

Adaptive prompts for learning Evolution with worked examples - Highlighting the students between the "novices" and the "experts" in a classroom

Charlotte Neubrand^a, Christoph Borzikowsky^b and Ute Harms^a

^aIPN - Leibniz Institute for Science and Mathematics Education, Kiel, GERMANY ^bInstitute of Medical Informatics and Statistics, Kiel, GERMANY

ABSTRACT

Evolutionary theory constitutes the overarching concept in biology. There is hardly any other concept that is more complex, and causes more difficulties in learning and teaching. One instructional approach in optimizing the learning of complex topics is to use worked examples combined with self-explanation prompts that fit to the prior knowledge (knowledge adapted prompts). Especially from cognitive psychological research we know, that prior knowledge is a tremendously relevant factor for learning. However, corresponding studies so far mainly consider the domain specific prior knowledge of high knowledge (expert) versus low knowledge (novice) students. The majority of the learners in a classroom - namely students between these experts and novices - were hardly focused on. These students will be considered here. The aim of our study was to identify how these learners with average prior knowledge can be supported by prompts when learning with worked examples.

Using worked examples we analyzed how different types of self-explanation prompts (at novice and/or expert level) affect knowledge acquisition in evolution of learners with average prior knowledge. For determining the prior biological knowledge we used a general biological content knowledge test (GBCK). The learning gain was measured with an evolutionary biological content knowledge test (EBCK). Knowing what type of prompt is most effective for the learners with average knowledge we compared the benefits of this instructional combination between the three knowledge levels: novices, averages, and experts.

Results show that for learners with average knowledge, all types of prompts were equally effective. The Matthew effect was not reliable between the knowledge levels.

According to our results, learners with average prior knowledge did not require explicit measures of differentiation for learning evolution with prompted worked examples. Nonetheless, for the experts it seems not appropriate to use worked examples with adapted self-explanation prompts. Rather it may be advisable to use another instructional format than worked examples.

KEYWORDS

Evolution, worked examples, adaptive prompts, average knowledge, self-explaining

ARTICLE HISTORY

Received 18 February 2016

Revised 31 July 2016

Accepted 11 August 2016

CORRESPONDENCE Charlotte Neubrand Email:✉ neubrand@ipn.uni-kiel.de

© 2016 Neubrand, Borzikowsky and Harms. Open Access terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>) apply. The license permits unrestricted use, distribution, and reproduction in any medium, on the condition that users give exact credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if they made any changes.

Introduction

Learning and understanding evolutionary theory as a cognitive framework is of central importance in understanding the living world (Nehm et al., 2009; National Research Council and National Academy of Sciences [NRC and NAS], 2012). It is not only a matter of looking backward and trying to find out relationships (e.g. between modern humans and Neanderthales). The development of antibiotic resistance for instance shows that evolution is an ongoing process on our planet that affects our daily lives. However, evolutionary theory is one of the most complex concepts of biology (Mayr, 1982, p. 481). This is reflected in the fact that evolution is poorly understood by students (Bishop & Anderson, 1990; Brumby, 1979; Gregory, 2009; Nehm & Reilly, 2007; Opfer, Nehm, & Ha, 2012), and they only show very basic skills in argumentation on evolutionary topics (Basel, Harms, & Prechtel 2013; Basel, Harms, Prechtel, Weiß, & Rothgangel 2014). Even after taking courses in evolution, students harbor plenty of misconceptions (e.g. Brumby, 1979). Moreover, Yates and Marek (2014) have shown that teachers actually arouse such misconceptions in lesson. One reason may be that the teachers themselves did not achieve a deep understanding of evolutionary theory and show misconceptions that are commonly held by students (Nehm & Schonfeld, 2007). This may also explain why evolution is perceived as the most difficult topic to teach in biology (Bestermann & Baggott La Velle, 2007). Altogether, these findings indicate that (a) evolutionary misconceptions are highly stable over time, and (b) there are plenty of difficulties in teaching and learning evolution. One way of overcoming these difficulties is by ensuring that first of all, teachers have a profound knowledge of evolution (Großschedl, Konnemann, & Basel, 2014). Accordingly, research has focused on what particularly needs to be taught for enhancing an accurate understanding of evolutionary processes, and for clarifying the centrality of evolutionary theory. In this context, recent approaches have focused on identifying the central concepts that are fundamental for understanding biology in general and evolution in particular (e.g. threshold concepts; cf. Ross et al., 2010). However, after enlightening *what* has to be learned to grasp the evolution theory comes the question *how* to teach and learn this complex concept. We have addressed the latter question in our study. To promote students' learning of evolutionary issues, we have focused on particular instructional formats (i.e. learning with worked examples in combination with knowledge adapted self-explanation prompts), assuming that this approach will facilitate knowledge acquisition of evolutionary concepts.

Theoretical Background

Cognitive load theory and its consequences for instruction

The effectiveness of instructional formats is influenced by several factors. The cognitive determinant is described within the cognitive load theory (Sweller, 1988; Sweller, van Merriënboer, & Paas, 1998; van Merriënboer & Sweller, 2005). Cognitive load theory refers to a human cognitive architecture that is characterized by the working memory as a processor of information, interacting with a long-term memory in which the available knowledge is stored (Atkinson & Shiffrin, 1968). Within this model, learning (i.e. knowledge acquisition) can be described as altering long-term memory. The aim of instructions correspondingly is



to facilitate these changes in the knowledge base of learners (Kirschner, Sweller, & Clark, 2006). However, the extent of immediate changes in long-term memory is limited by the capacity of the working memory. That is because working memory is constrained in processing novel information (Baddeley, 1968). As explained in the following paragraphs, these limitations of the working memory are closely associated with the effectiveness of instruction.

Cognitive load theory assumes that every task performance imposes load on the cognitive system. Thereby, cognitive load depends on the number of elements (i.e. independent information units) and the required relations between the elements that need to be available in the working memory for understanding and learning the task. If the elements that need to be processed simultaneously exceed working memory capacity, failure to understanding will arise. One essential aspect that affects the cognitive load is the nature of the materials or tasks that has to be learned (intrinsic cognitive load; Sweller et al., 1998). The extent of intrinsic cognitive load caused by a task is determined by the expertise of the learner. That is because the complexity of the learning matter is related to the prior domain knowledge. Tasks with high element interactivity for someone might be tasks with low element interactivity for people with more expertise. Consequently, the intrinsic cognitive load is not directly affected by the instructional design itself. However, the manner in which the tasks are presented has to be processed by the working memory as well, causing additional cognitive load. Extraneous cognitive load is defined as the load that arises by instructional design features which are not necessary for knowledge acquisition, and is therefore ineffective for learning (Sweller et al., 1998). Extraneous cognitive load can thus be altered by using particular instructional interventions. Cognitive load affected by the learning processes themselves is called germane cognitive load (Sweller et al., 1998). This implies that every mental effort that contributes directly to learning also requires additional working memory capacity. The germane cognitive load reflects this effort on knowledge acquisition. Contrary to extraneous load, germane cognitive load is a useful and learning-relevant demand on the working memory.

These memory structures must be considered while creating instructional designs. Sweller et al. (1998) even suggest that the cognitive load imposed by the instruction should be the pre-eminent consideration when deciding on the application of a particular instruction. Relevant for the effectiveness of instructional formats is the additive character of the three processors: intrinsic cognitive load, extraneous cognitive load, and germane cognitive load altogether constitute working memory (van Merriënboer & Sweller, 2005). The limiting factor in this interplay is the prior knowledge. The intrinsic cognitive load of a task varies depending on prior knowledge. Additionally, depending on intrinsic cognitive load, the extraneous cognitive load needs to be adapted by altering the instructional design. So, if the intrinsic cognitive load is high, it is inevitable that we lower extraneous cognitive load in order to enable more germane load.

