



Classifying Examples is More Effective for Learning Relational Categories Than Reading or Generating Examples

Tim M. Steininger¹ · Jörg Wittwer¹ · Thamar Voss¹

Received: 4 June 2021 / Accepted: 20 February 2022 / Published online: 24 March 2022
© The Author(s) 2022

Abstract

Successful teaching requires that student teachers acquire a conceptual understanding of teaching practices. A promising way to promote such a conceptual understanding is to provide student teachers with examples. We conducted a 3 (between-subjects factor *example format*: reading, generation, classification) x 4 (within-subjects factor *type of knowledge*: facts, concepts, principles, procedures) experiment with $N=83$ student teachers to examine how different formats of learning with examples influence the acquisition of relational categories in the context of lesson planning. Classifying provided examples was more effective for conceptual learning than reading provided examples or generating new examples. At the same time, reading provided examples or generating new examples made no difference in conceptual learning. However, generating new examples resulted in overly optimistic judgments of conceptual learning whereas reading provided examples or classifying provided examples led to rather accurate judgments of conceptual learning. Regardless of example format, more complex categories were more difficult to learn than less complex categories. The findings indicate that classifying provided examples is an effective form of conceptual learning. Generating examples, however, might be detrimental to learning in early phases of concept acquisition. In addition, learning with examples should be adapted to the complexity of the covered categories.

Keywords Teacher education · Category learning · Relational categories · Example-based learning · Overestimation

✉ Tim M. Steininger
tim.steininger@ezw.uni-freiburg.de

¹ Department of Education, University of Freiburg, Rempartstraße 11,
D-79098 Freiburg im Breisgau, Germany

Classifying Examples is More Effective for Learning Relational Categories Than Reading or Generating Examples

Teachers are a key element in student learning (Hattie, 2012). Accordingly, there is widespread agreement that it is necessary to provide high-quality teacher education (Darling-Hammond, 2006). A promising way to do so is to focus on core practices that teachers must execute to teach effectively (Ball & Forzani, 2009). Examples of such core practices are developing learning objectives in curriculum planning, addressing misbehavior in classroom management, or modeling a solution strategy in explaining (Hogan et al., 2003). To facilitate the acquisition of core practices, it is important that teacher education addresses the knowledge required for executing the to-be-learned core practices (Forzani, 2014). A key aspect of this knowledge refers to the conceptual aspects underlying a core practice. For example, when developing learning objectives for teaching, it is necessary that teachers are aware of the type of knowledge that needs to be acquired to reach an instructional objective. Only then can teachers select instructional methods that specifically support the acquisition of the target knowledge (e.g., Koedinger et al., 2012).

Research has shown that, compared with novice teachers, expert teachers usually possess more conceptual knowledge that helps them to flexibly adapt their teaching to a classroom situation (Feldon, 2007). In addition, there is empirical evidence suggesting that without explicit instruction on conceptual knowledge teachers might rely their teaching upon naïve conceptions of core practices (Calderhead & Robson, 1991). In this article, we present an experiment in which we instructionally supported the learning of relational categories necessary for successfully executing core practices. More specifically, we were interested in how different ways of learning through examples would influence student teachers' acquisition of relational categories.

Relational Categories in Teacher Education

Concepts are mental representations of categories that consist of entities being the same for some respect (e.g., Medin & Rips 2005). They support not only cognitive activities such as categorization, perception, or memory (Barsalou et al., 2003) but also goal-directed action (Barsalou et al., 2018). Research has catalogued different kinds of categories (Medin et al., 2000; Murphy, 2004). The majority of studies has focused on feature-based categories whose members share properties that are intrinsic to the entities of a specific category (e.g., barking is a property of dogs; Goldwater et al., 2018). In educational settings such as teacher education, however, many categories are not feature-based but relational (Goldwater & Schalk, 2016). Relational categories consist of entities whose membership is determined by a relational structure (Gentner & Kurtz, 2005). Therefore, the role that an entity plays in a relational system is critical for being a member of a relational category.

In the experiment reported in this article, we examined relational categories in the context of lesson planning. More concretely, we focused on different types of knowledge that need to be acquired to achieve a learning objective. Merrill and Twitchell (1994) proposed the following types of knowledge: facts, concepts, principles, and procedures. For example, to understand which type of knowledge students must acquire to reach the learning objective *Can name the rectangle area formula*, a teacher needs to know that the learning objective

addresses knowledge about facts. This type of knowledge is a relational category because it specifies a relation between the condition under which the knowledge is applied and the type of response required when applying the knowledge (Koedinger et al., 2012).

Learning Relational Categories Through Examples

The literature on concept learning has primarily concentrated on feature-based categories (Goldwater et al., 2018). Learning relational categories, however, can be different from learning feature-based categories (Goldwater & Schalk, 2016; see, however, Little & McDaniel 2015). First, not intrinsic properties of entities but relational structures need to be identified. Second, the role that each entity plays in a relational structure together with the rule specifying the relationship between the different roles of the entities must be discovered. Third, surface features of entities are usually not relevant for understanding the relations among the entities. Therefore, it is important to abstract from these features and recognize the underlying regularities.

In a series of experiments, Rawson et al., (2015) examined the conditions under which learning of relational categories is most effective. More concretely, they contrasted (re-) reading the definition (i.e., the rule) of relational categories with reading the definition together with provided examples of relational categories. The results showed that reading the definition together with provided examples more effectively supported the acquisition of relational categories than (re-)reading the definition alone. In another series of experiments, Rawson and Dunlosky (2016) showed that not only reading provided examples but also generating examples was more beneficial for learning relational categories than (re-) reading the definition alone. In addition, Zamary and Rawson (2018) investigated whether generating examples would be more effective for learning than reading provided examples. Contrary to expectation, the reading of provided examples was not less but more effective than the generation of examples.

