Check for updates

Received: 8 April 2020 | Revised: 16 June 2020 | Accepted: 2 July 2020
DOI: 10.1002/hbe2.209

EMPIRICAL ARTICLE

WILEY

Realism as a retrieval cue: Evidence for concreteness-specific effects of realistic, schematic, and verbal components of visualizations on learning and testing

Alexander Skulmowski | Günter Daniel Rey

Psychology of Learning with Digital Media, Chemnitz University of Technology, Chemnitz, Germany

Correspondence
Alexander Skulmowski, Psychology of Learning with Digital Media, Chemnitz University of Technology, Straße der Nationen 12, 09111 Chemnitz, Germany.
Email: alexander.skulmowski@phil.tu-chemnitz.de

Abstract

Previous research in technology-enhanced learning suggests that realism in visualizations can overwhelm learners and lower their learning performance. However, some studies offer evidence that realism may be helpful in learning tests that require detailed visual knowledge. A potential mechanism may be that realistic details act as retrieval cues (i.e., visual cues that facilitate retrieval during testing). In order to assess the impact of realism in learning and testing, we conducted an experiment (N = 40) using realistic and schematic renderings generated for this study. Realism was varied as a between-subjects factor while participants completed both a schematic and realistic retention test. As hypothesized, a realistic test visualization primarily benefits those who had learned using a realistic rather than a schematic visualization. In addition, the realistic test led to higher retention scores regardless of the realism level of the learning materials. A second experiment (n = 50) was conducted to assess whether schematic visualizations can be made more learner-friendly by including verbal descriptions that enhance the less detailed visualizations with additional cues. The experiment revealed that learning with schematic visualizations can be fostered using verbal descriptions for a multiple-choice retention test, but at the cost of lower scores in an image-based retention test. Both studies suggest that matching the realism and concreteness in the learning and testing stages lead to the best performance. Our results have implications for the design of a range of technology-enhanced learning contexts such as virtual reality education.

KEYWORDS

e-learning, concreteness, human cognition, human-computer interaction, instructional design, learning, realism, retrieval cues, retrieval specificity, visualizations

1 | INTRODUCTION

Current technology opens up new avenues in the design of visualizations. Learning materials can be enhanced through realistic graphical representations using various techniques that are becoming more and more accessible, such as photogrammetry (i.e., the automated digitization of real-world objects using photographs) or open-source three-dimensional (3D) modeling and rendering packages. Yet, there still is no consensus on the usefulness of realistic visualizations concerning learning (see Lin, Holmqvist, Miyoshi, & Ashida, 2017; for an overview

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2020 The Authors. *Human Behavior and Emerging Technologies* published by Wiley Periodicals LLC.

on the design of visualizations, see Castro-Alonso, Ayres, & Sweller, 2019). In this paper, we investigate one potential advantage of realistic visualizations, namely their facilitating effect in the retrieval of information. In addition, a second study focuses on how to enhance learning with less detailed visualizations by making them more memorable through verbal descriptions.

1.1 | Realism and concreteness

Research investigating the role of realism in learning usually compares at least two representations differing in their level of visual fidelity to a real-world object (e.g., Dwyer Jr, 1967; Scheiter, Gerjets, Huk, Imhof, & Kammerer, 2009; Skulmowski & Rey, 2020b; for an overview on realism definitions, see Anglin, Vaez, & Cunningham, 2004). The often implicit basis of these comparisons is the notion of a *realism continuum* (Dwyer Jr, 1967). This continuum spans from photographs to abstract visual representations consisting of only a few and basic two-dimensional shapes (Dwyer Jr, 1967). One aspect that should receive appropriate consideration is the comparability of the different visualizations used in previous studies. For example, the studies by Dwyer Jr (1967) and Dwyer, (1971) compared visualizations that not only differed in their degree of realism but also represent entirely different real-world objects (for a meta-analysis on these and related studies by Dwyer, see Reinwein & Huberdeau, 1997). As an example, Dwyer (1971) used, among others, highly abstract line drawings of the heart, detailed drawings of the heart, and photographs of a heart model and an actual heart. Besides the degree of realism, there are other differences between these images used by Dwyer (1971), as the drawings and the heart model also differ in their visual style. As discussed by Skulmowski and Rey (2018), there are many degrees of freedom when choosing schematic and realistic counterparts for a comparison (see also Castro-Alonso, Ayres, & Paas, 2016). Hence, a better method for creating comparable learning materials is to use current 3D modeling and rendering software to generate visualizations that feature models with the same basic shapes and a similar style, yet are rendered in different ways (e.g., realism vs. a schematic, line-drawn look).