Due to this interaction of instruction and prior knowledge that affect learning, we assume that fitting compatibility between the learner's prior knowledge and the instructional format is crucial for the effectiveness. The expertise reversal effect described by Kalyuga, Ayres, Chandler, and Sweller (2003) confirms this

assumption. They have shown that an instruction that is beneficial for learners with little prior knowledge may lose its effectiveness or even be disadvantageous with more experience in the domain. One essential implication is that learners with different prior knowledge levels need different instructional methods. In this context, there is clear evidence that inexperienced learners benefit most from highly guided instructions (e.g. Kalyuga, Chandler, Tuovinen, & Sweller, 2001). An appropriate instructional format is critical especially when dealing with complex tasks like evolutionary theory, and it is recommended to use highly guided instructions in order to decrease extraneous cognitive load (Kalyuga, Chandler, & Sweller, 2001).

Learning with worked examples

Learning with worked examples is probably the most investigated fully-guided instruction format. Worked examples consist of a problem followed by the worked-out solution itself. All the solution details are presented in a step-by-step format to the learner, ending with a final answer to the problem. The learners can decide how long they deal with the given information because they work through the given solution by themselves.

Worked examples provide an exemplary solution to the learner by illustrating complex issues in a particular application. Using this instructional format, practicing autonomous problem solving (i.e. solving the task without any guidance) fades into the background. Learning with worked examples is more about imparting knowledge in application, and fostering the understanding of fundamental underlying principles. By linking examples to the principles, worked examples encourage principle-related learning. This was shown to be very important for knowledge acquisition (Wadouh, Liu, Sandmann, & Neuhaus, 2014). The basic understanding of the rationale of the solution in turn is a necessary condition for solving problems autonomously (Schwonke et al., 2009). If detached from the specific context, principles that have been already acquired can be applied to new problems. The benefits of learning with worked examples compared to problem-based learning, has been shown in many studies (e.g. Carroll, 1994; Cooper & Sweller, 1987; Hilbert & Renkl, 2009, Sweller & Cooper, 1985). Moreover, less learning time is required for achieving a comparable amount of learning gain (Schwonke et al., 2009; Salden, Koedinger, Renkl, Alevin, & McLaren, 2010). The effectiveness of worked examples can be explained by the cognitive load theory. Worked examples focus the learners' attention to the task and the associated correct solution. The learners can concentrate on understanding the problem solution and the underlying principle, and do not have to solve and understand the problem simultaneously. Thus, worked examples decrease extraneous cognitive load (Renkl & Atkinson, 2003; Sweller et al., 1998). However, in accordance with the expertise reversal effect, the advantage of worked examples compared to autonomous problem solving disappears with increasing expertise (Kalyuga, Chandler, & Sweller, 2001; Kalyuga et al., 2003; Renkl, 2005; Tuovinen & Sweller, 1999). Consequently, it is advisable to learn with worked examples in the initial phase of skill acquisition. Learning with problem-solving tasks should be preferred over worked examples for learning the autonomous application of the acquired knowledge. Furthermore, the structure of human cognition implies that next to the learner experience, the nature of the



matter to learn needs to be considered (Kalyuga, Chandler, & Sweller, 2001). Tasks that already have a high level of difficulty per se should not be presented in learning environments that cause additionally high demand on extraneous cognitive load. For this reason Kalyuga, Chandler, and Sweller (2001) recommended worked examples for learning situations that are high in complexity. Especially without lessons on evolution learning evolutionary theory causes high demands on the cognitive system and it is appropriate to use worked examples as instructional format.

Self-explanations and their role in learning with worked examples

The concept of self-explaining was originally described by Chi, Lewis, Reimann, and Glaser (1989). It is characterized as a constructivist learning activity which proceeds spontaneously and without any preconceived plan. By generating explanations to oneself, the process of integrating new information with existing knowledge in long-term memory is facilitated. Research has shown that the effectiveness of worked examples depends on the extent to which the learners deal with the given solution (Chi, De Leeuw, Chiu, & LaVancher, 1994; Chi, 2000; Nokes, Schunn, & Chi, 2010). It is not sufficient just to read through the worked out solution without willing to understand. The success of learning with worked examples is mainly influenced by the intensity with which the learner tries to understand the given solution, or tries to self-explain the worked example. Self-explaining is effective for understanding the underlying rationale and therefore in accordance with the theory of germane cognitive load (Paas & van Gog, 2006; Renkl & Atkinson, 2003). However, it was shown that the majority of learners are unlikely to engage spontaneously in self-explanations when learning with worked examples (Renkl, 1997). This implies that worked examples are not studied in an effective way, because free working memory capacity that arose from lowering the extraneous load is not used productively. Consequently, Renkl and Atkinson (2003) stressed the need for instructional techniques that foster effective self-explanations in order to increase germane cognitive load. One possibility would be to provide instructional aid by eliciting self-explanations while learning with worked examples. There is evidence that worked examples combined with self-explanation prompts leads to a deeper understanding than learning with worked examples alone (Chi et al., 1994; Crippen & Earl, 2007; Nokes-Malach, vanLehn, Belenky, Lichtenstein, & Cox, 2013). However, the empirical evidence for the benefits of fostering self-explanations by prompts has been mixed. For instance, Große and Renkl (2006) compared different ways of instructional support (non vs. self-explanations vs. instructional explanations) and did not find any positive effects of using both self-explanation prompts and instructional explanations. This finding has been confirmed by Lin, Atkinson, Saveney, and Nelson (2014) while comparing the different types of self-explanation prompts (non vs. prediction prompts (i.e. prompting questions before instruction) vs. reflection prompts (i.e. prompting questions after instruction)). Again, there was no advantage of using self-explanation prompts. A lack of prior knowledge may explain the missing effectiveness. If the prompted self-explanations do not fit to the learners' expertise, it is likely that they induce extraneous load instead of germane load (Paas & van Gog, 2006). This interaction of prior knowledge and self-explanation prompts was shown by Nückles, Hübner, Dümer, and Renkl (2010). At the

beginning of skill acquisition, the students benefited from the self-explanations prompts. But with increasing expertise, the prompts provided lost their effectiveness. So the expertise reversal effect was replicated for self-explanation prompts. Therefore, it would be important to consider learner's prior knowledge when adding self-explanation prompts to worked examples.

One possibility could be to tailor the prompts to the learner's knowledge level by using different kinds of prompts. Self-explanation patterns differ depending on prior knowledge in the domain (Chi et al., 1989; Kroß & Lind, 2001; Lind & Sandmann, 2003; Renkl, 1997). In contrastive approaches, Kroß and Lind (2001) as well as Lind and Sandmann (2003) investigated self-explanations of learners with high prior knowledge and low prior knowledge (we will refer to them as experts and novices, respectively). Experts tend to make inferences based on solution-relevant principles and rely on their existing knowledge to do that. They try to solve the problems by themselves and anticipate single solution steps before they use the given solution for assistance. Thus, the given solution can be perceived as some form of feedback. Additionally, the elaborations of experts go beyond the content of the worked examples more frequently. Self-explanations of experts can be categorized thereby as being solution based, connected with existing knowledge, and anticipative.

On the contrary, self-explanations of novices serve to gain a basic understanding of the example content more frequently. They tend to paraphrase the given information and rely on the knowledge provided by the worked example. When the information is presented by different sources, they spend much time on understanding their relationship. At a descriptive level, novices are characterized by the repeated reading of single text passages or solution steps. The self-explanation categories of novices can be summarized as being surface-based, stuck on example information, and reproductive.

