ABSTRACT
Teaching and learning evolution is challenging. Biology education research shows that the underlying evolutionary concepts are poorly understood among students. This prevents a meaningful understanding of the central biological concepts. The instructional format of self-explanation prompts seems to be promising to respond to these difficulties. However, previous research has shown that the interaction between prior knowledge and prompts is an important factor in enhancing effective self-explanations. Nevertheless, there are hardly any studies which focus on the majority of learners in the classrooms, ie students with average prior knowledge. The aim of this study was to analyse the requirements of adaptive self-explanation prompts specifically for these learners. This should enable students at all knowledge levels to be effectively fostered in self-explaining, while they are learning evolution. We investigated the effects of three kinds of prompts for fostering self-explanations: learners with average prior knowledge were prompted with low or/and high-knowledge prompts. Our analysis of 22 verbal protocols by middle-school German students shows that the prompts evoked the intended self-explanations. The most positive impact with regard to self-explanation quality was created by the prompting condition which combines low-knowledge and high-knowledge self-explanation characteristics.

Introduction
Evolutionary biology is a consistent principle which unifies all topics in biology (Dobzhansky 1973; Nehm et al. 2009) and explains both the diversity as well as the unity of life (Dobzhansky 1973). Understanding evolutionary theory 'helps us to understand the world around us, our place in it, and ourselves' (Futuyma 1998, 6). Based on these claims, it is clear that evolutionary theory is probably the most complex theory in biology (Mayr 1982, 481), which emphasises the importance of teaching and learning evolution in biology education.

It is likely that difficulties and misunderstandings in learning about all other areas of biology will arise in the absence of a proper understanding of evolutionary theory. However, a considerable number of studies have shown that evolutionary theory is poorly understood by students (e.g. Brumby 1979; Opfer, Nehm, and Ha 2012). This is further substantiated by the fact that lines of argumentation
remain superficial when it comes to evolutionary topics (Basel, Harms, and Prechtl 2013; Basel et al. 2014). It is of particular concern that even courses in evolutionary biology may not always eliminate these deficits (Brumby 1979; Yates and Marek 2014). These findings indicate that there is a great number of difficulties in teaching and learning evolution. Consequently, a profound and meaningful understanding of the underlying evolutionary concepts is only rarely reached.

A promising approach to ensure meaningful learning is to create active-learning environments. Prince (2004) defines active learning as 'any instructional method that engages students in the learning process.' Therefore, the main characteristics of active learning are student activity and engagement. Numerous studies have shown that learner engagement is an important influencing factor in the learning outcome (e.g. Freeman et al. 2014; Prince 2004). One type of active learning is the generation of self-explanations during the learning process (Catrambone and Yuasa 2006). The learners are actively involved in the learning process and develop a meaningful understanding by generating explanations to oneself (e.g. Chi 2000; Chi et al. 1989). Particularly in view of the call that learning strategies for evolution should integrate learners’ prior knowledge with the new knowledge which is presented in instructional materials (Schilders et al. 2009), it is worth to investigate the effect of using self-explanation prompts as a specific instructional format, on learning evolution.

Theoretical background

Self-explanations and their role in learning

Self-explanation was first described by Chi et al. (1989). It refers to the constructive activity of generating explanations to oneself instead of receiving the explanations passively from a text or a teacher. Self-explanation is a knowledge-building activity whose goal is not just to memorise the information that is given by any kind of external input. Rather, self-explanation ensures that the learners engage themselves with the new information in a meaningful way. In this fashion, they are able to attain a deep and robust understanding (Chi 2000; Chi et al. 1989, 1994; Wylie and Chi 2014). Hence, learners are actively involved in the knowledge construction process by linking the information which is contained in the instructional material with their existing knowledge via the self-explanations they generate. In this sense, self-explaining is a process during which inferences are created and used to fill gaps in the learners’ knowledge and to foster the understanding of concepts and principles (Chi 2000).

Depending on the domain-specific prior knowledge, differences in the self-explanation quality become evident. These differences are reflected in the use of different self-explanation categories. Since certain categories are used relatively more often by learners with high prior knowledge (high-knowledge learners) than by learners with low prior knowledge (low-knowledge learners), the quality of self-explanations are determined in this way.

High-knowledge learners elaborate on the instructional material more intensively and deeply (e.g. Best, Ozuru, and McNamara 2004; Chi et al. 1989; Kroß and Lind 2001; Lind and Sandmann 2003; Mackensen-Friedrichs 2005; Renkl 1997). Consequently, high-knowledge learners’ explanations go beyond the given information; these learners tend to make inferences based on solution-relevant principles. To do so, they rely on their existing knowledge. Moreover, high-knowledge learners distinguish themselves from the others by their anticipative approach; they try to solve tasks by themselves without any help. In summary, high-knowledge or high-quality self-explanations can therefore be categorised as being solution based, anticipative, and connected to existing knowledge.