As we want to include even more abstract representations in the form of verbal descriptions in the studies of this paper, we wish to enrich the classic realism continuum with recent theoretical accounts used to classify concrete (realistic) and abstract (simplified visual or textual) representations. For instance, Belenky and Schalk (2014) proposed a theoretical system that takes into account the strength of a representation's likeness to a real-world object on the dimension of *groundedness*. They contrast abstract representations (such as numbers) with more concrete representations (such as visualizations) that either include unnecessary details or not (Belenky & Schalk, 2014). Throughout this paper, we will utilize Belenky and Schalk's (2014) theoretical model of groundedness as a way to relate realistic, schematic, and verbal representations to another. While realistic visualizations that are highly detailed fall on the grounded end of the spectrum, schematic visualizations would be considered grounded representations without unnecessary details according to Belenky and Schalk (2014). Verbal descriptions can be seen as the most extreme example of abstraction and are thus least grounded in the real-world appearance of an object. To sum up the definitions and theoretical dimensions combined from Dwyer Jr (1967) and Belenky and Schalk (2014), a realistic visualization will contain more details (that may not necessarily be relevant for the learning task) and thus is the most concrete; schematic visualizations offer only the most relevant details and are less realistic and concrete; verbal descriptions are the most abstract forms of representation and thus feature no pictorial realism.

In the next sections, we will outline three different perspectives on learning with realistic and schematic visualizations. While the first perspective describes that a preference for realistic visualizations can be attributed to a layman's intuition that lacks empirical evidence for the usefulness of realism (Smallman & St. John, 2005), the second perspective claims that realism can be cognitively demanding (e.g., Scheiter et al., 2009). In contrast to these two stances, we highlight ways in which realism may help learners to retrieve information by acting as a memory aid (e.g., Lokka & Çöltekin, 2019).

1.2 | Realism preference as a fallacy

Naive realism assumes that laypeople have an intuitive preference for realistic visualizations that is based on widespread fallacies regarding the function of cognitive processing (Smallman & St. John, 2005). One of these fallacies is thought to be the idea that realistic visualizations are automatically more helpful due to their strong resemblance to their real-world counterparts (Smallman & St. John, 2005). In a variety of studies, it was shown that working with realistic visualizations can be harder, leading to worse results on performance measures (e.g., Hegarty, Smallman, & Stull, 2012; Smallman & St. John, 2005). In a series of studies, weather maps presented in different levels of detail were shown to participants (Hegarty et al., 2012). A substantial percentage of participants preferred maps with irrelevant (and often realistic) details although these maps led to more errors in retrieving and comparing information (Hegarty et al., 2012). In line with such results, research on the naive realism perspective advises against the use of realistic visualizations (Smallman & St. John, 2005). However, it should be noted that research on naive realism has been largely focused on the effects of perceptual cues such as judging the strength of meteorological phenomena based on visual cues in maps (e.g., Hegarty, Smallman, Stull, & Canham, 2009) or estimating the location of objects in terrains from two-dimensional or three-dimensional views (St. John, Cowen, Smallman, & Oonk, 2001). Results from such studies might not be fully generalizable to other fields such as learning.

1.3 | Realism as cognitive load and the problem of comparability

Cognitive load theory (Sweller, van Merrienboer, & Paas, 1998; for an overview in the context of anatomy learning, see Khalil, Paas, Johnson, & Payer, 2005) is one of the most important frameworks for instructional design and can be used to frame research on realism. Cognitive load theory is centered around the idea that learners have a certain mental capacity that can be filled using different types of cognitive load, *intrinsic cognitive load* and *extraneous cognitive load* (Sweller, van Merriënboer, & Paas, 2019). In this theory, intrinsic cognitive load is defined as the cognitive load stemming from the learning content itself while extraneous cognitive load encompasses all detrimental cognitive load arising from the nonoptimal presentation of that content (Sweller et al., 1998). Therefore, the theory holds that learning performance can be optimized through the minimization of extraneous cognitive load (Sweller et al., 1998). Realistic details have consequently been characterized as an example of unnecessary cognitive load (e.g., Brucker, Scheiter, & Gerjets, 2014). Some studies conducted within the field of learning with realistic visualizations provide empirical support for this view. For instance, Scheiter et al. (2009) compared two versions of dynamic visualizations of mitosis: a series of schematic drawings and actual microscopic photographs of the process. They found that the schematic presentation leads to better learning performance than learning with the photographs (Scheiter et al., 2009). A recent paper containing two studies on understanding morphological changes in the development of animals such as bugs similarly found that detailed illustrations interfered with transfer performance (Menendez, Rosengren, & Alibali, 2020). These results were explained with a supposed higher specificity of the information in the more detailed illustration that made it harder to apply this knowledge to new tasks (Menendez et al., 2020).