Another line of literature focuses on inductive forms of learning (relational) categories (Brunmair & Richter, 2019). In this research, learners typically observe sequentially presented examples and classify them into categories. After each classification, learners receive feedback and, if their classification was wrong, are presented with the correct category. In this way, learners acquire knowledge about categories by trial and error (Markman & Ross, 2003). In some studies (e.g., Goldwater et al., 2018), learners also receive explicit instruction about the rules determining category membership. Usually, this instruction improves classification and, thus, supports learning (e.g., Jung & Hummel 2015).

Overall, the reported studies (e.g., Goldwater et al., 2018; Jung & Hummel, 2015; Zamary & Rawson, 2018) demonstrate that examples are important in supporting the learning of relational categories. Their benefits might be attributed to the fact that examples can provide real-world contexts that illustrate how abstract relationships are to be applied (Rawson et al., 2015). In addition, studying examples of a relational category might evoke structural alignment processes by which learners develop a deep understanding of the relational structure instantiated by the examples (Rittle-Johnson & Star, 2011).

Cognitive Engagement in Learning Relational Categories Through Examples

To theoretically explain how different example formats such as reading, generating, or classifying examples affect learning of relational categories, the ICAP framework (Chi & Wylie, 2014) is informative. According to this framework, learners can exhibit four different modes of engagement in learning: passive, active, constructive, interactive. A passive mode of learning means that learners receive information without doing anything else. In active learning, learners manipulate information, for example, by restudying the learning materials. In a constructive mode, new information is created that goes beyond what was presented in the learning materials. Interactive learning occurs when learners cooperate and co-construct new knowledge. Usually, the more engaged learners are, the more they learn. Drawing on the ICAP framework (Chi & Wylie, 2014), it can be assumed that for learning relational categories a constructive mode of engagement such as generating new examples is more effective than a passive mode of engagement such as reading provided examples. However, Rawson and Dunlosky (2016) as well as Zmary and Rawson (2018) found that the generation of examples was less beneficial for learning than the reading of provided examples. This seemingly counterintuitive result can be explained by the fact that learners generated only partially correct examples, which might have undermined the potential of example generation for learning relational categories. Interestingly, Rawson and Dunlosky (2016) additionally found that interventions to improve the correctness of generated examples such as prompting learners to recall the definition of a relational category or providing learners with the definition of a relational category resulted in more but still limited success. From these results one can conclude that generating examples is cognitively too overwhelming for learners in early phases of concept acquisition. Therefore, it seems to be more effective for learning when reading high-quality examples instead.

In line with the ICAP framework (Chi & Wiley, 2014), classification with initial instruction as examined in research on inductive learning of categories (e.g., Jung & Hummel 2015) can also be conceptualized as a constructive mode of engagement. This is because when learners classify examples into categories, they can do so by comparing their already existing knowledge about the rules determining category membership with the example being classified. As a result, they create new information by assigning a category to the example (see also Zepeda & Nokes-Malach, 2021). It can be assumed that, even though example generation and classification are both modes of constructive engagement, classifying is cognitively less burdensome than generating examples. If this is true, it can be expected that during learning examples are more often classified correctly than examples are generated correctly. Accordingly, example classification might more effectively lead to the acquisition of relational categories than example generation. At the same time, due to its constructive nature, example classification can be assumed to be more helpful in learning relational categories than only reading provided examples.

Monitoring the Learning of Relational Categories Through Examples

For learning to be effective, it is important that learners accurately monitor their learning activities (Prinz et al., 2019). Thus, when generating examples to learn relational categories,

it is necessary that learners judge the correctness of their generated examples. Otherwise, misconceptions are likely to occur. Zmary et al., (2016) found that learners overestimated the correctness of their generated examples. Even when learners received information that could be used to judge the quality of example generation, they remained overconfident. Obviously, example generation results in an illusion of understanding. This illusion probably occurs because learners primarily pay attention to the ease with which they can retrieve information about a generated example without acknowledging whether the information is correct (Prinz et al., 2018; Zmary et al., 2016).

In contrast, when learning relational categories by classifying or reading examples, the amount of actively generated information is rather low. Therefore, learners who monitor their learning under these circumstances cannot intensely retrieve incorrect information. As a result, monitoring activities in these cases might be less error-prone than monitoring the correctness of generated examples. Indirect evidence for this assumption comes from studies (e.g., Miesner & Maki 2007) showing that monitoring accuracy is higher for multiple-choice questions (i.e., questions with provided answer options) than for short-answer questions (i.e., questions without provided answer options).

Present Study

We conducted an experiment to examine how student teachers acquire relational categories in the context of lesson planning. More concretely, student teachers learned through examples which type of knowledge needs to be applied to achieve a learning objective. Following the taxonomy proposed by Merrill and Twitchell (1994), four different types of knowledge, namely, facts, concepts, principles, and procedures, were to be learned. According to Koedinger et al., (2012), the four types of knowledge differ in the complexity of their condition-response relationship. Facts have a constant condition (e.g., *name the rectangle area formula*) and constant response (e.g., $area = length \times width$) with a one-to-one mapping. Concepts, in contrast, have a variable condition (e.g., *What is a sheepdog? What is a poodle?*) but a constant response (e.g., *dog*) with a many-to-one mapping. Principles and procedures both have a variable condition (e.g., *Find the area of rectangle A. Find the area of rectangle B.*) and a variable response (e.g., $A: 5 = 5 \times 3, B: 8 = 2 \times 4$) with a many-to-many mapping. Procedures can be even more complex than principles when they require the understanding of a principle underlying the procedure (Wittwer & Renkl, 2010).