Low-knowledge learners, conversely, are only able to use their knowledge in a less flexible way (Mackensen-Friedrichs 2005; Sweller, van Merrienboer, and Paas 1998). Their elaborations remain on the surface, and the explanations more frequently serve to only gain a basic understanding of the given information (Best, Ozuru, and McNamara 2004; Chi et al. 1989; Kroß and Lind 2001; Lind and Sandmann 2003; Mackensen-Friedrichs 2005; Renkl 1997). Low-knowledge learners tend to paraphrase the given information; they rely on the knowledge provided by the instructional material. When the information is presented by different sources, they spend a large amount of time on
understanding the relationship between these sources. At a descriptive level, low-knowledge learners are characterised as those who re-read single text passages or statements. Low-knowledge or low-quality self-explanation categories is summarised as being surface based, strongly attached to example information, and reproductive.

Even though the quality of self-explanations varies, all self-explanation categories appear to be effective for learning. While high-knowledge learners show a greater learning gain with regard to problem-solving abilities, the learning gain of low-knowledge learners, on the other hand, refers more to content knowledge. Due to differences in prior knowledge, both groups of learners use different self-explanations. Both, however, achieve learning success (Mackensen-Friedrichs 2009).

**Fostering self-explanations**

The depth to which learners engage in the activity of self-explaining significantly influences the learning success (Chi and Bassok 1989; Chi et al. 1994). In particular, because learners often have difficulties engaging spontaneously in self-explanations (Renkl 1997), it seems to be advisable to foster self-explanations during the learning process.

An increase in the number of self-explanations is easy to achieve by simple requests and was shown to have a positive impact on knowledge acquisition (e.g. Chi et al. 1994; Nokes-Malach et al. 2013). These open-ended self-explanation prompts are more general in nature and without any expectations about the explanation. The advantage of these prompts is that the learners are free in their explanations and not restricted or influenced by preconceived ideas which are given by the prompts (Wylie and Chi 2014). However, open-ended prompts rather influence the quantity and not the quality of self-explanations; what is more prompting may not always work (eg Lin et al. 2014).

What is striking is that these previous studies normally did not consider the prior knowledge of the learners. While it was indeed found that self-explanation characteristics differ depending on prior knowledge, this was not reflected in the prompting conditions. This may explain the missing effectiveness of self-explanation prompts. If the prompted self-explanations do not fit the learners’ expertise, it is likely that students will fail to understand (Paas and van Gog 2006). This assumption was confirmed by Nückles et al. (2010), who investigated the effect of self-explanation prompts in writing learning journals. The findings showed that the benefits of self-explanation prompts varied depending on prior knowledge. At the beginning of the skill-acquisition process, the students benefited from the self-explanation prompts. With increasing expertise, however, the provided prompts lost their effectiveness. Hence, the expertise reversal effect was replicated for self-explanation prompts. This result suggests the need for the variation in self-explanation prompts within a group based on its learners’ prior knowledge.

In order to optimise the effectiveness of self-explanation prompts, the interaction of prior knowledge and prompts need to be considered. However, the effects of different kinds of prompts on different learner knowledge levels have not been studied adequately so far. In one of the few existing studies, Yeh et al. (2010) investigated how two different self-explanation prompts affect the learning outcome of low- and high-knowledge learners. They have shown that the benefits of self-explanation prompts vary in relation to expertise differences. Prompts which facilitate self-explanations that are typical of the knowledge level prove to be successful.

These findings were also supported by Mackensen-Friedrichs (2009). She designed two types of self-explanation prompts: low-knowledge prompts (LP) and high-knowledge prompts (HP). The prompts were based on the self-explanation characteristics of low-knowledge learners and high-knowledge learners, respectively. LP required the learners to paraphrase, to rely on information which was given in the instructions and to search for relations between pieces of information which were provided in different representations. HP asked the learners to anticipate the next solution step, to rely on their own prior knowledge, and to build inferences. The HP were compared to the LP with regard to the learners’ prior knowledge (high-knowledge learners vs. low-knowledge learners). High-knowledge
learners performed better when they were prompted with HP. The same applied to the low-knowledge learners who learned with LP.

The application of adaptive self-explanation prompts seems to have a promising future, even though their adaptability needs to be investigated in more detail (Wylie and Chi 2014). One aspect which has largely been disregarded is the focus on learners with average prior knowledge (average-knowledge learners). Although these learners form the majority in classrooms, there is, to our knowledge, hardly any research which focuses on this group. To address this gap, we designed the study at hand, which analyses the self-explanations of learners with average prior knowledge. Since the basis of adaptive self-explanation prompts (the selfExplanation categories) had been extensively described in the literature, we use adaptive selfExplanation prompts based on the model of Mackensen-Friedrichs (2009).

Research questions

This study takes a first step towards a stronger focus on average-knowledge learners and how they can be supported in learning evolution. The main objective of this study was to investigate how average-knowledge learners react to different adaptive selfExplanation prompts. We designed three selfExplanation prompt learning environments: the low knowledge prompt, the high knowledge prompt, and the transition prompt, which uses LP at the beginning and HP at the end. The research questions we addressed were:

1. Do differences in the distribution of the selfExplanation characteristics of average-knowledge learners arise under different prompting conditions?
2. Which selfExplanation are initiated by the different prompting conditions?
3. Do average-knowledge learners differ with regard to correct and incorrect selfExplanation under different prompting conditions?