However, a few design choices of some of the aforementioned studies need to be considered. For instance, Scheiter et al. (2009) used schematic visualizations in the learning tests regardless of the realism level used during the learning phase. As discussed above, there may be issues of comparability between some forms of realistic and schematic visualizations. This issue is exacerbated by differing opinions around the definitions of realism (for an overview, see Anglin et al., 2004). Similarly, it can be hard to determine which information in a learning environment is irrelevant, and therefore, a contributing factor to extraneous load. For instance, the realistic details in the heart photograph used by Dwyer Jr (1967) may be unnecessary for learning factual information concerning the function of the heart but have high relevance for preparing students to recognize the different parts of the heart in their later medical profession. In line with this reasoning, Dwyer Jr (1967) found that learning with photographs aided in image-based tests such as recognition tests while not affecting performance in terminology tests. Such examples underline that the context of a learning task (including the nature of the learning tests) needs to be considered (Hegarty, 2011) when trying to determine which aspects might be extraneous. This aspect further complicates comparisons between different studies and their results.

1.4 | Realism as a benefit for learning and retrieval for specific tests

After considering perspectives favoring a more restrained use of realism, we will now turn to approaches that imply a positive influence of realism during learning and testing. Realism can guide learners' attention by highlighting information (Lokka & Çöltekin, 2019; Skulmowski & Rey, 2020b). Lokka and Çöltekin (2019) demonstrated that a navigational task can be made easier by presenting virtual streets in which only important buildings are rendered realistically, with the majority of irrelevant houses only shown as blocks without textures. This combination of realistic and schematic content improved the learning performance of routes compared to fully realistic and fully schematic virtual displays of the routes in question (Lokka & Çöltekin, 2019). Building on this result, Skulmowski and Rey (2020b) found that in a combined presentation of realistic and schematic visualizations, retention performance was higher for the realistic portions of the learning materials. As presented above, earlier studies in the field of anatomy learning suggest benefits of learning with more realistic and detailed visualizations for specific tests (e.g., Dwyer Jr, 1967). Other authors emphasize that realism can have divergent effects on different types of tests. For example, Belenky and Schalk (2014) conclude that more concrete representations may particularly help learners during the initial comprehension stage. In contrast, more abstract representations are said to be more beneficial for transfer tasks and to comprehend the essential relations in more complex learning materials (Belenky & Schalk, 2014).

In this paper, we will focus on the aspect of cues during learning and testing. It is a well-established finding that the presentation of a secondary stimulus during learning and testing—known as *retrieval cue*—can improve recall results for a primary stimulus (Tulving & Osler, 1968). This finding was transferred into the field of multimedia learning by showing that added images or videos can help in a word-learning task (Plass, Chun, Mayer, & Leutner, 1998). It has also been long known that the encoding processes active during learning play a role during retrieval (Tulving & Thomson, 1973). A related study in which paired word lists were learned resulted in higher recall performance for semantically related words (Roediger & Adelson, 1980). When applied to the field of multimedia learning, a recent series of studies found that even primarily decorative pictures can enhance retrieval (Schneider, Nebel, Beege, & Rey, 2019). An older study by Dwyer Jr (1967) found that learning with a photograph led to better performance in a comprehension test and a drawing test concerning the human heart compared to more schematic alternatives. An explanation for these positive effects of realism is that more detailed visualizations offer a higher number of perceptual cues (Dwyer Jr, 1967). Given these results, we assume that richer visual details contained in realistic visualizations may act as a retrieval cue in the context of an anatomy learning task. This research question will be the main focus of Experiment 1.

It needs to be noted that the more positive perspectives on realism do not necessarily contradict the aforementioned findings related to naive realism and cognitive load theory but rather suggests that perceptual richness may be helpful in certain learning situations (see Hegarty, 2011, for an overview of task-dependent factors in the design of visualizations). Realism might be used as a means of presenting more perceptual information that may help in particular types of learning tasks that heavily rely on details found in visualizations (see Dwyer Jr, 1967). Thus, it may still be a valid result that realism can be detrimental when learning a biological *process* as shown by Scheiter et al. (2009), but when creating a basic mental model of an anatomical component (including the generation of mental relations between parts and their names), we assume that realistic details can be helpful depending on the type of test. In Experiment 1, the retention tests consist of cued-recall tests that ask learners to assign labels to an image. Based on previous research (e.g., Dwyer Jr, 1967; Skulmowski & Rey, 2020b), this type of test should result in a particularly strong benefit of realistic visualizations when learners are tested using similarly realistic retention tests.