Since experts occasionally use self-explanations at the novice level, just as novices show some self-explanations at the expert level (Mackensen-Friedrichs, 2009), it can be assumed that the spectrum of learning relevant self-explanations was completely captured by Kroß and Lind (2001). Thus, self-explanation for both the novices and experts can be considered in developing prompts. Such prompts are usually present in the form of short questions or incomplete sentences and they are related to the example content. Thereby, the prompts should be appropriately designed in a way that they evoke self-explanations that are typical for the associated knowledge level (Lind & Sandmann, 2003). Expert prompts should tend to encourage self-explanations regarding an understanding of the underlying principles and try to activate a linkage to the existing knowledge. Furthermore, expert prompts are characterized by an anticipative form asking the learners to generate the next solution step by themselves. In contrast, novice prompts elicit self-explanations dealing with example content. Prompts at the novice level ask the learners to paraphrase text passages in their own words, or make inferences based upon the information given in the text. They focus their attention on relevant information in the text, and help to connect the information presented in different sources. In a framework of the expert-novice paradigm, Mackensen-Friedrichs (2009) showed that learners benefit from worked examples which include prompts that are adjusted to their prior domain knowledge in the way



described above. Novices learning with novice prompts acquired more content knowledge compared to novices prompted at the expert level. Likewise, experts acquired more content knowledge using expert prompts than from learning with novice prompts. Furthermore, Mackensen-Friedrichs (2009) provided evidence that depending on prior knowledge level, the learning gain varies. When prompts were adapted to their knowledge level, the experts benefited more than the novices. This can be seen as a form of the Matthew effect (Merton, 1968; Walberg & Tsai, 1983), where “the rich get richer and the poor get poorer” on their knowledge acquisition.

Aim and Research Questions

The positive effect of combining worked examples with knowledge adapted prompts has only been shown in an expert-novice paradigm so far. There is hardly any research focusing on the majority in a learning group, i.e. the learners between the novices and experts with average prior knowledge. Thus, they cannot be supported with instructions that are adapted to their prior knowledge level. The aim of this study is to investigate how learners with an average knowledge level (the assignment given to the students in our study takes place normatively and is determined by test performance; operationalization is described in the “Procedure” section) can effectively be supported in learning evolutionary topics by using worked examples and self-explanation prompts. Our first research question was:

- (1) What combination of self-explanation prompts (novice- and/or expert-level) is most effective for learners with average knowledge level in order to foster the acquisition of evolutionary content knowledge when learning with worked examples?

Transition-Hypothesis. We anticipated that learners with average knowledge may be overwhelmed by exclusively learning with prompts at the expert level. Their existing knowledge about the relevant biological topics is likely not sufficient to self-explain at the expert level. At least initially, the expert prompts may cause additional extraneous load resulting in difficulties to learn adequately. However, providing exclusively novice prompts may in turn underutilize learners with average prior knowledge after a certain time so that they cannot fully exploit their cognitive potential. Thus, we hypothesized that learners with an average knowledge level will benefit from a transition within a sequence of worked examples, starting with the novice prompts and moving to the expert prompts.

In the next step, we focused on comparing the three knowledge levels (low, average, and high). Our aim was to investigate the differences in their learning gain as a result of using worked examples and knowledge adapted prompts. Because we assumed that all participants would hold a very limited scientifically correct content knowledge of evolution, the prompts were tailored to the general biological domain knowledge. Therefore self-explanation prompts were designed with respect to the characteristics described by Kroß and Lind (2001). Accordingly, our second research question was:

- (2) When prompts are adapted to prior biological knowledge, which learning group (novices, averages or experts) benefits most from learning evolutionary topics with worked examples?

Matthew-Hypothesis. We assumed that the Matthew effect (Merton, 1968, Walberg & Tsai, 1983) will become evident. Learners with high prior knowledge would benefit more from knowledge adapted worked examples than learners with average prior knowledge or low prior knowledge. Also, learners with average prior knowledge will have a greater learning gain than learners with low prior knowledge.

Methods

Sample

The sample consisted of 23 classes from 11 secondary schools (i.e. Gymnasium) from northern Germany. Altogether, N = 622 students from tenth grade aged between 15 and 17 participated in the study (53% female). Although none of the students had taken evolutionary biology course before participating in the study, it can be assumed that they already had some familiarity with this topic. Even though it is not explicitly mentioned in the curriculum, many topics in biology lessons deal with aspects of the evolutionary theory. Variation and adaptation, for example, are the basic ideas of the vertebrates unit taught in the sixth grade. Furthermore, evolutionary processes (e.g. antibiotic resistance) are quite popular in the public media. However, evolutionary theory is not specifically included in the curriculum before the tenth grade.

Within the group of participating students, the expert-novice paradigm was applied. That means the terms novices, averages, and experts are used in relative terms (cf. Chi, 2006; Kalyuga, 2007, 2008). Our study sample was used as the reference standard. The assignments for the different knowledge level groups took place on a normative way by establishing limit values in performance measure. In doing so, the relative novices (low-knowledge learners), relative averages (average-knowledge learners), and relative experts (high-knowledge learners) were compared.

Design

We used a pre-post design with three experimental groups (averages with novice or/and expert prompts) and two control groups (novice group and expert group).

Independent variable.

Knowledge level. The prior biological knowledge (i.e. content knowledge on various biological topics) of the students represents the first independent variable in this study (IV1: Knowledge level). It is normatively differentiated between three levels: low prior knowledge (novices), average prior knowledge (averages), and high prior knowledge (experts). To operationalize this quasi-experimental variable, an appropriate instrument measuring the existing biological knowledge of students in the tenth grade was required. Because the students had no evolutionary course before testing we decided to assign the students to the knowledge level groups on the basis of their general biological knowledge. However, the knowledge tests in biology usually focus on only one topic. Our aim was to investigate the general



biological content knowledge of students in the tenth grade for adequately distinguishing between the three knowledge levels of low, average, and high. For this, we developed a test reflecting the topics of the curricula up to the tenth grade. In the first step, we selected items from existing instruments (TIMSS-items by Baumert, 1998; Mackensen-Friedrichs, 2005; Schmiemann, 2010). The items were adapted linguistically and in their complexity for the learners at the tenth grade. In the next step, we created additional items dealing with topics which were not considered yet. After piloting 19 items (16 multiple choice items, two matching task items, and one open response item), considering a wide range of biological topics were selected. One sample item depicting the topic of human biology at the eighth grade is given in Table 1 (please contact the authors for more information on the test instrument).

TABLE 1

Each item was scored one point. Whereby, three items were staggered in score. Thus, the total score of the general biological content knowledge test (GBCK) ranges from 0 to 19. The reliability (measured with Cronbach's alpha) of the scale was .51. For group comparisons provided herein, the internal consistency can be regarded as still adequate (Lienert & Raatz, 1994). Thus, statements relating to group comparisons will be possible.

Type of prompting. The second independent variable is the type of prompting which is integrated in the worked examples (IV2: Type of prompting). Levels of this variable are: novice prompts, expert prompts, and the transition from novice to expert prompts (transition). The implementation of the different types of prompting was carried out on basis of the worked examples that are described in detail in the "Procedure" section below. It means that the content of worked examples did not differ within the intervention. However, the self-explanation prompts were varied. Therefore two different types of prompts were used: novice prompts and expert prompts. Based on the results of Lind and Sandmann (2003), the prompts targeted self-explanations that were identified to be typical for novices and experts, respectively. Accordingly, the novice prompts encouraged self-explanations that were shown to be effective for the novices (Kroß & Lind, 2001; Mackensen-Friedrichs, 2009). These prompts initiated paraphrasing, recourse to information given in the text, and searching for relations between information provided in different representations. However, expert prompts encouraged self-explanations that were shown to be effective for experts (Kroß & Lind 2001; Mackensen-Friedrichs, 2009). Herein prompts were integrated that caused anticipative approaches, drawing inferences, and recourse to prior knowledge. An overview of the different kind of prompts is given in Table 2.