Method

Sample

The results presented in this paper refer to a sample of participants in an intervention study. In total, the intervention study included 433 tenth-grade students aged between 15 and 17. Participation took place on a voluntary basis within the regular biology classes. The participants were normatively grouped into three groups according to their performance measure in a general biology knowledge test (Neubrand, Borzikowsky, and Harms Forthcoming): low knowledge level group (low-knowledge learners), average knowledge level group (average-knowledge learners) and high knowledge level group (high-knowledge learners). The operationalisation of the assignment is described in the ‘Procedure’ section. For the examination of the selfExplanation characteristic of learners with average knowledge, we analysed think-aloud protocols of 22 average-knowledge learners (16 females). This experimental sample consisted of students from eight different classes at five different high schools in Schleswig-Holstein, Germany. None of the students had taken evolutionary biology courses before their participation in the study.

Learning material

As the instructional format for the intervention, we decided to use worked examples. This instructional format consists of a problem followed by the worked-out solution. All the solution details are presented to the learner in a step-by-step format, ending with a final answer to the problem. We primarily chose worked examples as the instructional format for two reasons. First, because of participant prerequisites: due to their structure, worked examples are particularly suited for the initial phase of skill acquisition (Renkl 2014). This was applied to the participants in this study because they had not taken evolutionary
biology courses in school before. Second, because of evolution as a learning topic: Worked examples are recommended for learning concepts or subjects with a higher level of difficulty (Kalyuga, Chandler, and Sweller 2001); the subject of evolutionary theory fits this category based on its high complexity.

In total, we designed eight worked examples which deal with evolutionary topics. These eight worked examples were divided into two sequences in which each consist of four worked examples. In both sequences, all four worked examples show the same evolutionary principle in its application. In the course of the sequences, the complexity grows. A short introductory text was provided prior to each worked example sequence, giving the relevant definitions and the theoretical framework of the solution-relevant principle. This was to ensure that all the learners had the necessary information which they needed in order to understand the worked examples. An overview of both worked example sequences is given in Table 1.

### Design

A between subject design was used with average-knowledge learners who were randomly assigned to one of the three prompting conditions: exclusive LP, exclusive HP and transition, which started with LP and ended with high-knowledge prompts (LHP).

### Independent variables

**Knowledge level**

The students’ prior biological knowledge was used as a quasi-experimental independent variable in this study. The existing biological knowledge of the students was assessed by a knowledge test which consisted of 19 items (16 multiple-choice items, two matching-task items, and one open-response item).

### Table 1. Structure and content of the two worked example sequences.

| Principle     | Solution steps                                      | Sequence 1                                      | Sequence 2                                      |
|---------------|-----------------------------------------------------|-------------------------------------------------|------------------------------------------------|
| **Homology and analogy** | (1) Consideration of homologous traits                  | Selection                                        | (1) Looking at differences                      |
|               | (2) Distinction between derived and ancestral homologies | (2) Looking at the chances of survival and reproduction |
|               | (3) Conclusion on relationships                       | (3) Looking at the consequences on biological fitness |
| **Learning goal** | Learners are able to use the ‘homology and analogy’ principle at different organisational levels and correspondingly conclude on relationships | Learners understand the principle of ‘selection’ as a non-random process and are able to describe existing traits as the consequence of adaption processes |
| **Introductory text** | • Definition of homologous traits                     | • Darwin’s observations and conclusions          |
|               | • Definition of analogous traits                      | • Variation as an important factor affecting chances in survival and reproduction |
|               | • Definition of initial and derived characteristics   | • Definition of fitness                          |
|               | • Relationship on the basis of derived characteristics| • ‘survival of the fittest’ is declared          |
|               |                                                      | • Selection as consequence of different chances of survival and reproduction and heredity |
| **Worked example 1** | ‘The fish inside of us’                                | ‘How species originate’                         |
| **Worked example 2** | ‘The monkey inside of us’                              | ‘What women want and how men compete’           |
| **Worked example 3** | ‘The Neanderthal man inside of us’                    | ‘Aggression’                                     |
| **Worked example 4** | ‘The land of the hyraxes’                              | ‘The upright stance of the naked ape’           |
asking a wide range of biological content knowledge-based questions (Neubrand, Borzikowsky, and Harms Forthcoming). Based on the test scores, we normatively differentiated between three knowledge levels: low, average and high. This experimental study only applies to the average-knowledge learners. The factor levels ‘low’ and ‘high’ were not relevant within the framework of this analysis.

**Type of prompting**

The second independent variable was the type of prompting. This included: exclusive LP, exclusive HP and the transition from low-knowledge to high-knowledge prompts (LHP). The implementation of the different types of prompting was carried out on the basis of two inherently closed sequences of four worked examples which concerned evolutionary topics. The content of the worked examples did not differ within the intervention. Only the self-explanation prompts varied.

The instructional measures were based on the findings of Lind and Sandmann (2003), who identified self-explanation characteristics depending on the prior knowledge of students. The prompts targeted self-explanations which were shown to be typical of low-knowledge learners and high-knowledge learners, respectively. Accordingly, LP initiated paraphrasing, recourse to information given in the text and searching for relations between pieces of information provided in different representations (cf. Kroß and Lind 2001; Mackensen-Friedrichs 2009). However, HP encouraged self-explanations which are anticipative, inference-building, and with recourse to prior knowledge which was shown to be effective for high-knowledge learners (Kroß and Lind 2001; Mackensen-Friedrichs 2009). An overview of the implementation of the different kinds of prompts is given in Table 2.

