated classroom environments are suitable instruments for the assessment of action-oriented PK in teachers.

From these generic studies we conclude that simulated classroom environments represent appropriate tools to capture action-orientated, content-related professional knowledge (i.e., ePCK, CK) of pre-service teachers. To test this assumption, we developed the Simulated Classroom Biology (SCR^{Bio}), a tool measuring (pre-service) teachers' content-related professional knowledge, that is, ePCK and CK focusing on the central biological content evolution. Furthermore, validity evidence based on the test content and on known-groups comparisons is needed to conclude that the SCR^{Bio} actually captures pre-service biology teachers' knowledge about student understanding in an action-oriented setting (ePCK).

2 | THEORETICAL BACKGROUND

2.1 | Teachers' professional knowledge

Professional knowledge is described as the core of professional behaviour, a prerequisite for good teaching and predictive of student outcomes (e.g., Baumert & Kunter, 2006). Shulman's (1986) conceptualization of professional knowledge as CK, PCK, and PK has recently served as a framework for several large-scale studies that assessed and described professional knowledge in science (e.g., *ProwiN*: Fischer et al., 2012; *KiL*: Kleickmann et al., 2013) and in mathematics (e.g., *COACTIV*: Kunter et al., 2013). CK and PCK are mainly understood as the content-related domains of professional knowledge, whereas PK can be seen as a non-content-related (e.g., Förtsch et al., 2018; Jüttner et al., 2013; Mahler et al., 2017).

In general, CK refers to knowledge of facts, terms and the conceptual understanding of these contents (Shulman, 1986) and is essential for effective teaching (e.g., Ball et al., 2001; Baumert et al., 2010; Friedrichsen et al., 2009) and student performance (Sadler et al., 2013). In general, the PCK domain comprises two facets. The first facet describes knowledge that the teacher uses to formulate and represent the subject matter in an intelligible way (Shulman, 1986) and is summarized as knowledge about instructional strategies. The second facet of PCK focuses on knowledge about students' understanding, which is mainly operationalized through the diagnosis of student misconceptions (Förtsch et al., 2018). In addition, individual research groups have added further facets, which are covered by the PCK, such as curricular knowledge and knowledge of assessment methods (e.g., Tamir, 1988). In contrast to the CK and PCK domain, the PK domain is not content-related and covers several facets, such as general knowledge (Kunter et al., 2013) and has a focus on general aspects of teaching (Shulman, 1987). Regarding PK, teachers' judgement accuracy is key and describes the relationship between teachers' judgements of students' achievement and actual student performance (Südkamp et al., 2012). For example, a student in biology class may answer some questions correctly or incorrectly. Here the teacher needs biological CK and PCK to make a diagnosis.

The assessment of the student's performance and student's participation during the lesson requires teacher PK. The judgement accuracy can be measured by three components, (1) the ranking component, which shows whether the teacher can correctly assess the characteristic expression of students, (2) the level component, which indicates the teachers' ability to correctly assess the average level of a student characteristic, and the (3) differentiation component, which shows the ratio of real performance to assessed performance (Südkamp et al., 2012).

A fundamental model of professional competence in the German-speaking countries is developed in the COACTIV project and includes cognitive (e.g., professional knowledge) as well as affectively motivational (e.g., self-regulation, motivational orientation) aspects (Kunter et al., 2011). However, this conceptualization neglects the action-oriented facets which Blömeke et al. (2015) integrated into a new model. The authors describe the observable teaching action of a person as shown performance. For a long time there was no differentiation between tacit and action-oriented knowledge within professional knowledge in science education, whereas following psychological approaches, types of knowledge include declarative knowledge (knowing what) and action-related knowledge, consisting of procedural knowledge (knowing how) and conditional knowledge (knowing when and why) (Paris et al., 1983). In 2017, the Refined Consensus Model of PCK (RCM: Carlson & Daehler, 2019) advanced among others, the framework of Shulman (1986) and the Magnusson Model (Magnusson et al., 1999). The RCM provides researchers with a differentiated analysis of teachers' PCK that goes beyond static knowledge and focuses on dynamic application in action-orientated situations. A core characteristic of this model is the differentiation of three distinct realms of PCK—the collective PCK (cPCK), the personal PCK (pPCK), and the enacted PCK (ePCK)—ranging from the professional knowledge of the community of educators in a field (cPCK), the personal professional knowledge of individual science teachers (pPCK) to specific knowledge and skills of teachers in particular teaching situations (ePCK). Still, the model accounts for the domains of teachers' professional knowledge (i.e., content knowledge, pedagogical knowledge, knowledge of students, curricular knowledge, and assessment knowledge) from previous models (e.g., Magnusson et al., 1999; Shulman, 1987). The categorization of the PCK into three different knowledge realms (cPCK, pPCK, and ePCK) provides a clear theoretical framework for the research within PCK (Carlson & Daehler, 2019). The RCM facilitates the assessment of different PCK domains—including the knowledge of students' understanding—in an action-oriented framework (ePCK). Enabling pre-service teachers to apply their static knowledge in specific teaching situations and measuring ePCK simulated classroom environments as action-oriented settings might be a successful approach (Hixon & So, 2009).

2.2 | Simulated classroom environments

Simulated learning environments have recently become increasingly popular in the context of teacher education (Bradley & Kendall, 2014). Simulated classrooms provide solutions to bridge the gap between real teaching situations in the classroom at school and theory-bound teacher education at university (Hixon & So, 2009). A simulation is a simplified but valid, accurate, and dynamic model of reality implemented in a digital system (Sauvé et al., 2007). New digital technologies have motivated many research groups to design classroom simulations and provide them as a tool for observing pre-service teachers' application of their professional knowledge under controlled conditions (e.g., Cheong, 2010; Gurvitch & Metzler, 2009). Most classroom simulations in teacher education focus on the facets of PK, including classroom management, self-efficacy, and the professional interpersonal competence (e.g., *TeachME™*: Bautista & Boone, 2015; *simSchool*: Christensen et al., 2011; *VirtualPREX*: Dalgarno et al., 2016; *CS-TGCTS*: Yeh, 2004), and neglect the content-related facets like PCK. Here, the computer-based simulated classrooms facilitate remote access role-plays of the pre-service teacher and virtual students, in which the teacher faces real, classroom-specific problems (Clapper, 2010). In a role-play simulation the user slips into the role of the individuals involved in teaching (teacher or student) and adopts their tasks in this simulated scenario (Veletsianos et al., 2010). Generally, role-play simulations can be distinguished into the non-immersive simulation which uses texts and images to create a classroom situation and the immersive simulation which provides visual representations of a classroom including the application of avatars (Dalgarno et al., 2016). The realistic integration of the participants through 3-D representations in the classroom by a role-playing avatar enables them to identify more closely with their assigned teacher role (Jamaludin et al., 2009). In contrast, non-immersive classroom simulations have the advantage of providing complexity reduced environments under controllable conditions (e.g., Fiedler et al., 2007; Südkamp et al., 2008).

Non-immersive simulation, like the Simulated Classroom (SCR) is a suitable tool to illustrate classroom interaction (i.e., role-play between teacher and students) in an action-oriented setting. Here, the pre-service teachers slip into the role of the teacher and can pose questions to the virtual students in a virtual lesson. Additionally, the SCR can capture different facets of professional knowledge, which must be applied by the teacher to achieve an accurate assessment of the virtual students. Initially, the SCR was a computer-based tool to measure action-oriented diagnostic knowledge for the PK domain in an experimental environment on judgement biases such as teachers' error of central tendency and naïve theories about the behaviour of individuals and stereotypes (Fiedler et al., 2002; Fiedler et al., 2007). Südkamp et al. (2008) expanded the existing SCR in order to handle more complex educational psychological contents. Now the researchers could integrate their own tasks, questions and answers into the system, which in turn increased the ecological validity of the instrument. This made it possible to analyse the diagnostic judgements of pre-service teachers in a more differentiated way. Recently, the SCR was used to disentangle effects of virtual students' achievement and motivation on teachers' assessments of both constructs (Kaiser et al., 2013) and to analyse how minority status impacts teachers' assessments (Kaiser et al., 2017).