TABLE 2

For novice prompts, all worked examples of both the sequences included only novice prompts. The same applied to the expert prompts. The transition from novice prompts to expert prompts was implemented by using novice prompts in the first two worked examples and expert prompts in the last two worked examples of the sequence. It was taken care to apply an even number of novice and expert prompts.

Time. The measurement of time (IV3: Time) is divided into the two levels Pre (performance assessment before intervention) and Post (performance assessment after intervention).

For our first research question, we used a two-way factorial design with repeated measures focusing on the averages (between factor as *Type of prompting*, within factor as *Time*). For the second research question, we used a two-way factorial design with repeated measures but with the within factor as *Time* and the between factor as *Knowledge level*.

Dependent variable.

The knowledge on evolution was operationalized by the content knowledge. In order to investigate the evolutionary knowledge before and after instruction and correspondingly the knowledge gain as a result of learning with the worked examples, one important step was the development of an appropriate instrument. Most of the existing tests concentrate on evolutionary knowledge about natural selection (e.g. CINS by Anderson, Fisher, & Norman, 2002). However, the knowledge provided by the worked example sequences is not just dealing with natural selection. Test construction was managed by the same procedure as described for the GBCK: Development started with selecting and adapting already existing items (Johannsen & Krüger, 2005; Rutledge & Warden, 2000), followed by creating additional items. After piloting and statistical item analyses in the main study, the test consisted of six multiple choice items in evolutionary biology which focused on the content of the worked examples. A sample item is shown in Table 3 (please contact the authors for more information on the test instrument).

TABLE 3

Again, each item is scored one point. Thus, the total score of the evolutionary content knowledge test (EBCK) ranged from 0 to 6 points. Like the GBCK, the reliability ($\alpha = .51$) satisfied the requirements for our planned group comparisons (Lienert & Raatz, 1994).

Procedure

We started with creating two worked example sequences on evolution. Providing a sequence of four worked examples and the design of this sequence as well as the design of each worked example is consistent with the guidelines found in the literature (for an overview, see e.g. Atkinson, Derry, Renkl, & Wortham, 2000). For topic selection, various factors were considered. The main framework for instance was given by the German curriculum. First, it is determined by the content and the level of complexity of the examples. That means that the examples were chosen with respect to the topics that were already introduced and learned in school. We also took into account the curriculum predefined for teaching evolution in the tenth grade, where the focus is on human evolution, especially on the relationships between humans. In addition, it was taken care that the worked example within a sequence were interlinked with regard to content and related to each other. This ensured that the learners not only elaborated the underlying principle of the single worked example, but also of the whole sequence. In this way, the learners had the opportunity to compare the worked examples of a sequence, finding similarities beyond surface features.

Two sequences of four worked examples were implemented as intervention. Within a sequence, all worked examples had the same underlying evolutionary principle that was structured in three solution steps. In accordance with the curriculum, the underlying principle of the first sequence was “homology and analogy” with the solution steps (I) Consideration of homologous traits, (II) Distinction between derived and ancestral homologies, and (III) Conclusion on relationship. In the second sequence, the principle of “selection” was relevant to all the solutions. In line with the core concepts formulated by Opfer et al. (2012), we determined the following solution steps for this principle: (I) Looking at differences, (II) Looking at the chances of survival and reproduction, and (III) Looking at the consequences on biological fitness. In order to stress the importance, the three solution steps were graphically highlighted in all worked examples.

Both sequences were introduced by an informational text that provided learners with the basic knowledge relevant for understanding the underlying principles and the belonging solution steps. In the “homology and analogy”-sequence, the worked examples, focused on identifying relationships based on homologous traits. The first worked example was about the relationships of the vertebrates. The problem to be solved was: “How does the relationships of the vertebrates look like?”. During problem solution, the family tree of the vertebrates was constructed. Thereby classification was based on morphological traits. In order to motivate the students, Besterman and Baggott La Velle (2007) suggested that it is functional to use the context of human evolution, and the curriculum also focuses on human evolution. Accordingly, the next two worked examples dealt with the relationships between humans (“What is the relationship between humans and great apes?” and “What was the role of Neanderthals in the evolution of modern humans?”). The complexity of problem solution grew because it was no longer sufficient to look at the morphological traits alone. Scrutinizing the relationships take place at the molecular level. The relevance of the genetic basis for the differences between species is carved out (cf. Kalinowski, Leonard, & Andrews, 2010). The last worked example of this sequence was concerned with the phenomenon that similarities do not automatically refer to relationship (“How closely related are rabbits and hyraxes?”). The first worked out example of the “selection”-sequence was about speciation in general. The relating problem formulation was: “How do species originate?”. This worked example served to clarify main conditional factors of natural selection (i.e. variation, heredity, and differential reproduction and survival) and the relevant factors for speciation (i.e. separation of sexual reproduction). The second worked example illustrated the mechanism of sexual selection (“How can the sexual dimorphism of the blue peafowl be explained?”). Again with respect to Besterman and Baggott La Velle (2007) and the curriculum, the mechanism of sexual selection and natural selection was presented in the context of human evolution in the last two worked examples (“How can we explain that men are much more physically aggressive than women?” and “How could bipedalism and loss of functional body hair in hominids evolve?”).

Data were collected about four weeks before (Pre) and immediately after (Post) the students learned with the worked example sequences. The pre-testing included two different tests: the general biological content knowledge test (GBCK;

to sort the students into the three different knowledge groups; see next paragraph), and the evolutionary biological content knowledge test (EBCK). The post-test consisted of the EBCK alone. In this way, we were able to conclude about the learning success by comparing the evolutionary knowledge before learning with the worked examples and afterwards. In the time between pre-test and intervention, the teachers did not answer questions referring to the items. Learning time on the worked examples was not limited. However, the sequences were constructed to be solved in about 90 minutes. During the intervention, it was up to the learners to make sketches and notes, to skip backwards and to underline text.

Based upon the GBCK results, we have sorted the learners into three groups of prior knowledge levels: low level (novices), average level (averages) and high level (experts). We used the 35th percentile and the 62nd percentile of the GBCK scores to differentiate between the three groups. Thus, the knowledge level was used as a quasi-experimental between variable. All learners worked on two sequences of four worked examples dealing with evolutionary topics. Depending on prior knowledge level, the prompts of the worked examples were varied. The novice and expert groups were exclusively prompted according to their knowledge level with novice prompts and expert prompts, respectively. Based on previous research, it can be expected that the students achieved their highest possible learning outcome under these prompting conditions (Mackensen-Friedrichs, 2009). In this way, the novices and experts served as control groups that can be compared to an appropriate average group. The learners with average knowledge were randomly assigned to the different prompting conditions (novice level vs. expert level vs. transition from novice to expert level). For the first research question, we investigated the influence of this experimental between variable on the evolutionary biological content knowledge of averages. The aim was to determine the best prompting condition for the learners on average knowledge level. Using these findings the further aim of this study was to assess how far knowledge adjusted worked examples facilitate learners at all knowledge levels. Therefore, we analyzed and compared the learning outcome of the three knowledge levels.