For LP, all worked examples in both sequences included only LPs. The same applied to HP. The LHP was implemented using LP in the first two worked examples and HP in the last two worked examples of the sequence. An even number of LP and HP was applied. An example of a self-explanation prompt is given in the ‘Procedure’ section.

**Dependent variable**

The self-explanation characteristics of the students constitute the dependent variable. Self-explanations were collected with a standard method in expertise research: the think-aloud method according to Ericsson and Simon (1993). The think-aloud method involves the verbalisation of thoughts while working on a task without focusing on certain aspects. The resulting think-aloud protocols were analysed based on Lind and Sandmann’s (2003) detailed system of categories. Our data analysis includes all self-explanations which can be assigned to the following self-explanation categories:

- Drawing inferences: this category was scored if learners made conclusions using information given in the worked examples.
- Anticipating: this category includes all self-explanations in which learners tried to generate the next solution step autonomously without reading the worked out solution.
- Recourse to own prior knowledge: self-explanations which relate to knowledge that was not imparted by the worked examples.
- Recourse to given knowledge: self-explanations which relate to knowledge that was imparted by the worked examples before.

|                           | LP       | HP       | LHP      |
|---------------------------|----------|----------|----------|
| Paraphrasing              | 79 times | –        | 48 times |
| Recourse to information   | 33 times | –        | 22 times |
| Searching for relations   | 32 times | –        | 20 times |
| Anticipative approach     | –        | 46 times | 24 times |
| Inference building        | –        | 70 times | 26 times |
| Recourse to prior knowledge| –       | 21 times | 6 times  |
• Paraphrasing: self-explanations were classified as paraphrasing if participants reproduced text passages in their own words or restated information given in diagrams without adding new information.
• Relation building: self-explanations which indicate that learners searched for relations between pieces of information provided in different representations (e.g. text and figure).

Besides these, self-explanation statements which concerned strategies of reduction (emphasis extraction and the generation of rules by forming analogies), repeating strategies (repeated reading and reciting for the purpose of remembering) and monitoring were captured. Additionally, we documented whether the prompted self-explanations were right or wrong in relation to their content. An example of a coding is given in the ‘Procedure’ section.

Procedure

This study took place in 2012 and consisted of three phases. First, the participants’ prior knowledge was captured, according to which students were assigned to one of the three treatment conditions. The sample was determined on the basis of a knowledge test in general biology which was administered one month before the study. The knowledge level groups were identified based on the number of correctly solved items. Students with less than 35% correctly solved items were assigned to the group with low prior knowledge. Students with more than 62% correctly answered items were assigned to the group with high prior knowledge. If 35–62% of the items were answered correctly, the students were assigned to the group with average biological knowledge. Of the total 433 participants, 312 students had an average prior knowledge. These students were randomly assigned to one of the three prompting conditions. In order to analyse the self-explanation characteristics of the average-knowledge learners, 120 students were randomly selected and asked to verbalise their thoughts while working on the worked examples.

As a second step, the students who had to verbalise their thoughts were introduced to the think-aloud method. The participants were instructed using a short explanation. With the help of exercises, they subsequently practised to verbalise their thoughts without explaining, analysing or interpreting (c.f. Ericsson and Simon 1993).

Third, all participants learned with the two worked example sequences. Learning time was not limited. Based on Renkl's (1997) as well as Lind, Friege, and Sandmann's (2005) results, it can be assumed that learning time would only influence the results to a very small extent. However, both sequences of four worked examples were designed to be solved in about 90 min each. During learning, it was up to the students to make sketches and notes, to skip backwards, and to underline text.

There have been no major difficulties with the think-aloud method. The verbal protocols were extremely extensive and diverse. Only for eight students there were indications that they could not handle the think-aloud method. They spoke unclearly and gave almost no self-explanations. Moreover, 21 verbal protocols were incomplete. Given the broad amount of the verbal protocols, not all of the remaining 91 protocols were transcribed. For our data analysis, every fourth verbal protocol was transcribed. Further analysis would have not constituted significant added value and these 22 verbal protocols with a total of 13,805 self-explanations were sufficient to answer the research questions.

The transcripts were then analysed with the help of a qualitative content analysis approach (Mayring 2010). The analysis was supported by the MaxQDA software (version 10). As already described above, the coding scheme referred to Lind and Sandmann's (2003) category system. An example of a self-explanation prompt and the corresponding self-explanation as well as the coding is given in Figure 1.

Results

Initially, we analysed the self-explanation patterns of the three groups which learned under different self-explanation conditions. Therefore, all self-explanations were considered without differentiating
The example of the lemurs shows that many new and quite different species can originate out of one single species. But how do species originate?’

Student: ‘How different species can originate (1). Well, the dog was bred from the wolf (2). So you can breed new species (3).’

Coding: (1): Paraphrase; (2): Recourse to prior knowledge; (3): Inference-building.

Figure 1. Section of a worked out solution from the selection sequence with a self-explanation prompt (presented in the balloon) which should encourage paraphrasing, followed by the expressed self-explanation of a student with average prior knowledge as well as the coding.