1.5 | Adding textual cues to facilitate the comprehension of visualizations

Besides the role of realistic visualizations in learning and testing, we also assess how to make schematic visualizations that may be lacking in distinctive details more memorable. Based on previous research that combined text and image cues (e.g., Beck, 1984), we investigate whether abstract diagrams can be easier to remember when they are enhanced with short verbal descriptions. In a recent series of studies, Schwartz and Yovel (2019) found that the learning performance for faces was higher when participants were prompted to answer questions triggering a more conceptual processing mode (e.g., asking participants to rate the intelligence of faces) rather than relying on perceptual cues (such as asking participants to rate the symmetry of faces). Based on their results, Schwartz and Yovel (2019) conclude that facial recognition can be enhanced by inducing a more elaborate processing mode beyond mere perceptual cues. As schematic anatomical visualizations such as simplified line drawings offer only a limited amount of perceptual cues, we figured that the results of Schwartz and Yovel (2019) could be relevant for learning with schematic visualizations.

Several forms of visual cues have been shown to facilitate the comprehension of biological learning content (e.g., Patrick, Carter, & Wiebe, 2005). In particular, the combination of visual and text-based cues has been found to foster learning compared with conditions featuring no cues (Beck, 1984). We conclude that adding verbal descriptions helps learners in the process of generating a mental model by offering a short, easy-to-remember statement concerning the shape to be learned. In contrast. as summarized by Schnotz (2005), presenting information pictorially and verbally at the same time may lead to a disadvantage due to redundancy (for a study on the negative effects of redundancy, see Kalyuga, Chandler, & Sweller, 1999). However, we assume that similar to the results of Schwartz and Yovel (2019), the verbal description will not act as a form of redundant load, but rather as a help in mentally organizing the content of a schematic visualization.

1.6 | The present studies

As we have presented above, the results of using different levels of realism in learning and testing are not in unison. We argue that some of the positive results of schematic visualizations are context-dependent and need to be reexamined in a different learning task with other learning tests. To improve comparability, we created learning materials that only differed in their level of realism for Experiment 1. Unlike older studies on the issue of realism in learning that, among others, compared drawings with photographs (e.g., Dwyer Jr, 1969), we used 3D renderings of anatomical models specifically created using current technology. Additionally, we used mixed repeated-measures designs to increase statistical power. Since we are interested in differences in performance due to the realism level in learning tests, our studies employ a design in which participants are tested using visualizations and representations in different levels of realism (see Experiment 1) and concreteness (varied through the use of image-based and verbal tests in Experiment 2).

2 | EXPERIMENT 1

As the learning content of Experiment 1, we selected the anatomy of the parotid gland. The parotid gland features few distinct shapes in its overall appearance and thus, we figured that a retention test featuring surface details may act as a retrieval cue for learners who had previously learned using a realistic visualization. We did not measure cognitive load using a multifaceted survey instrument due to previous findings that cognitive load scores for realistic visualizations can diverge from the corresponding learning outcomes (referred to as the *realism paradox* by Skulmowski & Rey, 2020b) and that the measurement of extraneous load can be affected by the phrasing of survey items (Skulmowski & Rey, 2020a). As previous studies comparing learning tests used ratings of difficulty (e.g., Prisacari & Danielson, 2017), we decided to add a matching hypothesis for difficulty ratings of the tests. Consequently, we arrived at two hypotheses:

Hypothesis 1 *A retention test featuring realistic visualizations will lead to particularly high scores for those learners that had used the realistic learning materials compared to schematic visualizations*.

Hypothesis 2 *A retention test featuring realistic visualizations will be rated as particularly easy by those learners that had used the realistic learning materials compared to schematic visualizations*.

2.1 | Methods

2.1.1 | Participants and design

For this study, we chose a 2×2 mixed factorial design in which the degree of realism during the learning and testing phases was varied. The between-subjects factor was the level of realism during learning (schematic vs. realistic; block-randomized assignment) while the within-subjects factor was the level of realism during testing (schematic vs. realistic; the order was alternated sequentially between participants).

Forty participants (28 female, 12 male) were tested in a laboratory study based on a power calculation using G*Power (Version 3.1.9.2; Faul, Erdfelder, Buchner, & Lang, 2009) at an alpha level of .05, power = .80, correlation between measures = .5, and a η_p^2 = .05. The participants were students at a German university enrolled in Media Communication, Computer Science and Communication Sciences, or the graduate course Media Psychology and Instructional Psychology. Their participation counted as partial fulfillment of course requirements. In order to participate, the students needed to be between 18 and 30 years old, native speakers of German, and should have no to little knowledge of parotid gland anatomy. The two between-subjects groups each had 20 participants assigned to them.