Results

Preliminary analysis

The GBCK was used to assign the students to one of the three knowledge levels and therefore to operationalize the quasi-experimental independent variable. According to their scores, we identified students' prior knowledge of general biology (<35%: Low prior knowledge (novices); 35-62%: Average prior knowledge (averages); >62%: High prior knowledge (experts)). The result of one-way ANOVA reveals a significant effect of the group ($F(2,417) = 512.50, p < .001, \eta^2 = .71$), indicating that the overall means differed across groups. Because of missing homoscedasticity this effect was confirmed with the Welch test ($t(111) = 920.52, p < .001$). Post-hoc tests (with Games-Howell adjustment; Field, 2009) shows that novices ($N = 54, M = 7.05, SD = 1.11$) significantly differed from the averages ($N = 312, M = 11.94, SD = 1.79; p < .001, d = 2.88$) and the experts ($N = 54, M = 17.30, SD = 1.37; p < .001, d = 8.29$). Likewise, averages significantly differed from the

experts ($p < .001$, $d = 3.1$). Overall, it can be assumed that the three samples were representative of the relevant populations, and the GBCK is sufficient for differentiating between the three biological knowledge levels.

Research question 1

To answer the first research question, we examined the three groups of averages who learn with different types of prompts (novice, experts, and transition from novice to expert prompts). Figure 1 shows the mean performance of EBCK before (Pre) and after (Post) learning with worked examples.

FIGURE 1

The assumptions for ANOVA were met. Looking at the evolutionary knowledge before instruction, results of one-way ANOVA indicate that the effect of the *Group* was not significant ($F(2,309) = 1.01$, $p = \text{n.s.}$). This implies that the evolutionary biological content knowledge of all three groups did not significantly differ (novice prompts: $M_{\text{Pre}} = 2.79$, $SD_{\text{Pre}} = 1.10$; expert prompts: $M_{\text{Pre}} = 2.91$, $SD_{\text{Pre}} = 1.18$; transition from novice to expert prompts: $M_{\text{Pre}} = 2.68$, $SD_{\text{Pre}} = 1.10$). The means and standard deviations for each group after instruction show that under all prompting conditions, the averages performed better than before (novice prompts: $M_{\text{Post}} = 4.53$, $SD_{\text{Post}} = 1.26$; expert prompts: $M_{\text{Post}} = 4.56$, $SD_{\text{Post}} = 1.26$; transition from novice to expert prompts: $M_{\text{Post}} = 4.61$, $SD_{\text{Post}} = 1.36$). A 3 (Prompting condition) \times 2 (Time) ANOVA with repeated measurement reveals a significant main effect of *Time* ($F(1,308) = 400.56$, $p < .001$, $\eta^2 = .57$). This indicates an increasing mean value of the overall evolutionary knowledge. Knowing that at least one group had a significant evolutionary knowledge gain, the simple effect of *Time* was determined. For all three prompting conditions, this effect was significant (novice prompts: $t(418) = 43.16$, $p < .001$, $\eta^2 = .82$; expert prompts: $t(418) = 44.26$, $p < .001$, $\eta^2 = .83$; transition from novice to expert prompts: $t(418) = 40.73$, $p < .001$, $\eta^2 = .80$). Thus, the students showed a significant learning success in all groups. However, the main effect of *Prompting condition* was not significant ($F(2,308) = .28$, $p = \text{n.s.}$). Accordingly, the interaction effect of *Prompting condition* and *Time* was not significant either ($F(2,308) = .80$, $p = \text{n.s.}$). Thus, the learning success of the averages did not differ between the three types of prompts. For prompt averages adjusted to their existing knowledge, there was no type of prompting preferable.

Research question 2

To answer the second research question, we looked at the different knowledge levels of the participants. For the averages, we will hereinafter no longer distinguish between the different prompting conditions because they have all been knowledge adapted. This group of averages was compared with the novices and the experts (who were also prompt adapted to their knowledge level). The EBCK performances of the three knowledge level groups are shown in Figure 2.

FIGURE 2

The assumptions for ANOVA were met. A one-way ANOVA reveals that at least two of the three knowledge level groups significantly differed in their prior knowledge of evolutionary biology ($F(2,417) = 102.82$, $p < .001$, $\eta^2 = .33$). Because of similar

group variance but very different sample sizes ($N_{\text{novices}} = 54$, $N_{\text{average}} = 312$, $N_{\text{experts}} = 54$), Hochberg's GT2 procedure was used for post-hoc tests (Field, 2009). Similar to their general biological knowledge, novices ($M_{\text{Pre}} = 1.56$, $SD_{\text{Pre}} = .82$) had a significantly less evolutionary knowledge compared to averages ($M_{\text{Pre}} = 2.80$, $SD_{\text{Pre}} = 1.13$; $p < .001$, $d = 15.72$), and experts ($M_{\text{Pre}} = 4.51$, $SD_{\text{Pre}} = 1.04$; $p < .001$, $d = 20.35$). The same applied to the averages and experts ($p < .001$, $d = 21.64$). Learners of all knowledge levels showed higher evolutionary knowledge after learning with the worked examples (novices: $M_{\text{Post}} = 4.04$, $SD_{\text{Post}} = 1.23$; averages: $M_{\text{Post}} = 4.57$, $SD_{\text{Post}} = 1.29$; experts: $M_{\text{Post}} = 5.42$, $SD_{\text{Post}} = .74$). A 3 (Knowledge level) \times 2 (Time) ANOVA with repeated measurement again reveals a significant main effect of *Time* ($F(1,416) = 291.15$, $p < .001$, $\eta^2 = .41$). This means that the knowledge gain was significant in the overall means. The simple effect of *Time* indicates that the growth of evolutionary knowledge was significant for all knowledge level groups (novices: $t(418) = 23.49$, $p < .001$, $\eta^2 = .57$; averages: $t(418) = 74.16$, $p < .001$, $\eta^2 = .93$; experts: $t(418) = 41.70$, $p < .001$, $\eta^2 = .81$). Looking at the main effect of *Knowledge level* reveals a significant difference in the overall means ($F(2,416) = 85.09$, $p < .001$, $\eta^2 = .29$; novices: $M = 2.80$, $SD = .12$; averages: $M = 3.68$, $SD = .05$; experts: $M = 4.97$, $SD = .12$). This effect is reflected in the fact that experts as a group had increased evolutionary knowledge compared to averages and novices. Averages in turn showed higher evolutionary knowledge than the novices. Furthermore, the interaction effect of *Knowledge level* and *Time* also became significant ($F(2,416) = 14.78$, $p < .001$, $\eta^2 = .07$). These results show that the learning success significantly differed depending on prior biological knowledge. Focusing just on the mean learning success (i.e. the knowledge gain calculated by building the difference of Pre- and Post-test scores) shows that contrary to our expectation, the novices ($M_{\text{dif}} = 2.49$, $SD_{\text{dif}} = 1.52$) outperformed the averages ($M_{\text{dif}} = 1.77$, $SD_{\text{dif}} = 1.56$), who in turn outperformed the experts ($M_{\text{dif}} = 0.91$, $SD_{\text{dif}} = 1.14$). In accordance with the results presented above, a one-way ANOVA regarding the learning success indicates a significant effect of *Group* ($F(2,416) = 14.78$, $p < .001$, $\eta^2 = .07$). In order to examine which groups significantly differed in their learning success, post-hoc contrasts were calculated. Because of a missing group variance, Games-Howell adjustment was used (Field, 2009). Substantial differences were observed between all the groups. The learning success of novices in evolution was significantly higher compared to the averages ($p < .01$, $d = 6.43$) and the experts ($p < .001$, $d = 7.75$). The same applied to the comparison of averages and experts ($p < .001$, $d = 7.72$).