Table 3. Mean number of self-explanations.

| Type of prompting | LC  | HC  | Total |
|-------------------|-----|-----|-------|
| LP                | 346 | 122 | 468   |
| HP                | 141 | 181 | 322   |
| LHP               | 180 | 660 | 840   |
| Total             | 667 | 963 | 1630  |

Notes: LP: exclusive low-knowledge prompts; HP: exclusive high-knowledge prompts; LHP: transition from low-knowledge to high-knowledge prompts; LC: low-knowledge characteristic; HC: high-knowledge characteristic.

between those self-explanations generated due to prompts and those generated spontaneously. The distribution of the mean number of self-explanations with low-knowledge (LC) and high-knowledge (HC) characteristics is shown in Table 3.

In order to analyse the distributions of the self-explanation characteristics, we used Chi-square tests to record the differences in the overall frequency distributions. The chi-square value was significant when comparing LP to HP ($\chi^2 (1) = 73.30, p < .001$) and LHP ($\chi^2 (1) = 344.60, p < .001$). The same applies to the comparison of HP to LHP ($\chi^2 (1) = 58.21, p < .001$). These highly significant values indicate that the type of prompting used had a significant effect on whether the learners show low-knowledge or high-knowledge self-explanation characteristics. These findings imply that the learners will show a high-knowledge self-explanation characteristic using HP or the combination of both low-knowledge and HP. However, the high-knowledge self-explanation characteristic is most pronounced under the LHP condition.

In a next step, we analysed how the average-knowledge learners respond to the different self-explanation prompts. The assessment took place for the first self-explanation following a prompt which was allocated to one of the considered categories. All the students’ statements which did not fit into one of
the categories are included in the Other category. For self-explanations which were assigned to Relation building, we differentiated between students who established a relationship between the different representations and students who just agreed that there is a relation. This was necessary because the prompts fostering relation building were the only one that worked with a yes answer and in this case, we could not ensure that the students indeed built a relation or just agreed without further reflection.

The spectrum of the self-explanations which refer to the different prompting conditions is shown in Tables 4–6. To improve comparability, the mean number of self-explanations per prompt is indicated. Self-explanations which were supposed to be elicited by the prompt are highlighted in bold print. Average-knowledge learners who exclusively learned with LP (Table 4) predominantly showed the required self-explanations. Further self-explanations mainly fall into the category of Paraphrasing. High-knowledge self-explanations represent only a very small proportion of the prompted self-explanations. What is noticeable is the high proportion of other self-explanations for the prompts which should initiate the recourse to given knowledge. These are primarily negative monitoring statements such as ‘I can’t remember’ or ‘I don’t know’.

Under the HP condition (Table 5), we see that the majority of the self-explanations belong to the required category. However, the proportion of the intended self-explanations is slightly lower compared to the LP condition, with a corresponding rise above the proportions of other statements (Other). As for the LP condition, these other statements primarily belong to negative monitoring. The proportion of the low-knowledge self-explanations is very small.

Our results show that under the LHP condition (Table 6), the average-knowledge learners mainly offered the required self-explanations. However, the distribution of the proportion referring the

### Table 4. Mean number and standard deviation of self-explanations per prompt (P: Paraphrasing; RGK: Recourse to given knowledge; R: Relation building; A: Anticipating; DI: Drawing inferences; RPK: Recourse to own prior knowledge; O: Other; ✓: Prompt without any self-explanation; ✓: Agreement) under the low-knowledge prompts condition.

|          | P        | RGK      | R         | A         | DI        | RPK       | O          | ✓         |
|----------|----------|----------|-----------|-----------|-----------|-----------|------------|-----------|
| LP       | 0.61 (0.10) | 0.04 (0.02) | 0.01 (0.02) | 0.02 (0.02) | 0.09 (0.04) | 0         | 0.11 (0.06) | 0.04 (0.07) | ✓         |
| RGK      | 0.11 (0.06) | 0.50 (0.10) | 0.03 (0.02) | 0.01 (0.02) | 0.04 (0.03) | 0.00 (0.01) | 0.25 (0.09) | 0.03 (0.04) | ✓         |
| R        | 0.14 (0.06) | 0         | 0.51 (0.14) | 0.01 (0.02) | 0.02 (0.02) | 0         | 0.08 (0.05) | 0.14 (0.10) | 0.22 (0.15) |

### Table 5. Mean number and standard deviation of self-explanations per prompt (P: Paraphrasing; RGK: Recourse to given knowledge; R: Relation building; A: Anticipating; DI: Drawing inferences; RPK: Recourse to own prior knowledge; O: Other; ✓: Prompt without any self-explanation; ✓: Agreement) under the high-knowledge prompts condition.

|          | P        | RGK      | R         | A         | DI        | RPK       | O          | ✓         |
|----------|----------|----------|-----------|-----------|-----------|-----------|------------|-----------|
| HP       | 0.03 (0.02) | 0.03 (0.02) | 0.01 (0.02) | 0.39 (0.12) | 0.14 (0.06) | 0.02 (0.02) | 0.26 (0.10) | 0.12 (0.09) | ✓         |
| DI       | 0.11 (0.05) | 0.05 (0.04) | 0.03 (0.03) | 0.09 (0.05) | 0.34 (0.09) | 0.01 (0.01) | 0.28 (0.13) | 0.09 (0.09) | ✓         |
| RPK      | 0.07 (0.07) | 0.05 (0.03) | 0.02 (0.02) | 0.07 (0.07) | 0.06 (0.08) | 0.40 (0.11) | 0.28 (0.13) | 0.05 (0.05) | ✓         |