Whilst in educational psychology teachers' accuracy of judgements, based on the differentiation, ranking, and level-component (Fiedler et al., 2002; Fiedler et al., 2007; Südkamp et al., 2008) is a measure of diagnostic knowledge (i.e., PK), models on teachers' PCK consider the diagnosis of subject-specific misconceptions as key (e.g., Großschedl et al., 2018; Ziadie & Andrews, 2018). Teachers need to know valid and reliable assessments to identify students' misconceptions in their subject (e.g., CANS, CINS, ACORNS for evolution in biology; Ziadie & Andrews, 2018). However, teacher-student interactions require consistent diagnosis of students' misconceptions (i.e., ePCK) during real classroom discussions on, for example, natural selection. In order to assess pre-service teachers' knowledge about students' understanding during student-teacher-interactions, the initial SCR was expanded to the Simulated Classroom Biology (SCR^{Bio}), which focuses the theory of evolution.

2.3 | Teaching and learning evolution

Within biology education, the diagnosis of students' understanding about the evolutionary theory is paramount because evolutionary theory is key to a conceptual understanding of all topics in the field of biology (e.g., Anderson et al., 2002; Basel et al., 2014; Bruckermann et al., 2021; Furtak, 2012; Opfer et al., 2012). Unfortunately, the evolutionary theory has always been considered very challenging in the science curriculum, and traditional teaching often fails when it comes to expanding students' understanding to scientific explanations (e.g., Bishop & Anderson, 1990; Gregory, 2009; Kampourakis & Zogza, 2008; McVaugh et al., 2011; Opfer et al., 2012; Shtulman, 2006). Educators and researchers identified common misconceptions among students as well as the public and instructional strategies to overcome these. Thus, a knowledge base is provided for the professional knowledge of teachers about the theory of evolution (for an overview: on misconceptions, see Gregory, 2009; on teaching strategies, see Harms & Reiss, 2019; Ziadie & Andrews, 2018).

Despite its functionality as conceptual organizer for learning about evolution (e.g., Opfer et al., 2012; Tibell & Harms, 2017) the key concept of natural selection itself evokes numerous misconceptions, especially in explaining the adaption of individuals and population to their environment (e.g., Bishop & Anderson, 1990; Ferrari & Chi, 1998). Most misconceptions are based on the belief that evolution occurs because individual organisms actively evolve and adapt during their lifetime (e.g., Bishop & Anderson, 1990; Nehm & Schonfeld, 2008; Settlage Jr, 1994). Teleological and anthropomorphic explanations as well as the misconception of use or disuse are widespread and intensively researched categories of misunderstanding. In teleological explanations the development of a trait is target-oriented and purposeful. Here, traits arise because they result in a survival advantage for the organism (e.g., Alters & Nelson, 2002; Andrews et al., 2011; Bishop & Anderson, 1990; Kampourakis & Zogza, 2008; Nehm & Reilly, 2007; Nehm & Schonfeld, 2008; Settlage Jr, 1994; Sinatra et al., 2008). Anthropomorphic misconceptions are characterized by the transfer of human characteristics—physical and mental—on other living organisms (Byrne et al., 2009). Trait changes result from intentional and purposeful action through the individual to cope with new environmental conditions (e.g., Demastes et al., 1995; Gregory, 2009; Sinatra et al., 2008; Tamir & Zohar, 1991). Similar to anthropomorphic misconceptions is the belief that individuals also have an influence on the physical change through their use or disuse of organs or abilities. After the development of the new traits these are passed on to future generations (e.g., Andrews et al., 2011; Brumby, 1984; Kalinowski et al., 2016; Kampourakis & Zogza, 2008; Prinou et al., 2008). All these misconceptions ignore several evolutionary mechanisms of natural selection and follow a causal explanatory pattern. Supporting students in acquiring conceptual knowledge about the theory of evolution requires teachers to diagnose the most common student misconceptions (e.g., Gregory, 2009), which the teachers' PCK on the theory of evolution presupposes (Ziadie & Andrews, 2018).

Different contexts influence the probability of an expressed misconception regardless of the specific misconception categories. For example, the biological domain (i.e., bacteria, plants, animals, and humans) influences the appearance of a misconception (Opfer et al., 2012). Thus, fewer misconceptions are associated with plants than with animals, for example (Nehm et al., 2012). Moreover, students are more adept at explaining a gained trait than the loss of a trait in the process of natural selection (Nehm et al., 2012; Nehm & Ha, 2011).

2.4 | Aspects of validity

In previous studies, the SCR was used to measure action-oriented PK (Fiedler et al., 2002; Fiedler et al., 2007; Südkamp et al., 2008). However, there is a lack of evidence on what findings need to be considered in research on teaching and learning evolution within biology education by developing a simulated classroom environment and, whether the ePCK of pre-service teachers in a simulated classroom (i.e., through the SCR^{Bio}) leads to valid conclusions. This is one of the key challenges of the present research, to be able to measure the ePCK of pre-service biology teachers via an action-oriented setting—the SCR^{Bio}.

In the development and evaluation of a test instrument to measure teachers' ePCK validity is a fundamental element to be considered (e.g., Großschedl et al., 2019). Validity indicates the degree to which theory and evidence support the interpretation of test results (Kane, 2013). Accordingly, validation concerns not the test instrument itself but the interpretation of results that the test provides (AERA et al., 2014). The current study focuses two aspects of validity, that is, evidence based on the test content and on the relation to other variables.

The first aspect is validity evidence based on test content. Here, the relationship between the test content and the construct to be measured by the test instrument is analysed. Test content includes, for example, the topics, the wording or format of the items, and the guidelines for processing and scoring. Expert ratings are often used to verify test content. Expert judgements give an indication of the content validity between the test items and the theoretical construct (AERA et al., 2014; Sireci & Faulkner-Bond, 2014).

The second aspect is validity evidence based on the analysis of the relationship between test results and other variables. This relationship can refer both to the consideration of different groups (known-groups) and to different constructs as external variables (Mathesius et al., 2019). The known-groups method requires that at least two different groups work on a test. Thus, the test values can be interpreted as valid if the two groups show different, theoretically predicted results with respect to the same construct, depending on their preconditions (Hattie & Cooksey, 1984).

3 | RESEARCH QUESTIONS

In this study, we investigate the validity evidence for the use of the SCR^{Bio} to measure pre-service biology teachers' action-oriented PCK (ePCK), operationalized via the diagnoses of students' misconceptions as part of their knowledge about student understanding. We consider (1) evidence based on the test content and (2) evidence based on known-groups comparisons as sources of validity evidence.

The SCR^{Bio} should provide a measure of pre-service teachers' ePCK, which is reflected in the diagnosis of misconceptions. The virtual students' answers must be reliably assigned to the operationalized misconceptions. Therefore, we used expert ratings to verify the content of the virtual student answers, that is, whether the answer actually contains a specific misconception or expresses a scientific way of thinking.

I. To what extent do the expert judgements support the validity evidence of the developed virtual student answers with regard to the specific categories of misconceptions (*validity evidence based on test content*)?

For the measurement of pre-service teachers' ePCK, scores on the measure should vary for groups known to differ in terms of the specific information prompts. We therefore created groups with specific PCK by prompting PCK-related or PCK-unrelated information.

I. To what extend is the SCR^{Bio} sensitive to the given PCK-related prompt in biology education (PCK) and educational psychology (PK) *(validity evidence based on known-groups comparison)*?

3.1 | Hypotheses

I. Experts who have not been involved in the item development can correctly judge the specific category of misconception or scientific explanations that the items in the SCR^{Bio} are intended to measure.

II. The SCR^{Bio} has good discriminatory power, that is, a PCK-related prompt:

a. increases pre-service teachers' diagnoses of specific misconceptions categories in virtual student answers (refer to PCK).
b. does not increase pre-service teachers' judgement accuracy of virtual students' lesson achievement (refer to PK).

4 | METHODS

4.1 | Sample

The participants were $N = 76$ pre-service teachers from two universities in Northern Germany enrolled in a teacher education program (24% male) with biology as one of their subjects. The teacher preparation in Germany is organized in two phases with phases leading to a degree after three years (bachelor) and five years (master). A total of $n = 38$ participants studied in the Bachelor teacher education program, $n = 37$ in the Master teacher education program and one participant did not provide any information. A proportion of 86.8% aspired to a teaching profession for the academic track (i.e., grammar school) and 13.2% for the nonacademic tracks (i.e., interdenominational school). The age of the study participants ranged between ages 18 and 42, mainly between ages 21–23 (51%), 24–26 (26%), and 27–30 (16%).