2.1.2 | Materials

During the learning phase, one of two renderings of the parotid gland with 11 labels was presented to participants (see Figure 1). The realistic version featured a high-resolution model with detailed texture maps and specular shading (see Figure 1a). For the schematic version (see Figure 1b), a lower resolution model with the same basic shape as the realistic model was used. Instead of realistic shading, a "toon"

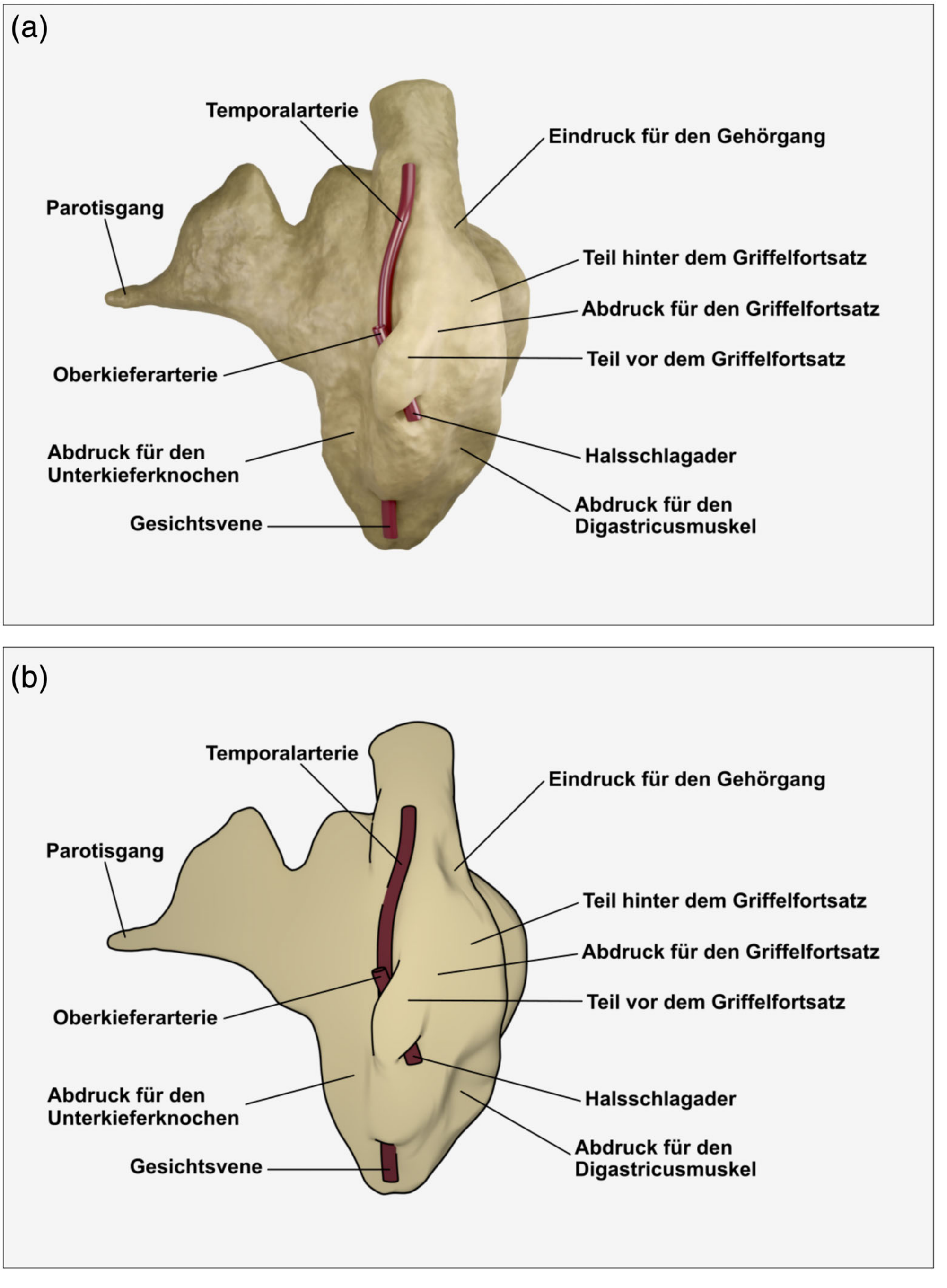


FIGURE 1 The learning materials presented in Experiment 1. Panels (a) and (b) depict the parotid gland in a realistic and a schematic rendering mode, respectively (based on Gray, 1918, with additional information taken from Bammes, 2009; Cole, 2015; Tillmann, 2016)

shader was applied to simplify the shading to look more like a drawing. A contour outline was applied to further transform the rendering into an image resembling a schematic diagram. Materials were created using Blender 2.79b and Inkscape 0.92.4. The study did not use additional spoken or written learning materials.

The learning tests consisted of two sets of two pages showing two different cropped versions of the renderings used in the learning phase (see Figure 2). Each participant was tested using two pages featuring the realistic rendering (see Figure 2a,b) and two pages showing the schematic version. Each rendering was labeled with eight letters (comparable to the test in Skulmowski & Rey, 2018). Participants were asked to drag the correct label to the corresponding letter below the respective test image. In addition, participants were notified that some of the labeled components of the test images might not have carried a label in the learning phase and should be left unlabeled. The testing phase did not have a time limit. The maximum score for each test (schematic and realistic) was 22 points. One point was awarded for each correctly assigned label and for each component that was correctly left unlabeled. In terms of reliability, the schematic test had a McDonald's ω (McDonald, 1999; calculated using JASP Version 0.9.2; JASP Team, 2019) of .71, the realistic test had a ω of .75.