We contrasted three groups of learning through examples: The first group read provided examples, the second group generated examples, and the third group classified provided examples. In line with Zmary & Rawson (2018), we predicted that reading provided examples would be more effective for learning than generating examples (*learning-by-reading-examples hypothesis*). In addition, we assumed that classifying would increase the effectiveness of studying provided examples for learning. Therefore, we expected that classifying provided examples would be more beneficial to learning than reading provided examples or generating examples (*learning-by-classifying-examples hypothesis*). In particular, we assumed classifying examples to be more effective for learning than generating examples because during learning the number of correctly classified examples should be higher than the number of correctly generated examples. Accordingly, more correct examples should result in more learning (*correctness-of-examples hypothesis*). When clas-

sifying examples, feedback was provided. Therefore, we also examined whether feedback improved the number of correctly classified examples during learning (*feedback research question*). Concerning the type of knowledge to be learned (e.g., Koedinger et al., 2012), we assumed that learning would be more difficult with increasing complexity of the type of knowledge (*type-of-knowledge hypothesis*). Therefore, we expected more knowledge about facts to be acquired than knowledge about concepts, more knowledge about concepts to be acquired than knowledge about principles, and more knowledge about principles to be acquired than knowledge about procedures. Zamary et al. (2016) showed that learners overestimated the correctness of their generated examples when learning relational categories. It can be assumed that this was because learners mainly used the new but partly incorrect information that they produced through example generation as a basis for judging the quality of their learning. When reading or classifying provided examples, there is limited opportunity to generate new information that could bias judgments of learning. Accordingly, we predicted that reading or classifying provided examples would result in more accurate judgments of learning than generating examples (*judgment-bias hypothesis*).

Method

Sample and Design

A total of $N=83$ student teachers participated in the experiment. All student teachers had a bachelor's degree with a teacher training component and were currently enrolled in a master's degree teacher education program. Student teachers studied natural sciences ($n=29$; i.e., biology, chemistry, mathematics, physics) or social sciences and humanities ($n=54$; i.e., history, languages, politics). We used an experimental design with two independent variables, namely *example format* and *knowledge type*. Example format was a between-subject factor. The student teachers were randomly assigned to one of three groups of example format: *reading* provided examples ($n=25$), *generating* examples ($n=27$), or *classifying* provided examples ($n=31$). Knowledge type was a within-subject factor. The student teachers learned about four different types of knowledge, namely *facts*, *concepts*, *principles*, and *procedures*.

A sensitivity analysis was conducted using G*Power (Faul et al., 2009). For the data analysis using MANOVA including repeated and between-subject measures, alpha level of 0.05, power of 0.80, and correlation among repeated measures of 0.16, the sample provided sufficient sensitivity to detect medium to small effects concerning the between-subject factor, $\eta^2=0.04$, as well as the within-subject factor, $\eta^2=0.03$.

Procedure

The experiment was embedded in an online lesson on classroom teaching in a teacher education program at a university in Germany. In this lesson, student teachers learned how to write and analyze learning objectives. The experiment consisted of a learning phase and a testing phase. Completion of all tasks was self-paced to allow student teachers to work on the tasks with sufficient time. The experiment took 30 to 45 min. All participants were provided information about the study and volunteered to participate.

In the learning phase, information about learning objectives and types of knowledge was provided. Each of the four types of knowledge according to the taxonomy proposed by Merrill & Twitchell (1994), that is, facts, concepts, principles, and procedures, was defined. In addition, an example of a learning objective together with each type of knowledge required to reach the learning objective was presented. After that, student teachers received a list of ten school subjects (e.g., mathematics, biology, history) and were asked to select a school subject they were studying. The examples that the student teachers received or generated for learning were taken from the selected school subject. This was done to make learning for every student teacher as relevant as possible. To further engage student teachers in learning, they were asked to have in mind possible commonalities and differences between examples.

Depending on example format, the student teachers read, generated, or classified subsequently two examples of a learning objective for each of the four types of knowledge, resulting in eight examples in total. According to the type of knowledge, we ordered the examples to be read, generated, or classified as follows: (1) concepts, (2) procedures, (3) principles, and (4) facts. In the group *reading*, the student teachers read provided examples of learning objectives together with the type of knowledge associated with it (e.g., *The following learning objective addresses knowledge about concepts: Students can identify action potentials in plots*). In the group *generating*, the student teachers generated examples of learning objectives on their own (e.g., *Please provide an example of a learning objective that addresses knowledge about facts*). In the group *classifying*, the student teachers classified the provided examples of learning objectives into the type of knowledge presumably addressed by each of the learning objectives. To do so, they selected one of the four types of knowledge presented (e.g., knowledge about facts). Feedback was given to indicate the correct type of knowledge for the example and whether the selected answer was correct.

In the testing phase, student teachers completed a test measuring their knowledge about the four types of knowledge. The test consisted of four different types of tasks. Then, they indicated the certainty with which they would identify the type of knowledge underlying a learning objective. Finally, they were debriefed.

Measures

Knowledge Test

The knowledge about the four types of knowledge was assessed by four different types of tasks (Markman & Ross, 2003). The first type was a classification task and required student teachers to classify eight sequentially and randomly presented learning objectives. The second type was a generation task. For each of the four types of knowledge, student teachers wrote two examples of a learning objective. They were told to generate the first example for the school subject they had selected and the second example for a second school subject they were studying. The third type was an inference task. Student teachers were provided with examples of learning objectives and had to write down missing features of the examples other than the category label (e.g., *A learning objective in the school subject biology refers to osmosis. The learning objective addresses knowledge about principles. Which tasks can students execute with the knowledge obtained through this learning objective?*). The fourth type was a definition recall task embedded in a real-world context. The student teachers were told to envision talking to a teacher colleague about learning objectives. They

were asked to write down for each of the four types of knowledge the commonalities that all learning objectives of a type of knowledge would possess, which corresponds to the definition of this type of knowledge.

The written answers to the questions in the generation, inference, and definition recall task were scored for correctness by two raters. Interrater reliability was good, all $\kappa > 0.81$. Every correct answer in the tasks was assigned 1 point. Every incorrect answer was assigned 0 points. The task score for each of the four types of tasks of the knowledge test was calculated as ratio of all correctly answered questions (0=no correct answers; 1=every answer correct).