### Table 6. Mean number and standard deviation of self-explanations per prompt (P: Paraphrasing; RGK: Recourse to given knowledge; R: Relation building; A: Anticipating; DI: Drawing inferences; RPK: Recourse to own prior knowledge; O: Other; ✓: Prompt without any self-explanation; ✓: Agreement) under the transition prompts condition.

|          | P        | RGK      | R         | A         | DI        | RPK       | O          | ✓         |
|----------|----------|----------|-----------|-----------|-----------|-----------|------------|-----------|
| LHP      | 0.48 (0.05) | 0.02 (0.02) | 0.01 (0.01) | 0.04 (0.03) | 0.18 (0.03) | 0         | 0.12 (0.06) | 0.08 (0.08) | ✓         |
| RGK      | 0.05 (0.04) | 0.47 (0.19) | 0         | 0.01 (0.02) | 0.06 (0.03) | 0.01 (0.02) | 0.42 (0.11) | 0.12 (0.09) | ✓         |
| R        | 0.12 (0.06) | 0.01 (0.02) | 0.41 (0.15) | 0.01 (0.02) | 0.04 (0.04) | 0         | 0.09 (0.07) | 0.19 (0.15) | 0.14 (0.05) |
| A        | 0.03 (0.03) | 0.02 (0.02) | 0         | 0.60 (0.13) | 0.05 (0.06) | 0         | 0.18 (0.10) | 0.13 (0.15) | ✓         |
| DI       | 0.14 (0.07) | 0.03 (0.04) | 0.03 (0.05) | 0.13 (0.05) | 0.41 (0.12) | 0.01 (0.02) | 0.16 (0.10) | 0.11 (0.11) | ✓         |
| RPK      | 0.07 (0.13) | 0.07 (0.10) | 0         | 0.02 (0.06) | 0.12 (0.13) | 0.50 (0.19) | 0.19 (0.15) | 0.02 (0.06) | ✓         |
intended self-explanations shifts: the proportions of the intended low-knowledge self-explanations decreased compared to the LP condition, whereas the proportions of the intended high-knowledge self-explanations increased compared to the HP condition. For instance, the Paraphrasing prompts increasingly caused inferences. The decrease of the intended Recourse to given knowledge self-explanations is reflected in the increase of other statements (Other). The increase of the intended high-knowledge self-explanation (i.e. Anticipating, Drawing inferences, and Recourse to own prior knowledge), on the other hand, is coupled with the decrease of other statements.

In order to determine significant differences between the groups with regard to the content-related correctness of the prompted self-explanations, we calculated a-priori contrasts. Because of the small sample size, we explored the distribution of the variables using the Shapiro–Wilk test. In each case, the sample was normally distributed. Against the background that the total number of prompts varied between the three prompting conditions, the mean number of correct and incorrect self-explanations per prompt was considered to facilitate comparability.

First, the correct self-explanations were analysed. The number of correct self-explanations was largest in the case of the LHP condition ($M = .68; SD = .11$), followed by the LP condition ($M = .57; SD = .14$). Only between HP and LHP, the number of correct self-explanations differed in a slightly significant way ($t(19) = -2.31; p < .05; d = 1.61$). The difference between LP and LHP ($p = .30$) as well as LP and HP ($p = .17$) was not significant. With regard to the incorrect self-explanations, there is an inverse relationship between the prompting conditions. The least incorrect self-explanations appeared under the LHP condition ($M = .22; SD = .07$). Significantly, more incorrect self-explanations occurred under the LP condition ($M = .36; SD = .09$; $t(19) = -3.18; p < .005; d = 1.94$) and the HP condition ($M = .44; SD = .11$; $t(19) = -4.93; p < .001; d = 2.8$). The difference between LP and HP was not significant ($p = .09$).

**Discussion**

In this study, we analysed students’ prompted self-explanations on worked examples in evolutionary biology. The focus was on learners with average prior knowledge. Although this group of learners represented the majority in the classroom, we are not aware of studies which investigate how their learning can be improved. Our objective was to address this gap by taking preliminary steps towards understanding how to elicit effective self-explanations from learners with average prior knowledge.

In summary, we found that the prompting condition has an impact on the self-explanation pattern beyond the prompted self-explanations. Not only the most additional self-explanations were generated, but also by far the most self-explanations with high-knowledge characteristics were present in the LHP group. Learners with average prior knowledge thus seemed to be able to deal with the learning objectives in a way which is detached from the surface features and enables them to elaborate more deeply. Moreover, they even showed a strongly defined high-knowledge characteristic, which indicates that they tended towards high-knowledge self-explanations. However, the self-explanation characteristic under the LP condition shows that they do not express themselves proactively. The self-explanation activity of average-knowledge learners thus needs to be supported in a better way. With regard to that, our results show that exclusive HP are not suitable for fostering high-knowledge self-explanations. Especially against the background that the students had to deal with the highly complex topic of evolution, we assume that the average-knowledge learners are overstrained under the HP condition. This assumption is supported by the structure of human cognition, which implies that the instructional format should fit the difficulty level of the learning matter (Kalyuga, Chandler, and Sweller 2001). The best enhancement to learning evolution seems to be the combination of LP and HP.