4.2 | Development of the simulated classroom biology (SCR^{Bio})

The initial developed non-immersive Simulated Classroom (Fiedler et al., 2002; Fiedler et al., 2007; Südkamp et al., 2008) was expanded to the SCR^{Bio}. Therefore, we created a virtual biology lesson for the application in the SCR^{Bio} with questions on evolution. We prepared ($N = 27$) questions on the process of natural selection covering all groups of organisms (i.e., bacteria, plant, animal, and human). Furthermore, we developed the questions to address the gain or loss of a trait as well as whether the trait has a morphological or physiological origin. For each individual question we designed possible virtual students answers ($N = 476$; between 8–30 words), based on theory and authentic student answers from exams (Baalmann et al., 2004; Gregory, 2009; see Table 1). For each question ($N = 27$; questions having between 30–84 words) the virtual students' answers represented a scientifically correct explanation or a misconception (anthropomorphic, teleological, use or disuse; see Table 2).

4.3 | Procedure

For this study, the pre-service teachers taught a virtual lesson within the computer-based simulation SCR^{Bio} on laptops. The laptops were set up in university lecture halls and pre-service teachers' participation in the study lasted 60 minutes. In the virtual lesson, the pre-service teachers performed a teaching sequence—following a question-answer scheme to diagnose virtual students' misconceptions—and a

TABLE 1 Scheme for the development of evolutionary biological questions and possible student answers in the SCR^Bio

Evolutionary direction	Type of trait	Domain			
		Bacteria	Plant	Animal	Human
Loss	Morphological Physiological	2 questions with 35 answers	3 questions with 50 answers	3 questions with 54 answers	2 questions with 28 answers
Gain	Morphological Physiological	5 questions with 93 answers	4 questions with 74 answers	4 questions with 76 answers	4 questions with 66 answers

TABLE 2 Example question and possible student answers about the red bentgrass (botanical organism; gain of a trait, physiological trait)

Red Bentgrass (*Agrostis tenuis*)
Possible teachers' question: "The red bentgrass is a common grass sometimes found close to copper mines. In these locations the soil is very coppery and has a toxic effect on many plants. Investigations show that the red bentgrass shows a high copper-tolerance that could not be established in conspecifics living beyond the coppery soils. How can the tolerance to copper be explained considering principles of evolutionary biology?"
Anthropomorphic students' answer: "The red bentgrass had the intention to develop a copper-resistance in order to be able to live on soils around copper mines."
Teleologic students' answer: "The copper-resistance of the grass had the purpose to grow even in areas around copper mines."
Use or disuse students' answer: "The intense utilization of the copper-resistance on polluted soils accelerated the development of this trait."
Scientific students' answer: "The change of the environmental conditions led to the shift of the evolutionary pressure whereby only those plants that are copper-resistant can survive in coppery soils."

grading sequence—to assess the overall performance of the class. The simulation of the virtual lesson adhered to the procedure described below.

Before starting the teaching and grading sequence, the SCR^Bio on the laptop screen informed the pre-service teacher of the duration of the study, information about data privacy, the virtual lesson's topic (evolution—natural selection) and the class level to be taught (12th grade). When the SCR^Bio simulation started, the pre-service teachers were randomly assigned to either a control group or the experimental group. In the control group receiving PCK-unrelated information prompts, pre-service teachers were given a brief overview on biology in school, science, and education. In the experimental group receiving PCK-related information prompt, pre-service teachers were given an overview of the three misconception categories and a scientific explanation.

The teaching sequence of the virtual lesson started after the pre-service teachers received the different information prompts. The teaching sequence followed a question-answer scheme in which the pre-service teacher asks a question and the virtual students answer. The pre-service teachers could ask these virtual students questions about natural selection that could be selected from a set of pre-defined questions (see Development section). In the simulated classroom, the virtual class consisting of nine virtual students was represented by students' photos and their names (see Figure 1). The virtual students' answers varied according to their pre-set ability profile. The class was divided into three groups of three virtual students. For each group a different ability profile was determined (i.e., 20%/50%/80% probability of correct answers). If the virtual students answered incorrectly, they always expressed a specific misconception about evolution (i.e., anthropomorphic, teleological or use/disuse). We combined each specific misconception category with each ability profile, what resulted in the nine students. For example, one of the virtual students answered only 20% correctly, while all other answers articulated a teleological misconception. Students who wanted to answer a question virtually raised their hand. This was indicated by a little hand-icon and orange-coloured background of the students' photo (see virtual students 1–7 in Figure 1). Pre-service teachers then selected one virtual student to answer the question. The virtual student's answer was provided in a box. The pre-service teacher immediately had to diagnose the answer after reading it. This means that the pre-service teachers had to assess whether the virtual students' answer was scientifically correct or not and, if it was wrong, which misconception was expressed. The pre-service teachers had 30 minutes to ask as many questions and call up as many virtual students as they needed to diagnose each virtual student's misconception and overall class performance regarding natural selection. The grading sequence immediately followed.

In the grading sequence, the pre-service teachers had to assess the virtual students' proportion of correct answers (on a scale from 0% to 100%), which is described as the overall performance (i.e., lesson achievement). Finally, pre-service teachers had to enter their demographical data and were informed about the aims of the SCR^Bio.

4.4 | Measures

During the teaching sequence we captured the pre-service teachers' assessment of the virtual students' misconceptions and scientific explanations by recording the number of accurate diagnoses as a measure of their ePCK. During the grading sequence we recorded the pre-service teachers' evaluation of the proportion of virtual students'

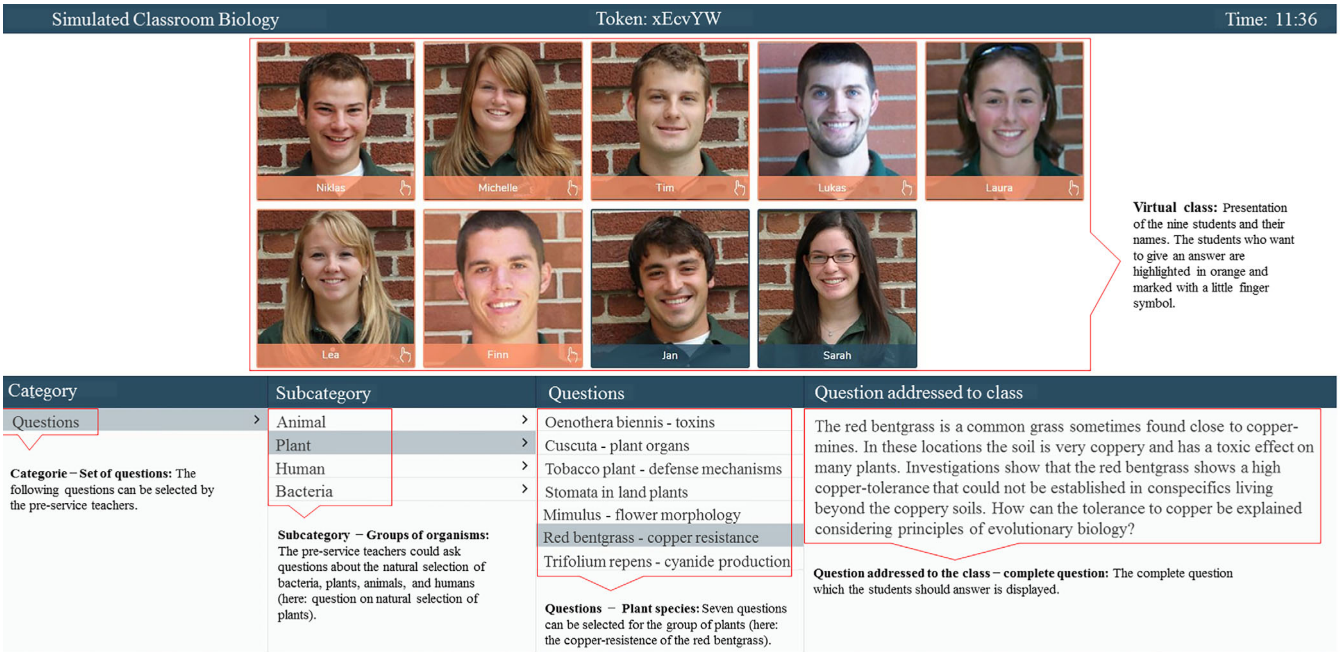

FIGURE 1 Screenshot of the teaching sequence in the SCR^{Bio}

correct answers over the entire virtual lesson as a measure of their PK. All data were recorded and stored as csv-data sets.

ePCK: We operationalize ePCK as the knowledge about student understanding. This is reflected in the diagnosis of specific categories of misconceptions in virtual students' answers by the pre-service teachers during the teaching sequence in the SCR^{Bio}. The relative frequency of correct diagnosis of misconceptions in virtual student answers is quantified as a measure of ePCK.