Certainty and Judgment Bias

After completing the knowledge test, student teachers were asked to judge the certainty with which they would identify a type of knowledge underlying a learning objective on a 7-point rating scale for each of the four types of knowledge (1=very low; 7=very high). To examine the accuracy with which student teachers judged the certainty of identifying a type of knowledge underlying a learning objective, we computed a bias measure (see Schraw, 2009). To do so, we z -transformed the scores obtained for the certainty for each of the four types of knowledge and the scores obtained for performance in the knowledge test for each of the four types of knowledge. Then, we computed the signed difference between certainty and performance in the knowledge test for each of the four types of knowledge. A positive value indicated overconfidence whereas a negative value indicated underconfidence. A score of zero indicated a completely accurate judgment.

Learning Phase: Correctness of Example Generation and Classification

In the learning phase, student teachers in the group *generating* produced examples of learning objectives for each of the four types of knowledge on their own. The generated examples were scored for correctness by two raters (0=incorrect example; 1=correct example). Interrater reliability was good, $\kappa=0.87$. In addition, student teachers in the group *classifying*

Table 1 Means (and Standard Deviations) for Correctness of Examples in the Learning Phase and Knowledge Acquisition Measures

Variables	Example format			
	Overall	Read	Generate	Classify
Learning phase ^a	0.64 (0.20)	-	0.59 (0.24)	0.69 (0.16)
Correctness of first examples	0.68 (0.23)	-	0.59 (0.23)	0.76 (0.20)
Correctness of second examples	0.61 (0.30)	-	0.58 (0.32)	0.63 (0.28)
Classification task	0.73 (0.17)	0.77 (0.17)	0.72 (0.19)	0.72 (0.15)
Generation task	0.69 (0.23)	0.65 (0.22)	0.57 (0.24)	0.83 (0.15)
Generation for covered subject	0.70 (0.29)	0.69 (0.26)	0.56 (0.32)	0.84 (0.21)
Generation for second subject	0.67 (0.25)	0.60 (0.27)	0.57 (0.24)	0.81 (0.17)
Inference task	0.67 (0.25)	0.67 (0.26)	0.58 (0.24)	0.76 (0.24)
Definition recall task	0.83 (0.26)	0.79 (0.30)	0.77 (0.27)	0.93 (0.18)

Note. Numbers represent ratio of correct answers. Standard deviations are presented in parentheses

^a Subordinate rows represent mean correctness over all first/second examples for each knowledge type

assigned the provided examples to one of the four types of knowledge presented. Correctness of classification was scored (0=incorrect classification; 1=correct classification).

Results

Table 1 shows descriptive statistics, Table 2 displays bivariate correlations between variables. To account for testing multiple comparisons, we used a Bonferroni-adjusted alpha level of 0.0125.

Effects of Example Format and Type of Knowledge on Learning

We conducted a MANOVA with the between-subject factor *example format*, the within-subject factor *type of knowledge*, and the scores obtained in the four types of tasks of the knowledge test (i.e., *classification, generation, inference, definition recall*) as dependent variables. In a first step, we tested the *learning-by-reading hypothesis* that reading examples

Table 2 Intercorrelations for Correctness of Examples in the Learning Phase and Knowledge Acquisition Measures

	Variables	2	3	4	5
1	Correctness of examples in learning phase	0.37*	0.41*	0.37*	0.18
	Generated examples	0.46*	0.45*	0.25	0.29
	Classified examples	0.26	0.08	0.40*	-0.25
2	Classification task		0.27*	0.22*	0.15
3	Generation task			0.33*	0.29*
4	Inference task				0.42*
5	Definition recall task				

* $p < .05$

Table 3 Univariate Follow-Up Contrast Analyses Concerning Example Format and Type of Knowledge

Hypothesis and univariate measures	$F(1, 80)$	p	η_p^2
Learning-by-reading hypothesis			
Classification task	1.23	0.136	0.02
Generation task	1.73	0.096	0.02
Inference task	1.63	0.103	0.02
Definition recall task	0.09	0.380	0.00
Learning-by-classifying hypothesis			
Classification task	0.45	0.251	0.01
Generation task	21.81	<0.001	0.21
Inference task	5.61	0.010	0.07
Definition recall task	6.68	0.006	0.08
Type-of-knowledge hypothesis			
Classification task	2.99	0.044	0.04
Generation task	10.62	0.001	0.11
Inference task	41.33	<0.001	0.34
Definition recall task	6.86	0.006	0.08

Note. All p values are one-tailed

would be more beneficial for learning than generating examples. To do so, we computed a planned contrast with the following weights: reading: +1, generating: -1, classifying: 0. The multivariate analysis was not significant, $V=0.038$ (Pillai's trace), $F(4, 77)=0.76$, $p=.277$, $\eta_p^2=0.04$. Follow-up univariate analyses were also not significant (see Table 3). Hence, reading examples was not more effective than generating examples.

In a second step, we examined the *learning-by-classifying hypothesis* that classifying examples would support learning more effectively than reading or generating examples. We computed a planned contrast with the following weights: reading: -1, generating: -1, classifying: +2. The multivariate analysis was significant, $V=0.290$ (Pillai's trace), $F(4, 77)=7.86$, $p<.001$, $\eta_p^2=0.29$. Follow-up univariate analyses showed that the planned contrast was significant for all types of the tasks of the knowledge test with the exception of the classification task (see Table 3). Accordingly, classifying examples was more beneficial for learning than reading or generating examples.

In a third step, we tested the *correctness-of-examples hypothesis* to examine the correctness with which student teachers generated or classified examples in the learning phase. Student teachers generated 59% of all examples correctly ($SD=23.72$). In contrast, student teachers classified 69% of all examples correctly ($SD=15.76$). This difference was significant, $F(1, 56)=4.08$, $p=.048$, $\eta^2=0.07$. Student teachers who classified provided examples received feedback whereas student teachers who generated examples did not. To address the *feedback research question*, we analyzed whether feedback supported student teachers in improving their classification of the provided examples of the same type of knowledge from the first trial to the second trial in the learning phase. To this end, we compared the number of correctly classified examples between the first trial and the second trial across the four types of knowledge. We conducted a repeated-measures analysis with *example format* (i.e., generating and classifying) as between-subject factor, *trial* as within-subject factor and *correctness of generated or classified examples* as dependent variable. There was no significant effect of trial, $F(1, 56)=2.46$, $p=.12$, $\eta_p^2=0.04$, and no significant interaction effect between trial and example format, $F(1, 56)=1.84$, $p=.18$, $\eta_p^2=0.03$. Thus, feedback did not improve the number of correctly classified examples in the learning phase (see also Table 1).