From taking a closer look at the self-explanations which are generated by prompts only, it can be concluded that the prompts generally elicited the intended self-explanations and therefore affected the quality of the self-explanations. Although the proportion of the intended low-knowledge self-explanations decreased, the intended high-knowledge self-explanations increased and were higher than under
the HP condition. Bearing in mind that the students learned with two worked example sequences, this effect may be explained by the nature of their training. The first sequence already enabled the students to deal with the profound structure of the worked examples more quickly than the second sequence. This is also reflected in the fact that the students made more inferences when they were asked to paraphrase. Moreover, the LP might have increasingly appeared to be distracting; this may explain the increase of other statements and omitted self-explanations. The increase of high-knowledge self-explanations and the decrease of other statements (concerning the high-knowledge characteristic usually of negative monitoring) compared to the HP condition confirm the assumption that the students with average prior knowledge may feel overstrained under the HP condition and that the quality of the self-explanations is best supported using a combination of LP and HP.

Certainly, the frequency of the self-explanation statements in particular categories is a descriptive indicator of the self-explanation quality. When students generate self-explanations with high-knowledge characteristics only, it does not necessarily mean that they relate to the content in a correct way. The generation of frequently incorrect self-explanations indicates that the quality of the self-explanations is nevertheless low (cf. Mackensen-Friedrichs 2009). For that reason, we also analysed the content-related correctness of the prompted self-explanations as an additional quality feature. Although the differences between correct self-explanations among the three prompting conditions are slightly significant at best, it can be concluded that learners with average prior knowledge were able to generate correct self-explanations even when they were prompted on the high-knowledge level. However, looking at the incorrect self-explanations, the benefits of the LHP condition become apparent again. The combination of low-knowledge and HP significantly reduced the number of incorrect self-explanations and therefore raised the self-explanation quality.

Our study has a few limitations. We did not examine the effectiveness of the different prompting conditions. As our aim was to detect indications of effective and adaptive support measures for learning evolutionary theory, we solely focused on the analysis of the self-explanation characteristic of average-knowledge learners and how it may be influenced by different prompting conditions. Based on the results of previous studies, it can be assumed that differences in the self-explanation quality influence the learning success: Learning with knowledge-adapted self-explanation prompts raises the learning success (Mackensen-Friedrichs 2005; Yeh et al. 2010). The effect of the different prompting conditions on the learning gain will be analysed further and reported in a future article. This future analysis does not only take the learning progress of average-knowledge learners into account, but also compares the three knowledge levels, low, average and high.

A further limitation is the lack of generalisability. The findings can only be applied to learning evolutionary theory. Learning topics with less complexity may be better supported using other prompting conditions under which average-knowledge learners will possibly not feel overstrained by the usage of exclusively HP. However, the adjustment of instructional conditions based on the complexity of the learning matter is recommended (Kalyuga, Chandler, and Sweller 2001). That, of course, does not exclude also reviewing the findings for other topics in further research.

Despite these limitations, the study at hand takes a further step in the development of adaptive self-explanations prompts (cf. Wylie and Chi 2014). The findings suggest the combination of LP and HP to foster learners with average knowledge. Within the highly complex topic of evolutionary theory, it makes sense to start with prompts which structure the learning processes and focus the attention on the relevant aspects of the instructional material. Sophisticated self-explanations concerning deep structures should be promoted at a later time of instruction. In this way, the amount of high-quality self-explanations is considerably strengthened.

**Educational implications**

Up to now, it was hardly possible to transfer results within the field of adaptive self-explanation prompts onto school practice and the development of learning materials. The heterogeneity of individual differences with regard to learners’ prior knowledge in a classroom was not sufficiently studied.
by only focusing on learners with low and high prior knowledge. Our findings provide first indications about how to fill this gap and therefore examine the performance range in classrooms by adaptive self-explanation prompts in greater detail. While learners with low and high prior knowledge should be supported with exclusively low-knowledge and HP, respectively (Mackensen-Friedrichs 2009), a combination which starts with LP and ends with HP would be more useful for learners with average prior knowledge, at least for learning topics which are high in complexity.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG [German Research Foundation]) [grant number GZ MA 3912/3-1].

References

Basel, N., U. Harms, and H. Prechtl. 2013. “Analysis of Students’ Arguments on Evolutionary Theory.” *Journal of Biological Education* 47 (4): 192–199. doi:10.1080/00219266.2013.799078.

Basel, N., U. Harms, H. Prechtl, T. Weiß, and M. Rothgangel. 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.

Best, R., Y. Ozuru, and D. S. McNamara. 2004. “Self-explaining Science Texts: Strategies, Knowledge, and Reading Skills.” In *Proceedings of the 6th International Conference on Learning Sciences*, 89–96. Santa Monica, CA: International Society of the Learning Sciences.

Brumby, M. 1979. “Problems in Learning the Concept of Natural Selection.” *Journal of Biological Education* 13 (2): 119–122.

Catrambone, R., and M. Yuasa. 2006. “Acquisition of Procedures: The Effects of Example Elaborations and Active Learning Exercises.” *Learning and Instruction* 16: 139–153. doi:10.1016/j.learninstruc.2006.02.002.

Chi, M. T. H. 2000. “Self-explaining Expository Texts: The Dual Processes of Generating Inferences and Repairing Mental Models.” In *Advances in Instructional Psychology*, edited by R. Glaser, 161–238. Mahwah, NJ: Lawrence Erlbaum.