PK: We operationalize PK as the judgement accuracy regarding the assessment of the virtual students' lesson achievement by the pre-service teachers during the grading sequence in the SCR^{Bio}. We measure judgement accuracy in three components (Südkamp et al., 2012):

i. The ranking component, as the correlation coefficient between pre-service teachers' ratings of virtual students' relative frequency of correct answers and their actual relative frequency.
ii. The level component, as the difference between the class mean of pre-service teachers' ratings of virtual students' relative frequency of correct answers and the actual class mean.
iii. The differentiation component, as the ratio of the class variance of pre-service teachers' ratings of virtual students' relative frequency of correct answers and the actual class variance.

4.5 | Validity

4.5.1 | Validity evidence based on the test content

To proof the evidence based on the test content, we use expert ratings. The experts ($N = 10$) were mainly from the field of biology education and included both researchers and practitioners with various degrees in biology. They evaluated the items in terms of misconceptions by assessing each item's belonging to one of the three categories (i.e., anthropomorphic, teleological, use or disuse) or scientifically correct. Within SCR^{Bio}, ePCK is measured by diagnosing the misconceptions articulated in the answers of the students. Accordingly, it is mandatory that our student answers explicitly express the intended misconception. Each expert received a coding manual which in theory clearly defines the misconceptions to be judged and presents a number of examples. The content validity ratio ($\boldsymbol{CVR} = \left(\boldsymbol{n_l} - \left(\frac{N}{2}\right)\right)/\frac{N}{2}$) by Lawshe (1975) was adapted to quantify the agreement between the expert judgements. Here, $\boldsymbol{n_l}$ = number of experts who judge an item as theoretically intended (i.e., the misconception we integrated into the virtual student answer was successfully identified by the experts) and N = total number of experts ($N = 10$). A CVR of at least 0.78 is necessary to consider an item or scale as valid (Frey, 2018).

Due to the large number of items (i.e., questions and corresponding answers), each expert rated a baseline pool ($N = 8$) and an additional share of the remaining questions ($N = 19$) and answers. The experts rated eight questions (baseline pool; questions from the different groups of organisms: bacteria, plants, animals and humans) and the corresponding 128 possible virtual student answers. In the second part, the experts were divided into five groups (two experts per group) which had to evaluate another four questions and corresponding virtual student answers ($N = 64$) in addition to the baseline pool. We were able to cover all 27 questions and corresponding student answers for the SCR^{Bio} with this expert rating scheme. We revised or selected the questions and the corresponding virtual student answers based on the feedback from the experts.

4.5.2 | Validity evidence based on relation to other variables (known-groups comparison)

Another important source of validity evidence is the analysis of the relationship between test results and external variables. Here, the external variable is an information prompt (i.e., PCK-related or PCK-unrelated information) that is given to different groups (i.e., known-groups: experimental group and control group).

The sample is divided into two groups with different information prompts to provide evidence based on a known-groups comparison. The relative diagnosis rates for the identification of specific misconceptions and of scientific explanations in virtual student answers were calculated separately for both groups. A Pearson chi-square test was performed to show group differences of these rates. Differences indicate the extent to which the information influences the diagnosis of misconceptions and scientific explanations within biology education (ePCK). Additionally, we addressed the components of judgement accuracy between the groups. We applied *t*-tests to determine if the PCK prompt influenced judgement accuracy of the pre-service teachers as a measure of pedagogical knowledge (PK).

5 | RESULTS

5.1 | RQ1: Validity evidence based on test content

The experts evaluated the baseline pool of virtual student answers with an overall $CVR_{mean} = 0.915$. The CVR_{mean} exceeds the threshold of 0.78 (Frey, 2018), indicating that virtual student answers clearly expressed a specific misconception or scientific explanation. A differentiated view of the expert rating shows that the items (baseline pool) were evaluated with an overall $CVR_{mean} = 0.911$ and the remaining sets of items were evaluated with an overall CVR_{mean} between 0.859 and 1. Furthermore, the items were assessed regarding evolutionary direction of the trait (i.e., gain or loss) with an inter-rater agreement of 96.25% and the type of a trait (i.e., morphological or physiological) with an inter-rater agreement of 86.25%. This also confirmed the conceptual development of our question pool.

5.2 | RQ2: Validity based on relation to other variables (known-groups comparison)

Pre-service teachers in the control group who received PCK-unrelated information asked a total of $n = 1378$ questions, whereby 50.7% ($n = 699$) of the virtual student answers contained a misconception. They correctly diagnosed anthropomorphic misconceptions in 43.0% ($n = 99$), use or disuse misconceptions in 56.7% ($N = 123$), and teleological misconceptions with 29.4% ($n = 74$) (see Figure 2). Overall, the control group selected the appropriate category across misconceptions with 43.0% ($n = 300$; for a graphic representation see Figure 2).

In contrast, pre-service teachers in the experimental group who received PCK-related information asked a total of $n = 1556$ questions, whereas 51.2% ($n = 796$) virtual student answers were scientifically correct, and 48.8% ($N = 760$) expressed a misconception. They correctly diagnosed 76.2% ($n = 198$) of anthropomorphic misconceptions, 70.8% ($n = 177$) of misconceptions regarding use or disuse, and 56.4% ($n = 141$) of teleological misconceptions. In total, they correctly

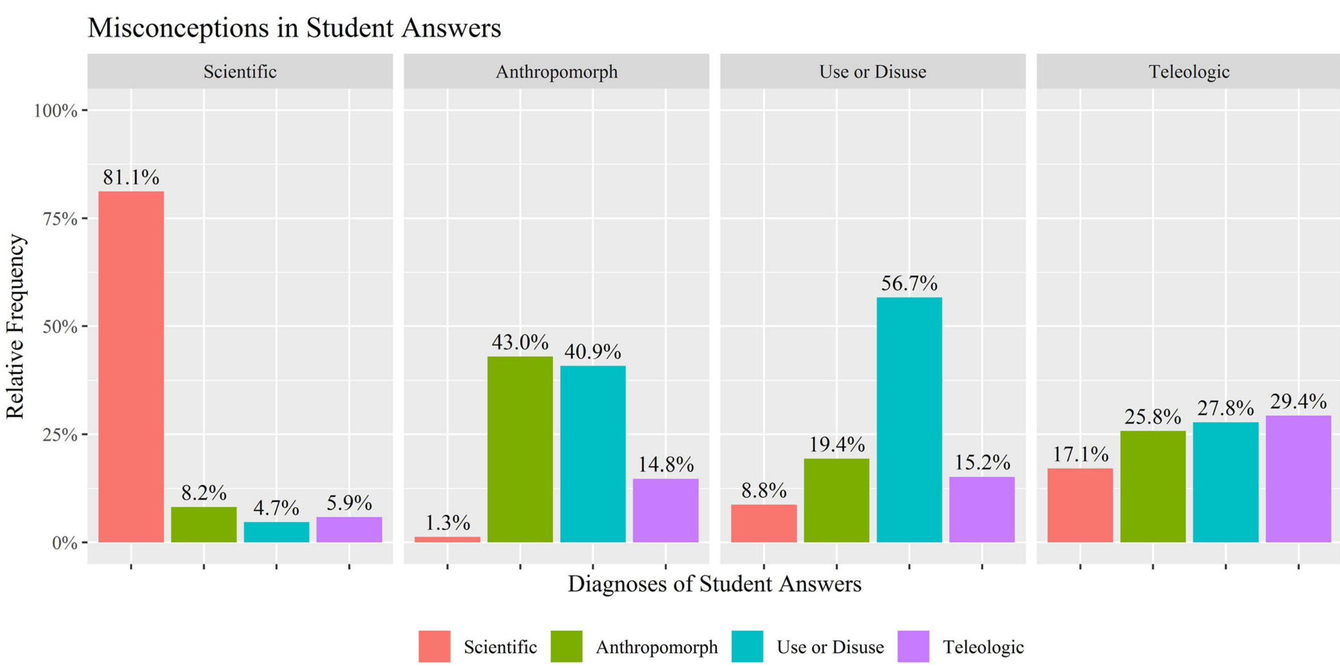

FIGURE 2 Control group: Relative frequency of diagnosis of students' answers with regard to expressed misconceptions categories or scientific explanations

diagnosed 67.8% of all misconceptions and showed a significantly higher rate than the pre-service teachers in the control group ($\chi^2(1) = 97.28$, $p < 0.001$). Regarding the correct diagnosis rate of the scientific explanations of the virtual students, the difference between both groups was not significant ($\chi^2(1) = 0.06$, $p = 0.804$) (for a graphic representation see Figure 3).