To investigate whether the correctness of generated and classified examples influenced learning, we computed correlations with performance in each of the four types of tasks of the knowledge test separately for both experimental conditions. As displayed in Table 2, correctly generating examples was significantly correlated with performance in the classification task and in the generation task but not in the inference task and in the definition recall task. Conversely, correctly classifying examples significantly correlated with performance only in the inference task.

In a fourth step, we tested the *type-of-knowledge hypothesis* that learning about the different types of knowledge would be more challenging with increasing complexity of the type of knowledge. For the planned contrast, we used the following weights to test a linear decrease in learning with increasing complexity of type of knowledge: facts: +3, concepts: +1, principles: -1, procedures: -3. In line with this hypothesis, the contrast was significant, $V=0.392$ (Pillai's trace), $F(4, 77)=12.44$, $p<.001$, $\eta_p^2=0.39$. Follow-up univariate analyses showed that the planned contrast was significant for all four types of the tasks of the knowledge test (see Table 3). Hence, learning outcomes were lower with increasing complexity of the type of knowledge. Descriptively, however, the learning outcome for knowledge about procedures was rather high in the classification task and in the generation task (see Table 1).

Effects of Example Format on Judgment Bias

We conducted a MANOVA with the between-subject factor *example format* and *judgment bias* for the four types of knowledge as dependent variables. According to the *judgment-bias hypothesis*, we assumed that overconfidence in identifying the type of knowledge underlying a learning objective would be higher when generating examples than when reading or classifying examples. To test this hypothesis, we computed a planned contrast with the following weights: reading: -1 , generating: $+2$, classifying: -1 . The multivariate analysis was significant, $V=0.174$ (Pillai's trace), $F(4, 77)=4.05$, $p=.003$, $\eta^2=0.17$. Follow-up univariate analyses showed that the planned contrast was significant for principles and procedures as types of knowledge (see Table 4). Thus, in line with the hypothesis, generating examples resulted in more overconfidence than reading or classifying provided examples particularly with regard to the certainty in identifying that a learning objective addresses knowledge about principles or procedures.

Discussion

In this study, we examined the influence of different example formats on learning relational categories. First, we found that reading examples was not significantly more beneficial for learning than generating examples. This result stands in contrast to the experiments conducted by Zamary and Rawson (2018) who found that reading examples supported learning more effectively than generating examples. The contradictory results might be explained by

Table 4 Univariate Follow-Up Contrast Analyses Concerning Judgment Bias

Example Format	Judgment Bias		Performance		Certainty	
	z	M (SD)	z	M (SD)	z	M (SD)
Facts: $F(1, 80)=1.75$, $p=.085$, $\eta^2=0.02$						
Reading	-0.06	0.86 (0.15)	0.08	6.00 (0.87)	0.02	
Generation	0.34	0.77 (0.17)	-0.52	5.78 (1.45)	-0.18	
Classification	-0.25	0.90 (0.11)	0.39	6.13 (0.96)	0.14	
Concepts: $F(1, 80)=0.03$, $p=.432$, $\eta^2=0.00$						
Reading	-0.33	0.75 (0.24)	0.06	4.28 (1.14)	-0.27	
Generation	0.28	0.66 (0.32)	-0.27	4.59 (1.31)	0.00	
Classification	0.02	0.78 (0.21)	0.19	4.84 (1.04)	0.21	
Principles: $F(1, 80)=7.13$, $p=.005$, $\eta^2=0.08$						
Reading	-0.12	0.59 (0.25)	-0.13	3.84 (1.49)	-0.24	
Generation	0.70	0.53 (0.21)	-0.38	4.59 (1.37)	0.32	
Classification	-0.51	0.71 (0.17)	0.44	4.06 (1.15)	-0.08	
Procedures: $F(1, 80)=11.74$, $p<.001$, $\eta^2=0.13$						
Reading	0.06	0.69 (0.27)	-0.23	4.96 (1.46)	-0.18	
Generation	0.59	0.69 (0.25)	-0.23	5.70 (1.49)	0.36	
Classification	-0.56	0.83 (0.18)	0.39	4.97 (1.20)	-0.17	

Note. All p values are one-tailed. Numbers for performance represent the mean ratio of correct answers over all tasks. Numbers for certainty represent the mean of the 7-point rating scale. Judgment bias was computed according to Schraw (2009) using z -transformed values to account for the different scales

differences in the learning materials. Zamary and Rawson (2018) used relational categories from social psychology (e.g., hindsight bias) whereas the relational categories examined in our study came from educational psychology. It can be assumed that the relational categories that we used in our study were somewhat easier to learn than the relational categories investigated by Zamary and Rawson (2018). As a result, example generation was rather successful in our study. This assumption is corroborated by the finding that 59% of the generated examples in our study were correct. In contrast, the correctness of the generated examples in the experiments conducted by Zamary and Rawson (2018) was clearly lower (Experiment 1: 22%; Experiment 2: 19%). Obviously, when the complexity of the learning material is high and, thus, impoverishes the quality of generated examples, reading examples is more advantageous than generating examples. However, the superiority of reading examples over generating examples seems to disappear when the complexity of the learning material is reduced.