Chi, M. T. H., and M. Bassok. 1989. “Learning from Examples via Self-explanations.” In *Knowing, Learning, and Instruction*. Essays in Honor of Robert Glaser, edited by L. B. Resnick, 251–282. Hillsdale, NJ: Erlbaum.

Chi, M. T. H., N. de Leeuw, M. Chiu, and C. LaVancher. 1994. “Eliciting Self-explanations Improves Understanding.” *Cognitive Science* 18 (3): 439–477.

Chi, M. T. H., M. W. Lewis, P. Reimann, and R. Glaser. 1989. “Self-explanations: How Students Study and Use Examples in Learning to Solve Problems.” *Cognitive Science* 13 (2): 145–182.

Dobzhansky, T. 1973. “Nothing in Biology Makes Sense Except in the Light of Evolution.” *The American Biology Teacher* 35 (3): 125–129.

Ericsson, K. A., and H. A. Simon. 1993. *Protocol Analysis: Verbal Reports as Data*. Cambridge: MIT Press.

Freeman, S., S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth. 2014. “Active Learning Increases Student Performance in Science, Engineering, and Mathematics.” *Proceedings of the National Academy of Sciences* 111 (23): 8410–8415.

Futuyma, D. J. 1998. *Evolutionary Biology* (3rd ed). Sunderland, MA: Sinauer Associates.

Kalyuga, S., P. Chandler, and J. Sweller. 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.

Kroß, A., and G. Lind. 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.

Lin, L. L., R. K. Atkinson, W. C. Savenye, and B. C. Nelson. 2014. “Effects of Visual Cues and Self-explanation Prompts: Empirical Evidence in a Multimedia Environment.” *Interactive Learning Environment*. doi:10.1080/10494820.2014.924531.

Lind, G., G. Friege, and A. Sandmann. 2005. “Selbsterklären und Vorwissen.” [Self-explaining and Prior Knowledge.] *Empirische Pädagogik* 19 (1): 1–27.

Lind, G., and A. Sandmann. 2003. “Lernstrategien und Domänenwissen [Learning Strategies and Domain Knowledge].” *Zeitschrift für Psychologie* 211 (4): 171–192.
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 diss., University of Kiel. Accessed September 2010. 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.

Mayr, E. 1982. The Growth of Biological Thought: Diversity, Evolution, and Inheritance. Cambridge: Harvard University Press.

Mayring, P. 2010. Qualitative Inhaltsanalyse. Grundlagen und Techniken (11. Aufl.) [Qualitative Content Analysis. Basics and Techniques]. Weinheim: Beltz.

Nehm, R. H., T. M. Poole, M. E. Lyford, S. G. Hoskins, L. Carrut, B. E. Ewers, and P. J. S. Colberg. 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.

Neubrand, C., C. Borzikowsky, and U. Harms. Forthcoming. “Adaptive Prompts for Learning Evolution with Worked Examples - Highlighting the Students Between the ’novices’ and the ’experts’ in a Classroom.” International Journal of Environmental and Science Education.

Nokes-Malach, T. J., K. VanLehn, D. M. Belenky, M. Lichtenstein, and G. Cox. 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., S. Hübner, S. Dümer, and A. Renkl. 2010. “Expertise Reversal Effects in Writing-to-learn.” Instructional Science 38 (3): 237–258. doi:10.1007/s11251-009-9106-9.

Opfer, J. E., R. H. Nehm, and M. Ha. 2012. “Cognitive Foundations for Science Assessment Design: Knowing What Students Know about Evolution.” Journal of Research in Science Teaching 49 (6): 744–777. doi:10.1002/tea.21028.

Paas, F., and T. van Gog. 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.

Prince, M. 2004. “Does Active Learning Work? A Review of Research.” Journal of Engineering Education 93 (3): 223–231.

Renkl, A. 1997. “Learning from Worked-out Examples: A Study on Individual Differences.” Cognitive Science 21 (1): 1–29.

Renkl, A. 2014. “The Worked Examples Principle in Multimedia Learning.” In The Cambridge Handbook of Multimedia Learning, edited by R. E. Mayer, 391–412, 2nd ed. Cambridge: University Press.

Schilders, M., P. Sloep, E. Peled, and K. Boersma. 2009. “Worldviews and Evolution in the Biology Classroom.” Journal of Biological Education 43 (3): 115–120. doi:10.1080/00219266.2009.9656165.

Sweller, J., J. G. van Merrienboer, and F. Paas. 1998. “Cognitive Architecture and Instructional Design.” Educational Psychology Review 10 (3): 251–296. doi:1040-726X/98/0900-0251S15.00/0.

Wylie, R., and M. T. H. Chi. 2014. “The Self-explanation Principle in Multimedia Learning.” In The Cambridge Handbook of Multimedia Learning, edited by R. E. Mayer, 413–432, 2nd ed. Cambridge: University Press.

Yates, T. B., and E. A. Marek. 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). doi:10.1186/s12052-014-0007-2

Yeh, Y.-E., M.-C. Chen, P.-H. Hung, and G.-J. Hwang. 2010. “Optimal Self-explanation Prompt Design in Dynamic Multi-representational Learning Environments.” Computers & Education 54: 1089–1100. doi:10.1016/j.compedu.2009.10.013.