We examined the differences between the experimental and the control group on the three judgement accuracy components to check whether the PCK-related information has an impact on the judgement accuracy of the pre-service teachers. Pre-service teachers in the control group reached a mean ranking component of $M_r = 0.31$ ($SD = 0.34$, Min $= -0.48$, Max $= 0.88$), whereas the experimental group has a mean ranking component of $M_r = 0.34$ ($SD = 0.35$, Min $= 0.30$, Max $= 0.90$), the difference was not significant, even with an increased alpha-level of $\alpha = 0.25$ ($t[73] = 0.27$, $p = 0.787$). For the mean level component, a result of $M_n = -0.02$ ($SD = 0.17$, Min $= -0.51$, Max $= 0.38$) was obtained for the control group and $M_n = 0.05$ ($SD = 0.16$, Min $= -0.35$, Max $= 0.33$) for the experimental group. However, the corresponding *t*-test does not allow rejecting group difference in the level component of judgement accuracy ($t[73] = 1.94$, $p = 0.056$). Within the control group there is a mean differentiation component of $M_d = 0.57$ ($SD = 0.25$, Min $= 0.04$, Max $= 1.10$) and for the experimental group of $M_d = 0.60$ ($SD = 0.33$, Min $= 0.06$, Max $= 1.61$). The *t*-test shows no significant group differences with respect to the differentiation component ($t[73] = 0.47$, $p = 0.643$; see Table 3).

6 | DISCUSSION

The present study reports the validity of using a German-language simulated classroom environment called the SCR[Bio] to investigate and measure the ePCK of pre-service biology teachers. More precisely, it reports the diagnosis of common misconceptions about the evolutionary process of natural selection. Thus, the SCR[Bio] must be developed in a way that the construct (i.e., knowledge about student understanding) is detected without the influence of construct-irrelevant variance (AERA et al., 2014). We focus on validity evidence based on test content and relationships to other variables.

The first research question is addressed using the validity evidence based on the test content gained by expert judgements. The test content of the items needs to adequately represent the construct

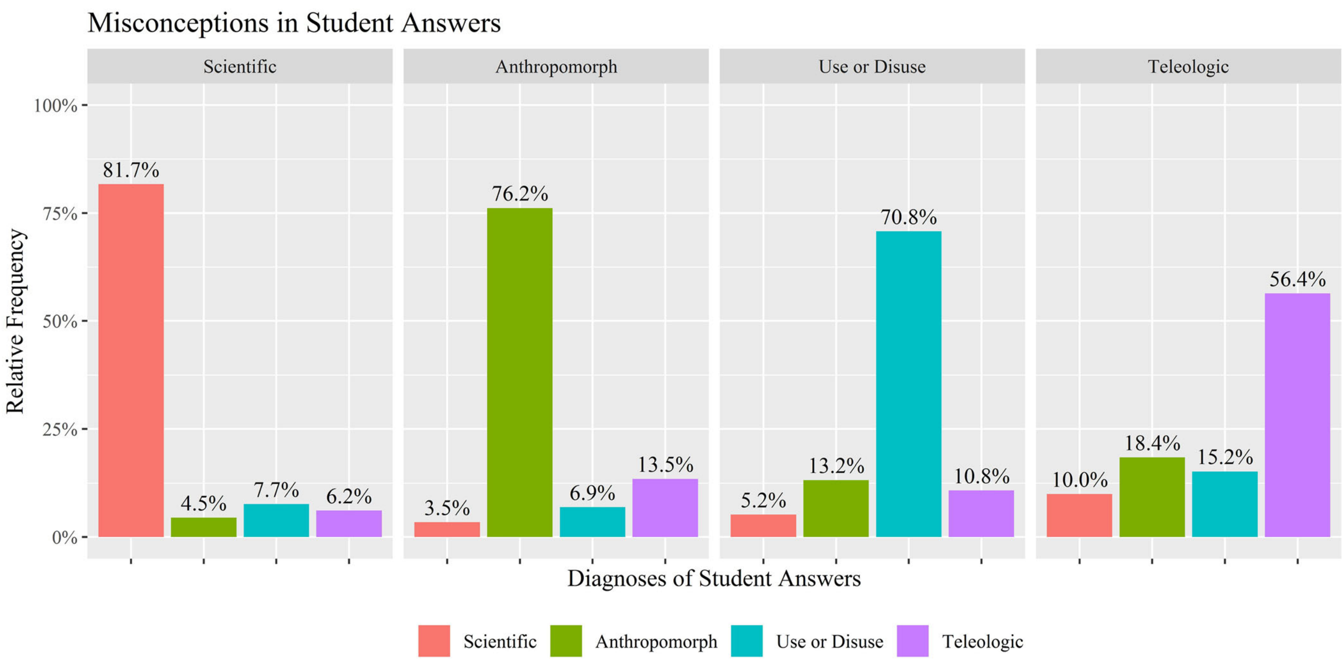

FIGURE 3 Experimental group: Relative frequency of diagnosis of students' answers with regard to expressed misconceptions categories or scientific explanations

TABLE 3 Results of the ranking, level, and differentiation component subdivided by control and experimental group

	Control group				Experimental group			
	M	*SD*	Min.	Max.	*M*	*SD*	Min.	Max.
Ranking component	0.31	0.34	−0.48	0.88	0.34	0.35	0.30	0.92
Level component	−0.02	0.17	−0.51	0.38	0.05	0.16	−0.35	0.33
Differentiation component	0.57	0.25	0.04	1.10	0.60	0.33	0.06	1.61

Sample. *N* = 76.

knowledge about student understanding which is necessary for the valid interpretation of the test results. In this study, researchers working in the field of biology education and biology teaching assigned the items (baseline pool) to common misconception categories or scientific ways of thinking. This resulted in an almost perfect agreement ($CVR_{mean} = 0.911$; see Ayre & Scally, 2014) with the intended wording of the items. Additional items have similar agreements (CVR_{mean} between 0.859–1). However, the overall CVR_{mean} of 0.915 is proposed to be sufficient for indicating validity based on test content (Ayre & Scally, 2014). Thus, the virtual student answers clearly expresses a specific category of misconception or a scientific way of thinking and we can assume the virtual student answers can be used as a valid measure of the knowledge about student understanding. Confirmation biases can be ignored (Guion, 1977) as the experts were not involved in the development of the SCR^{Bio}.

The second hypothesis, which focuses on the validity criterion of known groups, predicted that pre-service teachers who received a PCK-related information prompt (experimental group) show a higher diagnosis of specific misconception categories than their colleagues who received a PCK-unrelated information prompt (control group). In accordance with the hypothesized group differences, we found a significant difference in the diagnostic rate of the specific misconception categories between the groups ($\chi^2(1) = 97.28$, $p < 0.001$). Thus, the PCK-related information prompt influences the diagnosis of misconceptions positively and revealed the sensitivity of the SCR^{Bio} for ePCK. In this regard, the diagnosis of misconception is based on the ePCK of the pre-service teachers (e.g., Förtsch et al., 2018; Lucero et al., 2017). Within the specific misconception categories, some misconceptions were more likely to be diagnosed than others. Accordingly, pre-service teachers sometimes have difficulties in making correct assessments regarding their students. For example, the anthropomorphic misconception is assessed as supposedly more unscientific than the teleological misconception. At the same time, the teleological misconception was less frequently diagnosed as a non-scientific way of thinking than the other two misconceptions (anthropomorphic, use or disuse). An explanation for this could be given by empirical studies, which showed that many everyday problems are explained in a goal- or purpose-oriented way. If a transfer of these explanatory schemes to the evolutionary process of natural selection occurs, misconceptions may result (Gresch & Martens, 2019).

We conclude that this evidence supports the validity of our interpretation of the test results as measures for ePCK since the group means differ in the expected directions. In contrast, we observed no group differences in the diagnosis of scientifically correct virtual student answers ($\chi^2(1) = 0.06$, $p = 0.804$), indicating that the knowledge for the content-related assessment of an answer (i.e., scientifically correct or incorrect; refers to CK) differs from the knowledge about student understanding (i.e., diagnosis of specific misconception categories; refers to PCK; e.g., Förtsch et al., 2018; Großschedl et al., 2015). The PCK-related information prompt does not seem to influence the accuracy of pre-service teachers (based on their PK, Südkamp et al., 2008), in terms of the correlation between judgements of lesson achievement and actual achievement. Neither does it seem to influence the lesson judgement variance relative to the actual achievement variance. However, we cannot exclude that the PCK-related information on misconceptions led to different mean class judgement levels, possibly in terms of stricter judgements in the experimental condition. Although group differences do not appear to be major at first glance, these results provide clues to the distinction of knowledge needed to (1) diagnose common misconceptions in virtual student answers (ePCK) and (2) assess virtual students' lesson achievement (PK).