2.1.3 | Procedure

Participants provided informed consent and entered some personal information (prior knowledge, age range, gender, and course of study). Prior knowledge was assessed by having participants confirm that they only had no to little knowledge of the anatomy of the parotid gland. Participants indicated whether they fell into the required age range. They were instructed regarding the learning task on an instruction page that stated that the task was to memorize the names, shapes, and locations of the parts of the human parotid gland. The learning task lasted 60 s (a countdown was displayed) and participants were immediately directed to a filler task afterwards. In the 60 s filler task, they were asked to order the 16 German federal states according to their number of universities of applied sciences. The filler task was inserted to adjust the difficulty of the learning test in order to avoid ceiling effects. The duration of the learning phase and filler task was selected based on previous studies with a similar procedure (e.g., Skulmowski & Rey, 2020b). The order of the realistic and schematic tests was alternated sequentially between participants while the two pages of these two tests had a fixed order. After the presentation of the retention tests (without time limit), participants were asked to rate the difficulty of answering the schematic and realistic learning tests (each on a 9-point scale ranging from *very easy* to *very difficult*; using an adapted version of Syring et al.', 2015 translation of Marcus, Cooper, & Sweller's, 1996 difficulty question item). This approach follows previous research in the field of learning (e.g., Prisacari & Danielson, 2017). On the next page, participants were asked to indicate whether they were strongly distracted during the learning phase and whether there had been major technical issues during the learning phase (taken from Skulmowski & Rey, 2020b). The laboratory experiment used SoSci Survey (Version 3.2.03-i; Leiner, 2019; https://www.soscisurvey.de).

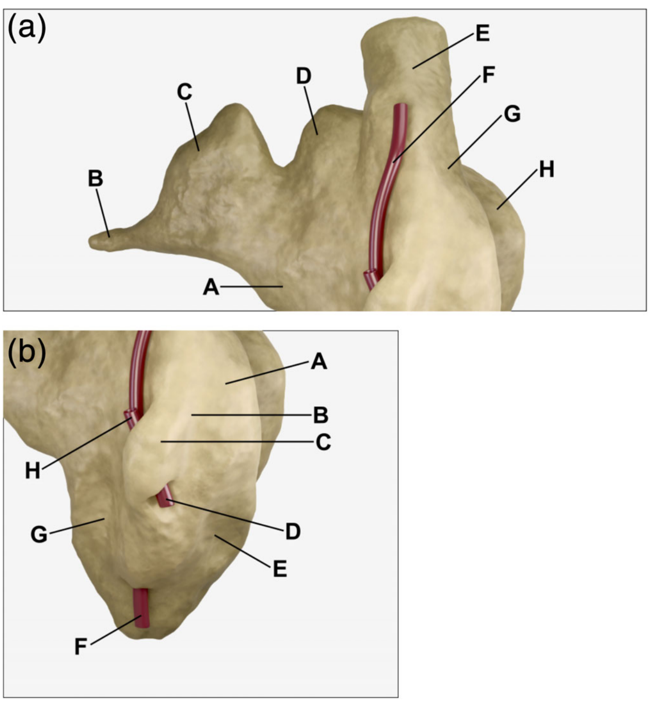


FIGURE 2 The learning tests used in Experiment 1. Panels (a) and (b) were shown in two separate pages. A second version of these two test pages was presented to all participants in the schematic style seen in Figure 1b

2.2 | Results and discussion

Unless otherwise noted, the analyses in this paper were performed using R (Version 3.6.0; R Core Team, 2019). We generated the plots using ggplot2 (Wickham, 2016). Untransformed descriptive data are presented in Figure 3. As no participant reported having been distracted or having faced technical difficulties, no data were removed from the analysis.

2.2.1 | Retention performance

Retention performance was analyzed using a mixed 2 × 2 ANOVA (realism during the learning phase × realism during the testing phase, see Figure 3a). As the Shapiro–Wilk test performed on the residuals of the 2 × 2 ANOVA reached significance (p = .026), we used nonparametric aligned rank transformations (Fawcett & Salter, 1984) using the ARTool R package (Version 0.10.6; Wobbrock, Findlater, Gergle, & Higgins, 2011).