Second, our study revealed that classifying examples was more effective for learning than reading or generating examples. More concretely, student teachers who classified examples in the learning phase generated examples, inferred missing information in examples, and communicated about knowledge types more effectively than student teachers who read or generated examples in the learning phase. In line with the ICAP framework (Chi & Wylie, 2014), the advantage of example classification can be attributed to the constructive nature of this form of learning. Student teachers who classified examples might have been more actively engaged in learning than student teachers who just read examples. At the same time, student teachers who classified examples received instructional guidance because they were provided with the labels of the relational categories (i.e., types of knowledge). This might have made example classification cognitively less demanding than example generation where student teachers were completely left to their own devices. The significant difference in the correctness between example classification and example generation in the learning phase adds to this picture: Whereas 69% of all examples were classified correctly, only 59% of all examples were generated correctly. It might be argued that example classification supported learning most effectively mainly because student teachers who classified examples received feedback in the learning phase whereas student teachers who generated examples did not. Our analysis of the correctly classified examples in the learning phase, however, revealed that student teachers did not improve the rate of correctly classified examples from the first to the second trial (see Table 1). Thus, it was not primarily the feedback but the classifying as such that resulted in higher learning outcomes. In other words, the mechanisms underlying improved learning as a result of classifying examples might refer to the processes that are triggered by classification (e.g., Goldwater & Schalk 2016). More concretely, classifying examples usually engages learners in abstracting from the concrete examples within a category to find commonalities. At the same time, it encourages the discrimination between examples of different categories. Reading or generating examples might induce similar processes but does so in much more unpredictable ways. Therefore, when the goal is to acquire knowledge about categories, processes that immediately result in the construction of mental structures in the form of such categories seem to be the most effective.

Third, our study showed that example generation made student teachers too optimistic in identifying the type of knowledge underlying a learning objective. Overconfidence primarily occurred when student teachers were asked to indicate their certainty in identifying that a learning objective addresses knowledge about principles or procedures. In contrast, when

reading or classifying examples, student teachers were more accurate or even tended to underestimate their performance. The observed overestimation that occurred as a result of example generation is in line with the findings obtained by Zamyar et al., (2016). Obviously, producing new information when generating examples makes learners prone to erroneously believe in the effectiveness of this form of learning. Thus, our results indicate a double curse of example generation in early phases of concept acquisition: Not only does example generation result in limited learning but it also produces an illusion of understanding.

Fourth, we found that the success with which student teachers learned depended on the complexity of the types of knowledge. Thus, it was easier to acquire knowledge about facts than, for example, about principles. However, student teachers were rather successful in applying their knowledge about procedures when asking to classify or generate examples in the posttest. This result can be explained by the fact that identifying a procedure in a learning objective or generating an example of a learning objective covering knowledge about a procedure primarily required student teachers to be aware of the steps necessary to execute a procedure. This might have been easier than identifying or describing a relationship between concepts as required for learning objectives covering knowledge about principles.

Limitations and Future Research

First, we examined relational categories taken from authentic learning materials relevant for students in their teacher education program. Thus, in contrast to the classical classification paradigm that often uses artificial categories, we focused on real-world categories (see also Goldwater & Schalk 2016). Even so, it is an open question whether the results obtained in our study are generalizable to other relational categories. Although there is an increasing interest in how to learn relational categories, research has mainly focused on classification as the predominant form of learning relational categories (e.g., Corral et al., 2018). Obviously, the implicit assumption in this research is that classification is the best way to learn relational categories. This assumption is in line with the results obtained in our study. Nevertheless, whether example classification is always more effective than example generation regardless of the relational categories to be learned is an interesting question for future research. Similarly, studies might examine how different example formats can be combined to exploit the potential of each example format for conceptual learning. Consistent with our results, example classification seems to be the most effective way to acquire concepts in early phases of concept acquisition. However, after learners have developed some initial understanding, example generation, as opposed to example classification, might be more helpful to actively apply the newly learned concepts to real-world contexts. Although we did not investigate how a combination of classifying and generating examples would affect learning, our findings show that example classification in the learning phase resulted in better example generation in the test phase. Also, research might investigate other forms of learning that are similar to example classification. For example, Goldwater et al., (2018) showed that inference learning, which asks learners to infer missing information in examples, as we studied in our inference task, might also serve as an effective way to learn relational categories.

Second, it is an open question whether the findings obtained in our study are confined to learning relational categories. Usually, it is assumed that it is important to construct a

mental representation of a relational structure to learn relational categories (e.g., Goldwater & Schalk 2016). However, the results of the experiments conducted by Corral et al., (2018) suggest that it can be sufficient to form a feature-based mental representation for learning relational categories. In this case, the relational structure making up a relational category is represented as a feature in the list of elements that form a relational category. For example, instead of constructing a mental representation of the relationship between a constant condition (e.g., *name the rectangle area formula*) and a constant response (e.g., *area = length x width*) to identify the learning objective *Can name the rectangle area formula* as knowledge about facts it might also be possible to directly associate with the rectangle area a fixed formula and, thus, a constant answer. Accordingly, a list making up this relational category could contain the elements *constant, rectangle area, formula*. Future research is encouraged to examine in more detail which type of mental representation learners form when learning relational categories and the circumstances under which they might select among different types of mental representations. Similarly, studies might seek to extend the results obtained in our study by directly comparing feature-based category learning with relational category learning.

Third, student teachers in our study learned the different types of knowledge in a pre-defined order (i.e., concepts, procedures, principles, facts). This order might have influenced the way student teachers learned. Therefore, the impact of different orders on learning relational categories should be examined in future research. Moreover, in our study, two examples of each of the four types of knowledge were to be read, generated, or classified subsequently. Thus, the examples were presented in a blocked format. Hence, an important next step would be to investigate how studying examples in an interleaved presentation influences learning. For example, the meta-analysis conducted by Brunmair and Richter (2019) suggests that in inductive forms of concept learning interleaving is superior to blocking when learning materials are complex. Similarly, Rawson et al. (2015) found that interleaving is particularly effective for learning by reading examples when definitions are not present during example study.

Fourth, we argued that example generation might be cognitively too overwhelming in early phases of concept acquisition. In contrast, example classification seems to be more beneficial to learning because it is cognitively less burdensome. To provide direct evidence for this assumption, future studies could use cognitive load measures to reveal differences in the amount of cognitive processing associated with an example format (e.g., Sweller, 2020). For example, it is plausible to conjecture that the intrinsic load, which results from the complexity of a learning task, is rather high when learners need to generate examples. Classifying examples, in contrast, might free up cognitive resources that can be used to devote attention to understanding what makes up a category. Therefore, when classifying examples, the intrinsic load might not only be lower, but also the germane load, which is associated with connecting new information with prior knowledge, might be higher than when generating examples.