Discussion and Implications

Based on the scores obtained in the GBCK, the students were assigned to one of the three knowledge level groups. In this way, the preliminary analysis showed that the novice, averages, and experts significantly differed in their performance. Learners of all knowledge levels showed an increase in content knowledge in evolution when working with the prompted worked examples. Based on formal research results, novices and experts exclusively learned with knowledge adapted prompts which were identified to facilitate knowledge acquisition most effectively. Thus, it can be assumed that students achieved their highest possible learning success. To identify the most effective type of prompting for averages, we compared the three conditions of learning with novice prompts only, expert



prompts only, and a transition from novice to expert prompts. For our first research question, results show that the learning success of averages is not influenced by the type of prompting. Contrary to our expectation of the transition-hypothesis, this finding suggests that all kinds of prompts foster knowledge acquisition similarly. There is no type of self-explanation prompt preferable for the averages. Viewing the current findings in the light of the cognitive load theory, it can be deduced that germane cognitive load was equally induced under all prompting conditions. Being exclusively prompted at the novice level or expert level seemed to cause no additional extraneous cognitive load. Learners with average knowledge appeared to be able to self-explain at the novice level as well as at the expert level without a loss of effectiveness. One explanation could be that when the novice prompts are not sufficient, the learners are able to switch to self-explanations at the expert level to gain an understanding of the underlying principles. Simultaneously, the averages provide additionally self-explanations at the novice level to understand the example content when exclusively prompted at the expert level. We assume that the self-explanation characteristic of the averages is a mixture of the novice and expert patterns. Depending on complexity of the subject matter and the compelling nature of the worked examples, the averages seemed to switch back and forth from self-explanations at the novice and expert level, respectively. However, this assumption needs to be substantiated with additional research. We therefore examined this aspect in an associated study where self-explanation patterns of learners with average knowledge were analyzed with think-aloud protocols.

Comparing the learners at the three knowledge levels, learning with knowledge adapted prompts revealed that the worked examples were most effective for the novices and least effective for the experts. The expected Matthew effect was not reliable. For novices and averages, learning with worked examples combined with knowledge adapted prompts seemed to be highly suitable for learning evolutionary theory effectively. Experts hardly benefited of learning with worked examples. These findings strongly support the redundancy effect within the cognitive load theory (cf. Bobis, Sweller, & Cooper, 1993; Mayer, Heiser, & Lonn, 2001; Nückles et al., 2010; Sweller, 2006). This means that the less knowledgeable learners need additional help provided by the worked examples and self-explanation prompts. For the experts, it may be redundant information that needs to be processed additionally in working memory and therefore increases extraneous cognitive load. In accordance with the expertise reversal effect (Kalyuga et al., 2003) it is conceivable that they would obtain a greater learning success when learning evolution is facilitated by other instructional formats. However, we did not include other instructions like autonomously problem solving. Thus, there is a lack of comparison and it is not possible to verify this assumption in our study.

Although the internal consistency of the GBCK and EBCK fulfill the requirements for group comparisons (Lienert & Raatz, 1994), the results are limited due to the relatively low reliability of the test instruments. Regarding the GBCK it was not expected to be otherwise because this test covered a large range of biological topics. However, the internal consistency of the EBCK was similarly low. This could be because the test only consisted of six items which directly impacts the reliability. Although we were able to analyze the knowledge gain of the learners,

the reliability of the EBCK is not satisfactory, and the test should be improved for further research. In this context, it would be interesting to expand the construct of evolutionary knowledge, since an understanding of evolution is not displayed only by the content knowledge. Regarding learning with worked examples, it would also make sense to assess problem solving abilities and how they emerge. Moreover, it may be advisable to investigate the cognitive load not only by the task performance, but also by subjective techniques. With respect to our research questions, we did not assess this additional variable in our study.

A further limitation of this study is that our findings are not generalizable. In the light of the unconfirmed Matthew hypothesis, which is inconsistent with the findings of Mackensen-Friedrichs (2009), the empirical evidence for the effectiveness regarding all knowledge levels is mixed at best. The statements here are strongly related to the content evolution. However, this is far from weakening the conclusion of Kalyuga et al. (2001) that the instructional format should fit to the complexity of the matter to learn, but moreover reinforces it. Future studies should therefore focus on other complex biological topics by considering the current findings and comparing the instructional combination to other instructional formats for all three knowledge levels. In doing so, it should be possible to find the most effective instructional format for all students in classroom.

Despite the limitations, our findings are particularly useful for the implementation of prompted worked examples for learning evolution in school. In order to prepare students for dealing with the changing world and building up a critical reflection with everyday life questions, an adequate knowledge of evolution is indispensable. Especially for complex issues like evolutionary theory, the prior knowledge has to be considered. However, the implementation of such internal differentiation is organizationally expensive, and thereby rarely done during the lessons (Wischer, 2008). To ensure a transposition in lesson, the differentiation of learning opportunities needs to remain practicable. Worked examples combined with self-explanation prompts accomplish this requirement. Worked examples can be tailored to the existing biological knowledge without significant expense by merely integrating the self-explanation prompts appropriate to the prior knowledge. The content of the worked examples can remain unchanged. According to our results, learners with average prior knowledge did not require explicit measures of differentiation when working with worked examples. The different types of prompts did not cause considerable differences in their learning success. Thus, there are only two variants of self-explanation prompts relevant for evolutionary lessons, namely exclusive prompts at the novice level and exclusive prompts at the expert level. In this manner, the organizational effort is relatively low. Nonetheless, account should be taken on the fact that the experts possibly would have performed better using another instructional format than worked examples for learning evolution, although they were adapted to their knowledge via self-explanation prompts on expert level. Under these conditions, it seems to make sense to use evolutionary worked examples just for the novices and averages. In this case only the exclusive novice prompts needs to be applied. For the expert it may be more preferable to use an instructional format that is less guided than worked examples.



Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Charlotte Neubrand is a Doctoral Candidate, Department Biology Education, IPN - Leibniz Institute for Science and Mathematics Education, Kiel, Germany.

Christoph Borzikowsky is a Doctor of Philosophy, Research Scientist at the Institute of Medical Informatics and Statistics, University Medical Center, Schleswig-Holstein, Kiel, Germany.

Ute Harms is a Professor, Head of the Department Biology Education, IPN - Leibniz Institute for Science and Mathematics Education, Kiel, Germany.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG [German Research Foundation], No. GZ MA 3912/3-1). We would like to thank Devasmita Chakraverty for proof-reading this article.

References

- 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.
- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, 70(2), 181-214. doi: 10.3102/00346543070002181
- Atkinson, R., & Shiffrin, R. (1968). Human memory: A proposal system and its control processes. In K. Spence & J. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89-195). New York: Academic.
- Baddeley, A. D. (1968). *Working memory*. Oxford: Oxford University Press.
- Basel, N., Harms, U., & Prechtel, H. (2013). Analysis of students' arguments on evolutionary theory. *Journal of Biological Education*, 47(4), 192-199. doi: 10.1080/00219266.2013.799078
- Basel, N., Harms, U., Prechtel, 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. doi: 10.1080/00219266.2013.849286
- Baumert, J. (Ed.) (1998). *Testaufgaben Naturwissenschaften TIMSS 7./8. Klasse (Population 2)* [Test items science TIMSS 7./8. grade (population 2)]. Berlin: Max-Planck-Institut für Bildungsforschung.
- Besterman, H., & Baggott La Velle, L. (2007). Using human evolution to teach evolutionary theory. *Journal of Biological Education*, 41(2), 76-81. doi: 10.1080/00219266.2007.9656066
- 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.
- Bobis, J., Sweller, J., & Cooper, M. (1993). Cognitive load effects in a primary-school geometry task. *Learning and Instruction*, 3, 1-21.
- Brumby, M. (1979). Problems in learning the concept of natural selection. *Journal of Biological Education*, 13(2), 119-122.
- Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology*, 86, 360-367. doi: 10.1037/0022-0663.86.3.360

- Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161-238). Mahwah, NJ: Lawrence Erlbaum.
- Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of expertise and expert performance* (pp.21-30). New York: Cambridge University Press.
- Chi, M. T. H., De Leeuw, N., Chiu, M., & LaVanher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439-477.
- Chi, M. T. H., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13(2), 145-182.
- Cooper, G., & Sweller, J. (1987). The effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, 79, 347-362. doi: 10.1037/0022-0663.79.4.347
- Crippen, K. J., & Earl, B. L. (2007). The impact of web-based worked examples and self-explanation on performance, problem solving, and self-efficacy. *Computers & Education*, 49, 809-821. doi: 10.1016/j.compedu.2005.11.018
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage Publications.
- Gregory, T. R. (2009). Understanding natural selection: Essential concepts and common misconceptions. *Evolution: Education & Outreach*, 2, 156-175. doi: 10.1007/s12052-009-0128-1
- Große, C. S., & Renkl, A. (2006). Effects of multiple solution methods in mathematics learning. *Learning and Instruction*, 16, 122-138. doi: 10.1016/j.learninstruc.2006.02.001
- Großschedl, J., Konnemann, C., & Basel, N. (2014). Pre-service biology teachers' acceptance of evolutionary theory and their preference for its teaching. *Evolution: Education and Outreach*, 7(18), 1-16. doi: 10.1186/s12052-014-0018-z
- Hilbert, T. S., & Renkl, A. (2009). Learning how to use a computer-based concept-mapping tool: Self-explaining examples helps. *Computers in Human Behavior*, 25, 267-274. doi: 10.1016/j.chb.2008.12.006
- Johannsen, M., & Krüger, D. (2005). Schülervorstellungen zur Evolution - eine quantitative Studie [Students conceptions of evolution - a quantitative approach]. *IDB - Berichte des Institutes für Didaktik der Biologie*, 14, 23-48.
- Kalinowski, S. T., Leonard, M. J., & Andrews, T. M. (2010). Nothing in evolution makes sense except in the light of DNA. *CBE - Life Sciences Education*, 9, 87-97. doi: 10.1187/cbe.09-12-0088
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychological Review*, 19, 509-539. doi: 10.1007/s10684-007-9054-3
- Kalyuga, S. (2008). Relative effectiveness of animated and static diagrams: An effect of learner prior knowledge. *Computers in Human Behaviour*, 24, 852-861. doi: 10.1016/j.chb.2007.02.018
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23-31. doi: 10.1207/S15326985EP3801_4
- Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and efficiency of instructional guidance. *Educational Psychology: An International Journal of Experimental Educational Psychology*, 21(1), 5-23. doi: 10.1080/01443410124681
- Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, 93(3), 579-588. doi: 10.1037/0022-0663.93.3.579
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, experimental, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75-86. doi: 10.1207/s15326985ep4102_1
- Kroß, A., & Lind, G. (2001). Einfluss von Vorwissen auf Intensität und Qualität des Selbsterklärens beim Lernen mit biologischen Beispielaufgaben [The impact of prior knowledge on the intensity and quality of self-explanations during studying worked-out examples from the domain of biology]. *Unterrichtswissenschaft: Zeitschrift für Lernforschung*, 29(1), 5-25.
- Lienert, G. A., & Raatz, U. (1994). *Testaufbau und Testanalyse* [test structure and test analysis]. Weinheim: Beltz.
- Lin, L. L., Atkinson, R. K., Saveney, W. C., & Nelson, B. C. (2014). Effects of visual cues and self-explanation prompts: empirical evidence in a multimedia environment. *Interactive Learning Environments*. doi: 10.1080/10494820.2014.924531
- Lind, G., & Sandmann, A. (2003). Lernstrategien und Domänenwissen [Learning strategies and domain knowledge]. *Zeitschrift für Psychologie*, 211(4), 171-192. doi: 10.1026//0044-3409.211.4.171



- Mackensen-Friedrichs, I. (2005). Förderung des Expertiseerwerbs durch das Lernen mit Beispielaufgaben im Biologieunterricht der Klasse 9 [Enhancing acquisition of expertise by learning with worked examples in biology lesson of ninth grade]. *Doctoral dissertation, University of Kiel*. Retrieved from http://e-diss.uni-kiel.de/diss_1303/
- Mackensen-Friedrichs, I. (2009). Die Rolle von Selbsterklärungen aufgrund vorwissensangepasster, domänenspezifischer Lernimpulse beim Lernen mit biologischen Beispielaufgaben [The role of self-explanations caused by domains-specific learning-stimuli that are adapted to the learners pre-knowledge while learning with biological worked-examples]. *Zeitschrift für Didaktik der Naturwissenschaften*, *15*, 155-172.
- Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, *93*(1), 187-198. doi: 10.1037/0022-0663.93.1.187
- Mayr, E. (1982). *The growth of biological thought: Diversity, Evolution, and Inheritance* Cambridge: Harvard University Press.
- Merton, R. K. (1968). The Matthew Effect in Science. *Science*, *159*(3810), 56-63.
- National Research Council and National Academy of Sciences. (2012). *Thinking Evolutionarily: Evolution Education Across the Life Sciences. Summary of a Convocation*. Steve Olson, Rapporteur. Planning Committee on Thinking Evolutionarily: Making Biology Education Make Sense. Board on Life Sciences, Division on Earth and Life Studies, National Research Council, and National Academy of Sciences. Washington, DC: The National Academies Press.
- Nehm, R. H., Poole, T. M., Lyford, M. E., Hoskins, S. G., Carruth, L., Ewers, B. E., & Colberg, P. J. S. (2009). Does the segregation of evolution in biology textbooks and introductory courses reinforce students' faulty mental models of biology and evolution? *Evolution: Educations and Outreach*, *2*, 527-532. doi: 10.1007/s12052-008-0100-5
- Nehm, R. H., & Reilly, R. (2007). Biology majors' knowledge and misconceptions of natural selection. *BioScience*, *57*(3), 263-272.
- Nehm, R. H., & Schonfeld, I. S. (2007). Does increasing biology teacher knowledge of evolution and the nature of science lead to greater preference for the teaching of evolution in schools? *Journal of Science and Teacher Education*, *18*, 699-723. doi: 10.1007/s10972-007-9062-7
- Nokes, T. J., Schunn, C. D., & Chi, M. T. H. (2010). Problem solving and human expertise. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International Encyclopedia of Education* (Vol. 5, pp. 265-272). Oxford: Elsevier
- Nokes-Malach, T. J., VanLehn, K., Belenky, D. M., Lichtenstein, M., & Cox, G. (2013). Coordinating principles and examples through analogy and self-explanation. *European Journal of Psychology of Education*, *28*, 1237-1263. doi: 10.1007/s10212-012-0164-z
- Nückles, M., Hübner, S., Dümer, S., & Renkl, A. (2010). Expertise reversal effects in writing-to-learn. *Instructional Science*, *38*(3), 237-258. doi: 10.1007/s11251-009-9106-9
- 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. doi: 10.1002/tea.21028
- Paas, F., & van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction*, *16*, 87-91. doi: 10.1016/j.learninstruc.2006.02.004
- Renkl, A. (1997). Learning from worked-examples: A study on individual differences. *Cognitive Science*, *21*(1), 1-29.
- Renkl, A. (2005). The worked-out-example principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 229-246). Cambridge, UK: Cambridge University Press.
- Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist*, *38*(1), 15-22. doi: 10.1207/S15326985EP3801_3
- Ross, P. M., Taylor, C. E., Hughes, C., Whitaker, N., Lutze-Mann, L., Koford, M., & Tzioumis, V. (2010). Threshold concepts in learning biology and evolution. *Biology International*, *47*, 47-54.
- Rutledge, M. L., & Warden, M. A. (2000). Evolutionary theory, the Nature of Science & high school biology teachers: critical relationships. *The American Biology Teacher*, *62*(1), 23-31. doi: 10.1662/0002-7685(2000)062[0023:ETTNOS]2.0.CO;2
- Salden, R., Koedinger, K. R., Renkl, A., Aleven, V., & McLaren, B. M. (2010). Accounting for beneficial effects of worked examples in tutored problem solving. *Educational Psychology Review*, *22*(4), 379-392. doi: 10.1007/s10648-010-9143-6