The nonparametric ANOVA resulted in a significant interaction effect, $F(1, 38) = 6.27$, $p = .017$, $\eta_p^2 = .14$. A realistic retention test was particularly helpful for learners who had learned using the realistic version of the learning materials. Hypothesis 1 was thus confirmed by the data of our experiment. In addition, we found a significant main

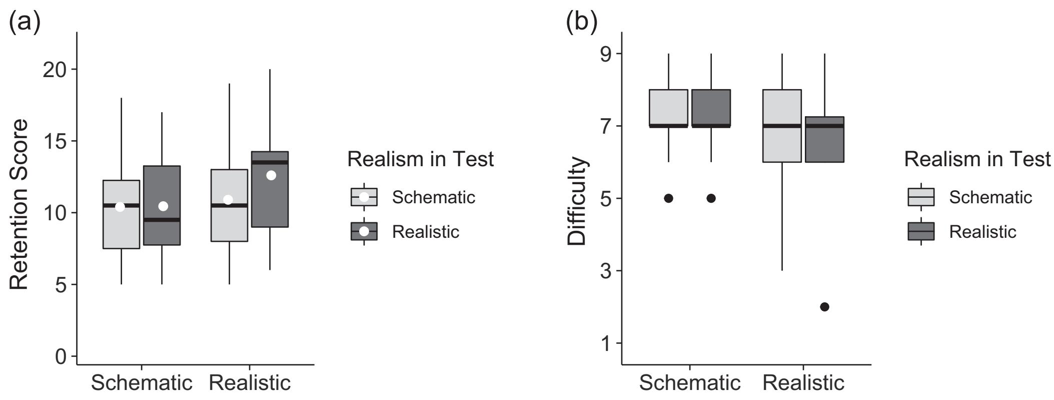


FIGURE 3 The untransformed descriptive results of Experiment 1. The realism of the learning materials is the between-subjects factor while the level of realism of the tests is the within-subjects factor. (a) Boxplot of retention scores with a minimum of 0 and a maximum of 22 per retention test. The white points denote the means. (b) Boxplot of difficulty ratings with a minimum of 1 and a maximum of 9

effect of the test type, $F(1, 38) = 7.63$, $p = .009$, $\eta_p^2 = .17$, indicating that the realistic learning test resulted in a higher retention performance. The factor of realism during the learning phase did not reach significance ($p = .319$). These results suggest that a higher level of realism during testing can be beneficial, perhaps by providing unambiguous visual points of reference that aid in retrieval. As the positive effect of realism during testing was particularly high for those learners who had already learned using the realistic version, we assume that there is a specificity effect of the concreteness of visualizations. A match between a high level of realism during learning and testing appears to lead to a superior retention performance.

2.2.2 | Difficulty ratings

The 2 × 2 ANOVA (realism during the learning phase × realism during the testing phase, see Figure 3b) for the difficulty ratings also required a nonparametric test due to a significant Shapiro–Wilk test performed on the residuals ($p = .002$). The interaction effect of the aligned rank transformed ANOVA data did not reach significance ($p = .338$). Therefore, Hypothesis 2 was not confirmed.

3 | EXPERIMENT 2

Based on the results of the first experiment, we sought to investigate how to improve learning and testing with less detailed visualizations that are often found in technology-enhanced learning environments. As we found that schematic visualizations may lead to worse results when they are used in the testing phase, we intended to assess whether a major drawback of less detailed visualizations—their lack of distinct visual information—can be remedied by including verbal descriptions in the learning materials (for overviews on combining visualizations and texts, see Castro-Alonso et al., 2019; Schnotz, 2005).

As we chose the human scapula as the learning content, an example of how such a statement could help learners would be the shape of the supraspinatous fossa, a triangular shape in the top portion of the scapula (see Figure 4). Without a verbal description, a novice learner looking at a schematic drawing of the scapula with several labels must first determine which shapes the labels refer to. In the case of the supraspinatous fossa, a learner may assume that the label applies to the entire triangular shape or perhaps just the part the label is connected to (in our learning materials shown in Figure 4, this could be just the center of the triangular shape). Similar to the study by Schwartz and Yovel (2019) a verbal description of the supraspinatous fossa as a triangular shape on the top right side of the scapula might save learners from unnecessary visual searches and inferences. Furthermore, the verbal description might help learners to generate an even more abstract, verbal model of the learning targets that can be particularly helpful for completing a text-based retention test. However, given the time limit of our learning task, we assume that the limited cognitive capacity postulated within cognitive load theory may lead to a diminished focus on the diagram if verbal descriptions are presented alongside the drawing. This assumption is in line with previous results that pictures are primarily viewed in the very first stages of learning if they are presented alongside texts (Lenzner, Schnotz, & Müller, 2013).

Due to the task-dependent usefulness of visualizations (see Hegarty, 2011), we hypothesized that the benefits of verbal descriptions would be test-dependent. Thus, we included two tests, one image-based labeling test as used in Experiment 1 and a text-based test.