As previously mentioned, validity cannot be stated universally but rather only for a specific interpretation of test results (AERA et al., 2014; Kane, 2013). In summary, this study provides evidence that the content of the items corresponds to the theoretical meaning of knowledge about student understanding (evidence based on test content) and indicates the sensitivity of the SCR^{Bio} for ePCK by differences in the relative frequency of diagnosing specific misconceptions categories in the virtual students answers between the experimental and control group (known- groups). Therefore, the present simulated classroom environment—the SCR^{Bio}—may be validly used to provide findings on the ePCK in biology education. Furthermore, the validation of results in the SCR^{Bio} contributes to closing the gap from the availability of simulated classroom environments, which only cover the pre-service teachers' action-oriented PK by now allowing the measurement of ePCK in biology.

6.1 | Limitations

The newly developed SCR^{Bio} is the first attempt to investigate ePCK of pre-service teachers in an action-oriented simulated classroom environment. In accordance with our study design we used the following sources for validity evidence: (1) evidence based on the test content and (2) evidence based on relation to other variables to provide a use argument for the valid interpretation of the test results in the SCR^{Bio}. However, the *Standards for Educational and Psychological Testing* provide various further sources of validity. These include the evidence based on internal structure, evidence based on response processes, and evidence based on the consequences of testing (AERA et al., 2014), which we did not include in this first validation approach.

Our approach for providing content validity was relatively straightforward compared to other approaches. We used a distinct operationalization of misconceptions which likely led to the very high agreement on the answer categories across expert raters. In the SCR^{Bio}, we used virtual student answers that are somewhat simplistic compared to more potential real-life student answers.

The known-groups comparison in the SCR^{Bio} was based on an information prompt in which a part of the sample received information helpful in processing the SCR^{Bio} while the remaining pre-service teachers received no helpful information. This comparison supports the valid interpretation of the results in the SCR^{Bio}. However, other group differences, which we have not considered in order to support this validity criterion, can also be taken into account, such as differences regarding the teaching career (e.g., academic or non-academic track), the status of the university career (e.g., graduate or undergraduate students), the teaching experience (e.g., pre-service or in-service teachers) or the studied subjects (e.g., one or two natural science subjects; e.g., Großschedl et al., 2019; Mathesius et al., 2019).

In the current SCR^{Bio}, we used student answers from three common misconception categories (i.e., anthropomorphic, teleological, and use or disuse). However, research has identified various categories of misconceptions about evolution (e.g., Bishop & Anderson, 1990; Gregory, 2009; Nehm & Schonfeld, 2008; Settlage Jr, 1994). For example, religious beliefs were abandoned because they do not have much influence in Germany (e.g., Großschedl et al., 2018), but in other countries they do (e.g., Berti et al., 2010; Billingsley et al., 2016).

Finally, the SCR^{Bio} is a complexity reducing simulated classroom environment designed to come close to reality of a real classroom (e.g., Fiedler et al., 2007; Südkamp et al., 2008). However, the instrument is still an environment of reduced information, blanking out other factors that influence teaching, such as student disruptions, a change in the student's ability during the virtual school lesson, the virtual students' tone, gaze, and gestures or answers that do not represent attempts to successfully answer the posed question (i.e., classroom management; PK). Thus, it is necessary to control the manipulated variables in the SCR^{Bio} to reduce the noise of the investigation. Nevertheless, we would like to emphasize that the standardization of the virtual school class (photos, associated name) as well as the possible virtual student answers contribute significantly to reducing the complexity of the simulated classroom environment. The resulting consequences, such as a lower realism of the SCR^{Bio}, are consciously accepted in order to ensure constant conditions.

6.2 | Implications for further research

Few simulated classroom environments have been developed to measure and promote action-oriented facets of PK (e.g., Bautista & Boone, 2015; Christensen et al., 2011; Dalgarno et al., 2016), but none that cover ePCK. We were successful in validating the SCR^{Bio} in German language so far, but still aim to use and compare the instrument in other countries and expand the range of participants. We therefore integrated new questions about the process of natural selection and corresponding answers from American college students (Nehm et al., 2012; *ACORNS*). Future studies with the SCR^{Bio} will focus on the extent to which pre-service teachers are able to diagnose misconceptions with regard to the evolutionary direction of a trait (i.e., loss or gain), the type of trait (i.e., physiological or morphological) or the domain (i.e., bacteria, plant, animal or human) of the biological example. Some studies have already examined the influence of these factors in relation to diagnostic ability (e.g., Nehm et al., 2012; Opfer et al., 2012). Furthermore, students' ability level was fixed during the teaching sequence of 60 minutes. Pre-service teachers, however, need to understand and diagnose changes in students' abilities as they are likely to change during teaching. Thus, other studies should use the SCR^{Bio} in a longitudinal design that presents the same class but with changing ability profiles on multiple occasions to test pre-service teachers' ability for detecting students' ability growth. Beyond the diagnosis of misconceptions (e.g., Gregory, 2009), pre-service teachers need to cover another key facet of PCK—instructional strategies (e.g., the conceptual change strategy)—in order to respond adequately to misconceptions regarding evolution (e.g., Ziadie & Andrews, 2018). New versions of the SCR^{Bio} will integrate adequate instructional strategies (i.e., strategies to overcome specific misconceptions), which must be selected by the pre-service teacher. We are currently conducting initial validation studies. In addition, future versions of the SCR^{Bio} should attempt to replicate previous results with similar research questions to generalize the findings as much as possible. In the future, the SCR^{Bio} will be developed from an assessment instrument to a training tool to simulate explicit teaching situations. In this training tool, university students will have the opportunity to train action-oriented professional knowledge. In this context, further challenges in the use of simulated classroom environments, such as more realistic representation (e.g., animated avatar), embodiment (e.g., through facial expressions, gestures, and tone of voice), and further background information (e.g., prior knowledge, experiences, and interests) of virtual students would also need to be intensively addressed (e.g., Jamaludin et al., 2009). A further goal is to provide direct feedback on performance in the SCR^{Bio} through a new developed report module. Following previous research on technological pedagogical content knowledge (TPACK), which describes a teacher's knowledge regarding the use of technologies in the teaching/learning process, the SCR^{Bio} can be used to examine if and how the use of this simulated classroom environment affects the development of TPACK (Cox & Graham, 2009). Furthermore, teachers need knowledge of how to use the technology when working with the SCR^{Bio} and that the TPACK model lends itself to this. In particular, the fact that the SCR^{Bio} itself, as a technology-intensive learning environment, meets the demands for integrating TPACK into teacher education programs emphasizes the need for attention to this concept in future simulated classroom environments (Niess, 2011). Thus, the SCR^{Bio} will significantly contribute to preparing future biology teachers for the challenges of teaching and learning evolution in class.

AUTHOR CONTRIBUTIONS

Julian Fischer, Nils Machts, Till Bruckermann, Jens Möller, Ute Harms drafted the research questions and Julian Fischer collected the data on several survey dates. Julian Fischer and Nils Machts conducted all the psychometric analyses, drafted the figures, and wrote the results. All authors contributed to the writing of the final manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors are also very grateful to Cornelia Gerigk for the language review. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

As part of the project ProSim (16DHL1014): Development of procedural professional knowledge with the Simulated Classroom, this project received funding from the Federal Ministry of Education and Research (BMBF).

CONFLICT OF INTEREST

The authors declare they have no competing interests.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12718.

DATA AVAILABILITY STATEMENT

Available from authors upon request.

PUBLICATION CONSENT

The authors received consent from all study participants to collect and publish the data they shared with us.

ORCID

Julian Fischer https://orcid.org/0000-0003-2552-9894
Till Bruckermann https://orcid.org/0000-0002-8789-8276
Jens Möller https://orcid.org/0000-0003-1767-5859
Ute Harms https://orcid.org/0000-0001-6284-9219

REFERENCES

Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell (Ed.), *Handbook of research on science education* (pp. 1105–1149). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203824696-44

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Alters, B. J., & Nelson, C. E. (2002). Perspective: Teaching evolution in higher education. *Evolution: International Journal of Organic Evolution*, *56*(10), 1891–1901. https://doi.org/10.1111/j.0014-3820.2002.tb00115.x

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching*, *39*(10), 952–978. https://doi.org/10.1002/tea.10053

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. https://doi.org/10.1007/s12052-011-0343-4

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development*, *47*(1), 79–86. https://doi.org/10.1177/0748175613513808

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 processes of adaptation-results of an interview study in the context of didactic reconstruction]. *Zeitschrift für Didaktik der Naturwissenschaften*, *10*(1), 7–28.

Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. *Handbook of Research on Teaching*, *4*, 433–456. https://doi.org/10.12691/education-3-4-18

Basel, N., Harms, U., Prechtl, H., Weiss, T., & Rothgangel, M. (2014). Students' arguments on the science and religion issue: The example of evolutionary theory and genesis. *Journal of Biological Education*, *48*(4), 179–187.

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften [Teachers' professional competence]. *Zeitschrift für Erziehungswissenschaft*, *9*(4), 469–520. https://doi.org/10.1007/s11618-006-0165-2

Baumert, J., & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (pp. 29–53). Waxmann.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., … Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*, 133–180. https://doi.org/10.3102/0002831209345157

Bautista, N. U., & Boone, W. J. (2015). Exploring the impact of TeachME™ lab virtual classroom teaching simulation on early childhood education majors' self-efficacy beliefs. *Journal of Science Teacher Education*, *26*(3), 237–262. https://doi.org/10.1007/s10972-014-9418-8

Berti, A. E., Toneatti, L., & Rosati, V. (2010). Children's conceptions about the origin of species: A study of Italian children's conceptions with and without instruction. *Journal of the Learning Sciences*, *19*(4), 506–538. https://doi.org/10.1080/10508406.2010.508027

Billingsley, B., Brock, R., Taber, K. S., & Riga, F. (2016). How students view the boundaries between their science and religious education concerning the origins of life and the universe. *Science Education*, *100*(3), 459–482. https://doi.org/10.1002/sce.21213

Bishop, B. A., & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching*, *27*(5), 415–427. https://doi.org/10.1002/tea.3660320509

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie*, *223*(3), 3–13. https://doi.org/10.1027/2151-2604/a000194

Bradley, E. G., & Kendall, B. (2014). A review of computer simulations in teacher education. *Journal of Educational Technology Systems*, *43*(1), 3–12. https://doi.org/10.2190/ET.43.1.b

Bruckermann, T., Fiedler, D., & Harms, U. (2021). Identifying precursory concepts in evolution during early childhood—A systematic literature review. *Studies in Science Education*, *57*(1), 85–127. https://doi.org/10.1080/03057267.2020.1792678

Brumby, M. N. (1984). Misconceptions about the concept of natural selection by medical biology students. *Science Education*, *68*(4), 493–503. https://doi.org/10.1002/sce.3730680412

Byrne, J., Grace, M., & Hanley, P. (2009). Children's anthropomorphic and anthropocentric ideas about micro-organisms: Educational research. *Journal of Biological Education*, *44*(1), 37–43. https://doi.org/10.1080/00219266.2009.9656190

Carlson, J., & Daehler, K. R. (2019). The refined consensus model of pedagogical content knowledge in science education. In *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 77–92). Springer. https://doi.org/10.1007/978-981-13-5898-2_2

Cheong, D. (2010). The effects of practice teaching sessions in second life on the change in pre-service teachers' teaching efficacy. *Computers in Education*, *55*(2), 868–880. https://doi.org/10.1016/j.compedu.2010.03.018

Christensen, R., Knezek, G., & Tyler-Wood, T. (2011). simSchool: An online dynamic simulator for enhancing teacher preparation. *International Journal of Learning Technology*, *6*(2), 201–220. https://doi.org/10.1504/IJLT.2011.042649

Clapper, T. C. (2010). Role play and simulation returning to teaching for understanding. *The Education Digest*, *75*(8), 39–43.

Cox, S., & Graham, C. R. (2009). Diagramming TPACK in practice: Using and elaborated model of the TPACK framework to analyze and depict teacher knowledge. *TechTrends*, *53*, 60–69.

Dalgarno, B., Gregory, S., Knox, V., & Reiners, T. (2016). Practising teaching using virtual classroom role plays. *Australian Journal of Teacher Education*, *41*(1), 126–154. https://doi.org/10.14221/ajte.2016v41n1.8

Demastes, S. S., Settlage, J., Jr., & 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*(5), 535–550. https://doi.org/10.1002/tea.3660320509

Depaepe, F., Torbeyns, J., Vermeersch, N., Janssens, D., Janssen, R., Kelchtermans, G. V., & van Dooren, W. (2015). Teachers' content and pedagogical content knowledge on rational numbers: A comparison of prospective elementary and lower secondary school teachers. *Teaching and Teacher Education*, *47*, 82–92. https://doi.org/10.1016/j.tate.2014.12.009

Ferrari, M., & Chi, M. T. (1998). The nature of naive explanations of natural selection. *International Journal of Science Education*, *20*(10), 1231–1256. https://doi.org/10.1080/0950069980201005

Fiedler, K., Freytag, P., & Unkelbach, C. (2007). Pseudo-contingencies in a simulated classroom. *Journal of Personality and Social Psychology*, *92*, 665–677. https://doi.org/10.1037/0022-3514.92.4.665

Fiedler, K., Walther, E., Freytag, P., & Plessner, H. (2002). Judgment biases in a simulated classroom—A cognitive-environmental approach. *Organizational Behavior and Human Decision Processes*, *88*, 527–561. https://doi.org/10.1006/obhd.2001.2981

Fischer, H. E., Borowski, A., & Tepner, O. (2012). Professional knowledge of science teachers. In *Second international handbook of science education* (pp. 435–448). Springer. https://doi.org/10.1007/978-1-4020-9041-7_30

Förtsch, C., Sommerhoff, D., Fischer, F., Fischer, M., Girwidz, R., Obersteiner, A., … Seidel, T. (2018). Systematizing professional knowledge of medical doctors and teachers: Development of an interdisciplinary framework in the context of diagnostic competences. *Education in Science*, *8*(4), 207–225. https://doi.org/10.3390/educsci8040207

Frey, B. B. (Ed.). (2018). *The SAGE encyclopedia of educational research, measurement, and evaluation*. Sage Publications. https://doi.org/10.4135/9781506326139

Friedrichsen, P. J., Abdell, S., Pareja, E., Brown, P., Lankford, D., & Volkmann, M. (2009). Does teaching experience matter? Examining biology teachers' prior knowledge for teaching in an alternative certification program. *Journal of Research in Science Teaching*, *46*, 357–383. https://doi.org/10.1002/tea.20283

Furtak, E. M. (2012). Linking a learning progression for natural selection to teachers' enactment of formative assessment. *Journal of Research in Science Teaching*, *49*(9), 1181–1210. https://doi.org/10.1002/tea.21054

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

Großschedl, J., Harms, U., Kleickmann, T., & Glowinski, I. (2015). Preservice biology teachers' professional knowledge: Structure and learning opportunities. *Journal of Science Teacher Education*, *26*(3), 291–318.

Großschedl, J., Mahler, D., Kleickmann, T., & Harms, U. (2014). Content-related knowledge of biology teachers from secondary schools: Structure and learning opportunities. *International Journal of Science Education*, *36*(14), 2335–2366.

Großschedl, J., Seredszus, F., & Harms, U. (2018). Angehende Biologielehrkräfte: Evolutionsbezogenes Wissen und Akzeptanz der Evolutionstheorie. [prospective biology teachers: Evolution-related knowledge and acceptance of the theory of evolution.]. *Zeitschrift für Didaktik der Naturwissenschaften.*, *24*(1), 51–70.

Großschedl, J., Welter, V., & Harms, U. (2019). A new instrument for measuring pre-service biology teachers' pedagogical content knowledge: The PCK-IBI. *Journal of Research in Science Teaching*, *56*(4), 402–439. https://doi.org/10.1002/tea.21482

Guion, R. (1977). Content validity: The source of my discontent. *Applied Psychological Measurement*, *1*, 1–10. https://doi.org/10.1177/014662167700100103

Gurvitch, R., & Metzler, M. W. (2009). The effects of laboratory-based and field-based practicum experience on pre-service teachers' self-efficacy. *Teaching and Teacher Education*, *25*(3), 437–443. https://doi.org/10.1016/j.tate.2008.08.006

Harms, U., & Reiss, M. J. (2019). The present status of evolution education. In U. Harms & M. J. Reiss (Eds.), *Evolution education re-considered* (pp. 1–19). Springer Verlag.

Hattie, J. (2009). Visible learning. A synthesis of over 800 meta-analyses relating to achievement. *Routledge*. https://doi.org/10.4324/9780203887332

Hattie, J., & Cooksey, R. W. (1984). Procedures for assessing the validities of tests using the" known-groups" method. *Applied Psychological Measurement*, *8*(3), 295–305. https://doi.org/10.1177/014662168400800306

Henze, I., & van Driel, J. H. (2015). Toward a more comprehensive way to capture PCK in its complexity. *Re-examining Pedagogical Content Knowledge in Science Education*, 120–134.