- Schmiemann, P. (2010). *Modellierung von Schülerkompetenzen im Bereich des biologischen Fachwissens* [Modeling of student competencies in the field of biological content knowledge]. Berlin: Logos.
- Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Alevon, V., & Salden, R. (2009). The worked-example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior*, *25*, 258-266. doi: 10.1016/j.chb.2008.12.011
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257-285.
- Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction*, *16*, 165-169. doi:10.1016/j.learninstruc.2006.02.005
- Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2*(1), 59-89. doi: 10.1207/s1532690xci0201_3
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251-296. doi: 1040-726X/98/0900-0251\$15.00/0
- Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology*, *91*, 334-341. doi: 10.1037/002-0663.91.2.334
- Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*(2), 147-177. doi: 10.1007/s10648-005-3951-0
- Wadoudh, J., Liu, N., Sandmann, A., & Neuhaus, B. J. (2014). The effect of knowledge linking levels in biology lessons upon students' knowledge structure. *International Journal of Science and Mathematics Education*, *12*(1), 25-47.
- Walberg, H. J., Tsai, S.-L. (1983). Matthew effects in education. *American Educational Research Journal*, *20*(3), 359-373.
- Wischer, B. (2008). Binnendifferenzierung ist ein Wort für das schlechte Gewissen des Lehrers [Internal differentiation is a word for the guilty conscience of a teacher]. *Erziehung und Unterricht*, *158*(9-10), 714-722.
- Yates, T. B., & Marek, E. A. (2014). Teachers teaching misconceptions: a study of factors contributing to high school biology students' acquisition of biological evolution-related misconceptions. *Evolution: Education & Outreach*, *7*(1), 7.

Figures

Figure 1. Mean EBCK test scores before (Pre) and after (Post) instruction for the learners of average knowledge level in the novice prompts, expert prompts, and transition from novice to expert prompts group. Standard error bars represent standard error of mean (SEM).

Figure 2. Mean EBCK test scores (+/-SEM) before (Pre) and after (Post) instruction for the Novices, Averages, and Experts.

Tables

Table 1. Sample item GBCK.

Table 2. Self-explanation prompts adapted to low and high prior knowledge.

Table 3. Sample item EBCK.

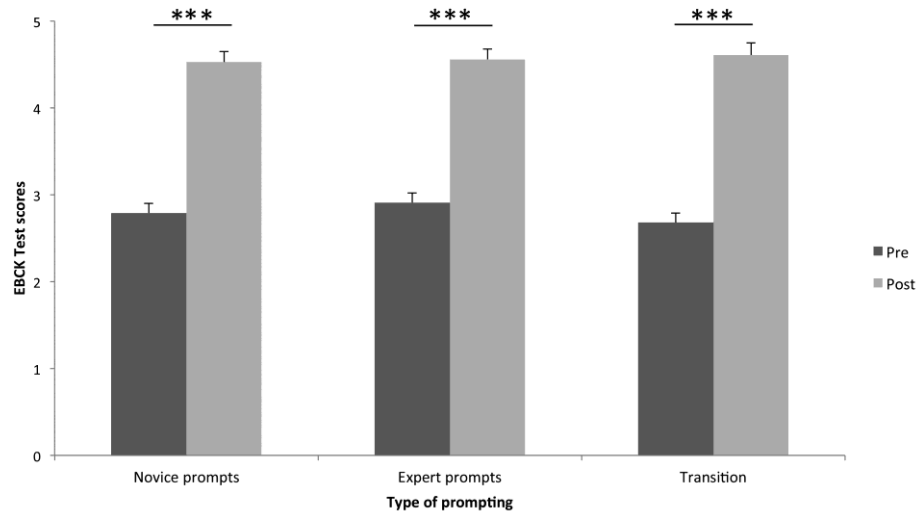


Figure 1. Mean EBCK test scores before (Pre) and after (Post) instruction for the learners on average knowledge level in the novice prompts, expert prompts, and transition from novice to expert prompts group. Standard error bars represent standard error of mean (SEM).

*** $p < .001$

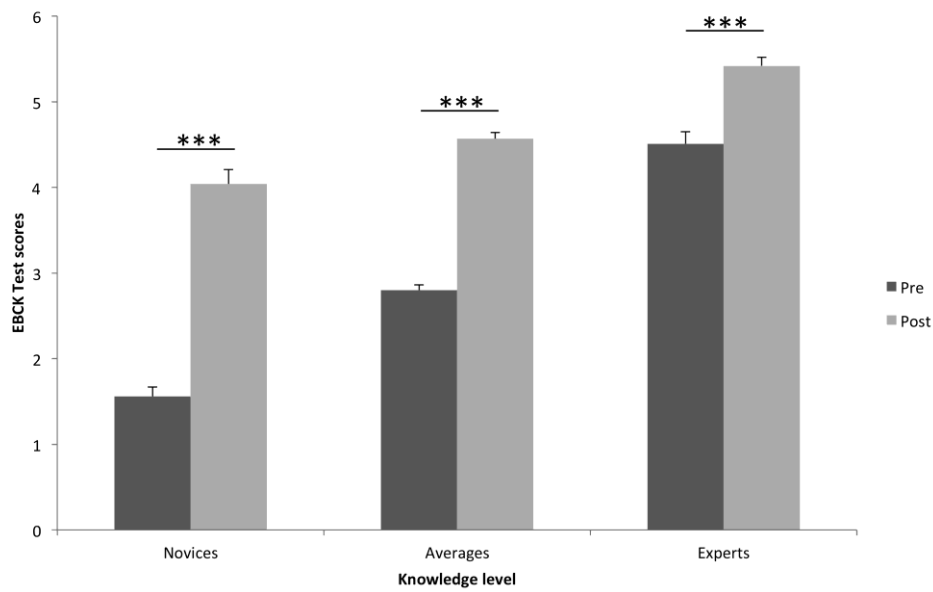


Figure 2. Mean EBCK test scores (+/- SEM) before (Pre) and after (Post) instruction for the Novices, Averages, and Experts.

*** $p < .001$

Table 1
Sample item GBCK

Which of the following is the task of tendons?

- Tendons transport stimuli from the brain to the muscles.
- Tendons keep the muscle fibers in a muscle together.
- Tendons transfer the power of the muscles onto the bones.
- Tendons stabilize two bones in a joint.

Table 2
Self-explanation prompts adapted to low and high prior knowledge

Novice prompts	Expert prompts
Paraphrase <i>“Now I know that...”</i>	Anticipative approach <i>“I think for myself before I go on reading.”</i>
Retrieval of knowledge provided by the worked example <i>“In the introduction I read about biological fitness that...”</i>	Retrieval of prior knowledge which is not provided by the worked examples <i>“Other mammals are...”</i>
Searching for relations <i>“I can find it in the following figure.”</i>	Solution based inferences <i>“If fishes are a monophyletic group...”</i>

Table 3
Sample item EBCK

The wing of an insect and the wing of a bat are analogous organs. This statement...

- is correct, because the wings have the same layout and fulfill the same function.
- is correct, because the wings have a different layout and fulfill the same function.
- is incorrect, because they are homologous organs.
- is incorrect, because bats are no insects and hence cannot be compared with them.