Fifth, although we found differences in learning outcomes as a function of example format, we did not examine long-term effects. Therefore, future studies might use delayed knowledge tests to investigate whether the effectiveness of the different example formats is sustainable. Classifying and generating examples are active forms of cognitive engagement. Therefore, they might produce more stable learning than reading examples which is a fairly passive form of cognitive engagement. Nonetheless, Zarny and Rawson (2016) who used

tests with a 2-day delay in their study, found greater learning for learners who read examples than for learners who generated examples.

Sixth, student teachers in our study learned the relational categories in a self-paced manner. Therefore, learning time might have varied depending on the example format. However, we did not collect data on learning time. Therefore, future studies might investigate whether possible differences in learning time might contribute to explaining the observed differences in learning between the example formats.

Seventh, we showed that classifying examples together with the provision of feedback was more beneficial for learning than reading or generating examples. However, more research is needed to examine the role of feedback for learning by classifying examples. In our study, feedback did not immediately improve classification in the learning phase. Nevertheless, given the general relevance of feedback for learning (e.g., Wisniewski et al., 2020), research is encouraged to systematically examine the influence of feedback on conceptual learning through different example formats. Therefore, studies might investigate the effect of feedback on learning by example classification in direct comparison to learning by example generation. Similarly, it would be important to study how to design feedback to optimally support learning from example classification and example generation. For example, Corral et al. (2021, Experiment 3) found that example classification with delayed feedback resulted in more learning than example classification with immediate feedback.

Eighth, how feedback exactly comes into play when learners self-assess their learning of relational categories is also an issue for future research. In our study, we found that example classification (with feedback) resulted in more accurate self-assessments of learning than example generation (without feedback). Accordingly, it might be that feedback helped learners who classified examples to form rather realistic self-assessments. However, Zamary et al. (2016) found that feedback did not prevent learners from being too optimistic in their conceptual learning. Also, our study revealed that learners who read examples (without feedback) were most often even more accurate in their self-assessments than learners who classified examples. Hence, feedback might not have played a major role for forming accurate self-assessments in learning by classifying examples. Nonetheless, future research is encouraged to use, for example, the think-aloud method to unveil how feedback influences self-assessments when learning relational categories.

Ninth and lastly, concerning the number of examples, we found that two examples for each of the four knowledge types did not result in perfect classification. Hence, it would be interesting to study which number of examples are necessary to optimize learning. In addition, it is worthwhile to examine how to adapt the number of examples specifically to the complexity of the relational category to be learned.

Practical Implications

The results of our study show that classifying provided examples is an effective form of learning relational categories. More common forms of learning relational categories such as reading provided examples or generating new examples might be less beneficial than intuitively assumed. This seems to be especially true for generating new examples because this form of learning obviously makes learners particularly prone to an illusion of understanding. Therefore, when teaching concepts, it is recommended to let learners generate their own

examples not in early phases of concept acquisition. Instead, it seems to be more beneficial to make heavier use of classification tasks in which learners can thoroughly reflect upon which example belongs to which relational category. Similarly, when learners engage in self-regulated learning of relational categories, they should refrain from generating examples too early.

The finding that the effectiveness of learning different types of knowledge depended on their complexity suggests that there is no *one-size-fits-all* approach in learning relational categories. Instead, it seems to be important to adapt instruction to the complexity of relational categories. One way to do so is to increase the number of examples that illustrate complex relational categories. For instance, while it might be sufficient to present two examples of knowledge about facts, more than two examples of knowledge about principles might be necessary to guarantee understanding of this type of knowledge.

Furthermore, our study revealed that student teachers who generated examples in the learning phase were too confident in identifying a learning objective addressing principles or procedures. From this result one can conclude that student teachers might have difficulties in writing learning objectives that cover these types of knowledge. Therefore, it is important that student teachers are aware that they might overestimate their understanding of relational categories when learning through example generation.

Authors' Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Tim M. Steininger and Jörg Wittwer. The first draft of the manuscript was written by Tim M. Steininger and all authors commented on previous versions of the manuscript and revised it critically for important intellectual content. All authors read and approved the final manuscript.

Funding This study is part of the *Qualitaetsoffensive Lehrerbildung*, a joint initiative of the German federal government and the Länder which aims to improve the quality of teacher training. The program is funded by the German Federal Ministry of Education and Research (BMBF). The authors are responsible for the content of this publication.

Open Access funding enabled and organized by Projekt DEAL.