Hixon, E., & So, H.-J. (2009). Technology's role in field experiences for preservice teacher training. *Educational Technology & Society*, *12*, 294–304.

Jamaludin, A., San Chee, Y., & Ho, C. M. L. (2009). Fostering argumentative knowledge construction through enactive role play in second life. *Computers & Education*, *53*(2), 317–329. https://doi.org/10.1016/j.compedu.2009.02.009

Jüttner, M., Boone, W., Park, S., & Neuhaus, B. J. (2013). Development and use of a test instrument to measure biology teachers' content knowledge (CK) and pedagogical content knowledge (PCK). *Educational Assessment, Evaluation and Accountability*, *25*(1), 45–67. https://doi.org/10.1007/s11092-013-9157-y

Jüttner, M., & Neuhaus, B. J. (2013). Validation of a paper-and-pencil test instrument measuring biology teachers' pedagogical content knowledge by using think-aloud interviews. *Journal of Education and Training Studies*, *1*(2), 113–127. https://doi.org/10.11114/jets.v1i2.126

Kaiser, J., Retelsdorf, J., Südkamp, A., & Möller, J. (2013). Achievement and engagement: How student characteristics influence teacher judgments. *Learning and Instruction*, *28*, 73–84.

Kaiser, J., Südkamp, A., & Möller, J. (2017). The effects of student characteristics on teachers' judgment accuracy: Disentangling ethnicity, minority status, and achievement. *Journal of Educational Psychology*, *109*(6), 871–888.

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*(4), 64. https://doi.org/10.1187/cbe.15-06-0134

Kampourakis, K., & Zogza, V. (2008). Students' intuitive explanations of the causes of homologies and adaptations. *Science & Education*, *17*(1), 27–47. https://doi.org/10.1007/s11191-007-9075-9

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73. https://doi.org/10.1111/jedm.12000

Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert, J. (2013). Teachers' content knowledge and pedagogical content knowledge: The role of structural differences in teacher education. *Journal of Teacher Education*, *64*(1), 90–106. https://doi.org/10.1177/0022487112460398

Kunter, M., Baumert, J., & Blum, W. (Eds.). (2011). *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV [professional competence of teachers: Results of the COACTIV research programme]*. Waxmann Verlag.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (2013). *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project*. Springer. https://doi.org/10.1007/9781-4614-5149-5

Lawshe, C. H. (1975). A quantitative approach to content validity 1. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Lucero, M. M., Petrosino, A. J., & Delgado, C. (2017). Exploring the relationship between secondary science teachers' subject matter knowledge and knowledge of student conceptions while teaching evolution by natural selection. *Journal of Research in Science Teaching*, *54*(2), 219–246. https://doi.org/10.1002/tea.21344

Magnusson, S., Krajcik, J. S., & Borko, H. (1999). Nature, sources and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 95–132). Springer. https://doi.org/10.1007/0-306-47217-1_4

Mahler, D., Großschedl, J., & Harms, U. (2017). Using doubly latent multilevel analysis to elucidate relationships between science teachers' professional knowledge and students' performance. *International Journal of Science Education*, *39*(2), 213–237.

Mathesius, S., Krell, M., Upmeierzu Belzen, A., & Krüger, D. (2019). Überprüfung eines tests zum wissenschaftlichen Denken unter Berücksichtigung des Validitätskriteriums relations-to-other-variables [review of a test of scientific thinking taking into account the validity criterion relations-to-other-variables]. *Zeitschrift für Pädagogik*, *4*, 492–510. https://doi.org/10.3262/ZP1904492

McVaugh, N. K., Birchfield, J., Lucero, M. M., & Petrosino, A. J. (2011). Evolution education: Seeing the forest for the trees and focusing our efforts on the teaching of evolution. *Evolution: Education and Outreach*, *4*(2), 286–292. https://doi.org/10.1007/s12052-010-0297-y

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*(2), 92–98. https://doi.org/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*(3), 237–256. https://doi.org/10.1002/tea.20400

Nehm, R. H., & Reilly, L. (2007). Biology majors' knowledge and misconceptions of natural selection. *Bioscience*, *57*(3), 263–272. https://doi.org/10.1641/B570311

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*(10), 1131–1160. https://doi.org/10.1002/tea.20251

Niess, M. L. (2011). Investigating TPACK: Knowledge growth in teaching with technology. *Journal of Educational Computing Research*, *44*, 299–317.

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*(6), 744–777. https://doi.org/10.1002/tea.21028

Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, *8*(3), 293–316. https://doi.org/10.1016/0361-476X(83)90018-8

Park, S., & Oliver, J. S. (2008). Revisiting the conceptualization of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, *38*, 261–284. https://doi.org/10.1007/s11165-007-9049-6

Plass, J. L., Homer, B. D., & Hayward, E. O. (2009). Design factors for educationally effective animations and simulations. *Journal of Computing in Higher Education*, *21*(1), 31–61.

Prinou, L., Halkia, L., & Skordoulis, C. (2008). What conceptions do Greek school students form about biological evolution? *Evolution: Education and Outreach*, *1*(3), 312–317. https://doi.org/10.1007/s12052-008-0051-x

Rozenszajn, R., & Yarden, A. (2014). Expansion of biology teachers' pedagogical content knowledge (PCK) during a long-term professional development program. *Research in Science Education*, *44*(1), 189–213. https://doi.org/10.1007/s11165-013-9378-6

Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal*, *50*(5), 1020–1049. https://doi.org/10.3102/0002831213477680

Sauvé, L., Renaud, L., Kaufman, D., & Marquis, J.-S. (2007). Distinguishing between games and simulations: A systematic review. *Journal of Educational Technology & Society*, *10*, 247–256.

Settlage, J., Jr. (1994). Conceptions of natural selection: A snapshot of the sense-making process. *Journal of Research in Science Teaching*, *31*(5), 449–457. https://doi.org/10.1002/tea.3660310503

Shtulman, A. (2006). Qualitative differences between naïve and scientific theories of evolution. *Cognitive Psychology*, *52*(2), 170–194. https://doi.org/10.1016/j.cogpsych.2005.10.001

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14.

Shulman, L. S. (1987). Knowledge and teaching: Foundation of the new reform. *Harvard Educational Review*, *57*(1), 1–22.

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*(2), 189–195. https://doi.org/10.1007/s12052-008-0037-8

Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, *26*(1), 100–107. https://doi.org/10.7334/psicothema2013.256

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(3), 743–762.

Südkamp, A., Möller, J., & Pohlmann, B. (2008). Der Simulierte Klassenraum: Eine experimentelle Untersuchung zur diagnostischen Kompetenz [The simulated classroom: An experimental investigation of diagnostic competence]. *Zeitschrift für Pädagogische Psychologie*, *22*(34), 261–276.

Tamir, P. (1988). Subject matter and related pedagogical knowledge in teacher education. *Teaching and Teacher Education*, *4*(2), 99–110. https://doi.org/10.1016/0742-051X(88)90011-X

Tamir, P., & Zohar, A. (1991). Anthropomorphism and teleology in reasoning about biological phenomena. *Science Education*, *75*(1), 57–67. https://doi.org/10.1002/sce.3730750106

Tibell, L. A., & Harms, U. (2017). Biological principles and threshold concepts for understanding natural selection. *Science Education*, *26*(7), 953–973.

Veletsianos, G., Heller, R., Overmyer, S., & Procter, M. (2010). Conversational agents in virtual worlds: Bridging disciplines. *British Journal of Educational Technology*, *41*(1), 123–140. https://doi.org/10.1111/j.1467-8535.2009.01027.x

Yeh, Y.-C. (2004). Nurturing reflective teaching during critical-thinking instruction in a computer simulation program. *Computers & Education*, *42*(2), 181–194. https://doi.org/10.1016/S0360-1315(03)00071-X

Ziadie, M. A., & Andrews, T. C. (2018). Moving evolution education forward: A systematic analysis of literature to identify gaps in collective knowledge for teaching. *CBE Life Sciences Education*, *17*(1), 11. https://doi.org/10.1187/cbe.17-08-0190

How to cite this article: Fischer, J., Machts, N., Bruckermann, T., Möller, J., & Harms, U. (2022). The Simulated Classroom Biology—A simulated classroom environment for capturing the action-oriented professional knowledge of pre-service teachers about evolution. *Journal of Computer Assisted Learning*, *38*(6), 1765–1778. https://doi.org/10.1111/jcal.12718