Availability of Data The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Compliance with Ethical Standards The procedure was approved by the ethics committee of the University of Freiburg (ID: 21-1090). Informed consent for participation and publication was obtained from each participant.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ball, D., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education*, 60(5), 497–511. <https://doi.org/10.1177/0022487109348479>
- Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, 7(2), 84–91. [https://doi.org/10.1016/S1364-6613\(02\)00029-3](https://doi.org/10.1016/S1364-6613(02)00029-3)
- Barsalou, L. W., Dutriaux, L., & Scheepers, C. (2018). Moving beyond the distinction between concrete and abstract concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1752), 20170144. <https://doi.org/10.1098/rstb.2017.0144>
- Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin*, 145(11), 1029–1052. <https://doi.org/10.1037/bul0000209>
- Calderhead, J., & Robson, M. (1991). Images of teaching: Student teachers' early conceptions of classroom practice. *Teaching and Teacher Education*, 7(1), 1–8. [https://doi.org/10.1016/0742-051x\(91\)90053-r](https://doi.org/10.1016/0742-051x(91)90053-r)
- Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- Corral, D., Kurtz, K. J., & Jones, M. (2018). Learning relational concepts from within- versus between-category comparisons. *Journal of Experimental Psychology: General*, 147(11), 1571–1596. <https://doi.org/10.1037/xge0000517>
- Corral, D., Carpenter, S. K., & Clingan-Siverly, S. (2021). The effects of immediate versus delayed feedback on complex concept learning. *Quarterly Journal of Experimental Psychology*, 74(4), 786–799. <https://doi.org/10.1177/1747021820977739>
- Darling-Hammond, L. (2006). Constructing 21st-century teacher education. *Journal of Teacher Education*, 57(3), 300–314. <https://doi.org/10.1177/0022487105285962>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160. <https://doi.org/10.3758/brm.41.4.1149>
- Feldon, D. F. (2007). Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational Psychologist*, 42(3), 123–137. <https://doi.org/10.1080/00461520701416173>
- Forzani, F. M. (2014). Understanding “core practices” and “practice-based” teacher education: Learning from the past. *Journal of Teacher Education*, 65(4), 357–368. <https://doi.org/10.1177/0022487114533800>
- Gentner, D., & Kurtz, K. J. (2005). Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *APA decade of behavior series. Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin* (pp. 151–175). American Psychological Association. <https://doi.org/10.1037/11156-009>
- Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin*, 142(7), 729–757. <https://doi.org/10.1037/bul0000043>
- Goldwater, M. B., Don, H. J., Krusche, M. J., & Livesey, E. J. (2018). Relational discovery in category learning. *Journal of Experimental Psychology: General*, 147(1), 1–35. <https://doi.org/10.1037/xge0000387>
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge. <https://doi.org/10.4324/9780203181522>
- Hogan, T., Rabinowitz, M., & Craven, I. I. I., J. A. (2003). Representation in teaching: Inferences from research of expert and novice teachers. *Educational Psychologist*, 38(4), 235–247. https://doi.org/10.1207/s15326985ep3804_3
- Jung, W., & Hummel, J. E. (2015). Making probabilistic relational categories learnable. *Cognitive Science*, 39(6), 1259–1291. <https://doi.org/10.1111/cogs.12199>
- Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, 36(5), 757–798. <https://doi.org/10.1111/j.1551-6709.2012.01245.x>
- Little, J. L., & McDaniel, M. A. (2015). Individual differences in category learning: Memorization versus rule abstraction. *Memory & Cognition*, 43(2), 283–297. <https://doi.org/10.3758/s13421-014-0475-1>
- Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, 129(4), 592–613. <https://doi.org/10.1037/0033-2909.129.4.592>
- Medin, D. L., Lynch, E. B., & Solomon, K. O. (2000). Are there kinds of concepts? *Annual Review of Psychology*, 51(1), 121–147. <https://doi.org/10.1146/annurev.psych.51.1.121>
- Medin, D. L., & Rips, L. J. (2005). Concepts and categories: Memory, meaning, and metaphysics. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (Vol. 137, pp. 37–72). Cambridge University Press. <https://doi.org/10.1093/oxfordhb/9780199734689.013.0011>
- Merrill, M. D., & Twitchell, D. (1994). *Instructional design theory*. Educational Technology Publications

- Miesner, M. T., & Maki, R. H. (2007). The role of test anxiety in absolute and relative metacomprehension accuracy. *European Journal of Cognitive Psychology*, 19(4–5), 650–670. <https://doi.org/10.1080/09541440701326196>
- Murphy, G. (2004). *The big book of concepts*. MIT press. <https://doi.org/10.7551/mitpress/1602.001.0001>
- Prinz, A., Golke, S., & Wittwer, J. (2018). The double curse of misconceptions: Misconceptions impair not only text comprehension but also metacomprehension in the domain of statistics. *Instructional Science*, 46(5), 723–765. <https://doi.org/10.1007/s11251-018-9452-6>
- Prinz, A., Golke, S., & Wittwer, J. (2019). Refutation texts compensate for detrimental effects of misconceptions on comprehension and metacomprehension accuracy and support transfer. *Journal of Educational Psychology*, 111(6), 957–981. <https://doi.org/10.1037/edu0000329>
- Rawson, K. A., & Dunlosky, J. (2016). How effective is example generation for learning declarative concepts? *Educational Psychology Review*, 28(3), 649–672. <https://doi.org/10.1007/s10648-016-9377-z>
- Rawson, K. A., Thomas, R. C., & Jacoby, L. L. (2015). The power of examples: Illustrative examples enhance conceptual learning of declarative concepts. *Educational Psychology Review*, 27(3), 483–504. <https://doi.org/10.1007/s10648-014-9273-3>
- Rittle-Johnson, B., & Star, J. R. (2011). The power of comparison in learning and instruction: Learning outcomes supported by different types of comparisons. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 199–225). Academic Press. <https://doi.org/10.1016/b978-0-12-387691-1.00007-7>
- Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, 4(1), 33–45. <https://doi.org/10.1007/s11409-008-9031-3>
- Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, 68, 1–16. <https://doi.org/10.1007/s11423-019-09701-3>
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, 3087. <https://doi.org/10.3389/fpsyg.2019.03087>
- Wittwer, J., & Renkl, A. (2010). How effective are instructional explanations in example-based learning? A meta-analytic review. *Educational Psychology Review*, 22(4), 393–409. <https://doi.org/10.1007/s10648-010-9136-5>
- Zamary, A., Rawson, K. A., & Dunlosky, J. (2016). How accurately can students evaluate the quality of self-generated examples of declarative concepts? Not well, and feedback does not help. *Learning and Instruction*, 46, 12–20. <https://doi.org/10.1016/j.learninstruc.2016.08.002>
- Zamary, A., & Rawson, K. A. (2018). Which technique is most effective for learning declarative concepts—Provided examples, generated examples, or both? *Educational Psychology Review*, 30(1), 275–301. <https://doi.org/10.1007/s10648-016-9396-9>
- Zepeda, C. D., & Nokes-Malach, T. J. (2021). Metacognitive study strategies in a college course and their relation to exam performance. *Memory & Cognition*, 49, 480–497. <https://doi.org/10.3758/s13421-020-01106-5>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.