Hypothesis 3 *Learners receiving an enhanced diagram with verbal descriptions achieve higher retention scores on a text-based test but lower scores on an image-based test compared to learners not receiving verbal descriptions.*

3.1 | Methods

3.1.1 | Participants and design

We used a 2 × 2 mixed factorial design with the between-subjects factor verbal descriptions (with vs. without; block-randomized assignment) and the within-subjects factor test type (image-based vs. text-based; the order was alternated sequentially between participants) for this laboratory experiment.

The data of 50 participants (42 female, 8 male) were collected in order to reach the target sample size computed using G*Power (Version 3.1.9.2) at an alpha level of .05, power = .80, correlation between measures = .5, and a η_p^2 = .04. Twenty-four participants were assigned to the group without verbal explanations while the remaining 26 participants were presented with the additional descriptions during the learning phase. The datasets of three additional participants were excluded due to their response of having been strongly distracted during the learning phase. The courses of study and participation conditions were identical to Experiment 1 except that participants needed to have no to little knowledge of the anatomy of the scapula.

3.1.2 | Materials

The learning phase featured a labeled schematic drawing of the human scapula (see Figure 4) created using Inkscape 0.92.4 that either had additional verbal descriptions (see Figure 4b) or not (see Figure 4a). A translated example for a verbal explanation would be "Small protuberance in the upper part" for the coracoid process. The learning tests were divided into one page of multiple-choice questions and another page featuring a similar labeling task as utilized in Experiment 1 (Figure 5).

In the creation of the multi-choice questions, we avoided using exact verbatim copies of the verbal descriptions displayed during the learning phase by using synonyms, similar phrases, or relying on entirely different descriptions. Hence, the group who was presented with the verbal descriptions was still required to study the diagram to fully understand the learning content. Example question items would be "What is the name of the small, hook-shaped process in the upper part?" for the coracoid process and "What is the name of the bony process at the top of the spine of the scapula?" for the acromion process. Participants received one point per correctly answered multiple-choice question, for each correctly labeled item, and for each item that was correctly left unlabeled. Reliability scores of the two tests (calculated using JASP Version 0.9.2) were ω = .76 for the image-based test (one item had to be dropped due to a variance of 0) and ω = .59 for the text-based test.

3.1.3 | Procedure

As in Experiment 1, participants provided their informed consent and were asked to enter some personal information (prior knowledge, age range, gender, and course of study). On an instruction page, the learning task was outlined similar to Experiment 1. The duration of the learning task was increased by 30 s compared to Experiment 1 to a total of 90 s for both groups in order to make it possible to read the verbal descriptions. In the filler task page to which participants were redirected automatically, their objective was to order the 16 German federal states according to their number of elementary schools within 60 s. The multiple-choice and image-based tests (presented without any time limit) were filled out in a fixed order to avoid an additional chance to restudy the diagram before answering the text-based test. Afterward, participants answered the questions regarding distractions and technical difficulties as in Experiment 1. SoSci Survey (Version 3.2.03-i; Leiner, 2019; https://www.soscisurvey.de) was used for this laboratory experiment.

3.2 | Results and discussion

3.2.1 | Retention performance

As no ANOVA assumptions were violated, we computed a 2×2 mixed ANOVA with the factors verbal descriptions and test type (see Figure 6). The hypothesized interaction effect reached significance, $F(1, 48) = 4.10$, $p = .049$, $\eta_p^2 = .08$ (no other effects reached significance, $ps > .601$). As hypothesized, retention performance in the text-based multiple-choice test was higher if verbal descriptions had been seen during the learning phase while the image-based test elicited higher scores from those participants who had learned without the additional verbal cues. This result implies that the effects of verbal descriptions in schematic diagrams are dependent on the format of testing.

4 | GENERAL DISCUSSION

The results of our two laboratory experiments generally suggest a specificity effect of learning with abstract and concrete representations. In Experiment 1, we compared the effects of using realistic and schematic visualizations during the retrieval phase. We additionally varied whether participants learned using one of the realistic or schematic visualizations they would later see during the testing phase. The experiment revealed that working with a realistic visualization during the testing phase can enhance retention performance, in particular for learners that already used a realistic visualization during learning. As the highest performance was achieved when both the learning phase and the testing phase used a realistic rendering, we assume that the knowledge gained from visualizations is quite specific; favoring a match between the degree of realism used in the learning and retrieval phases. A similar match was found in Experiment 2 for schematic visuals, verbal descriptions and different test formats. Learning with less detailed schematic visualizations can be enhanced using verbal descriptions, but only for a text-based test and at the expense of lower performance in an image-based test. Taken together, the results of both studies have several implications for learning with visualizations in the context on technology-enhanced learning.

In Experiment 1, a higher level of realism may have contributed to a more comprehensive mental model that can be retrieved particularly well if the visualizations of the retention test matches the level of

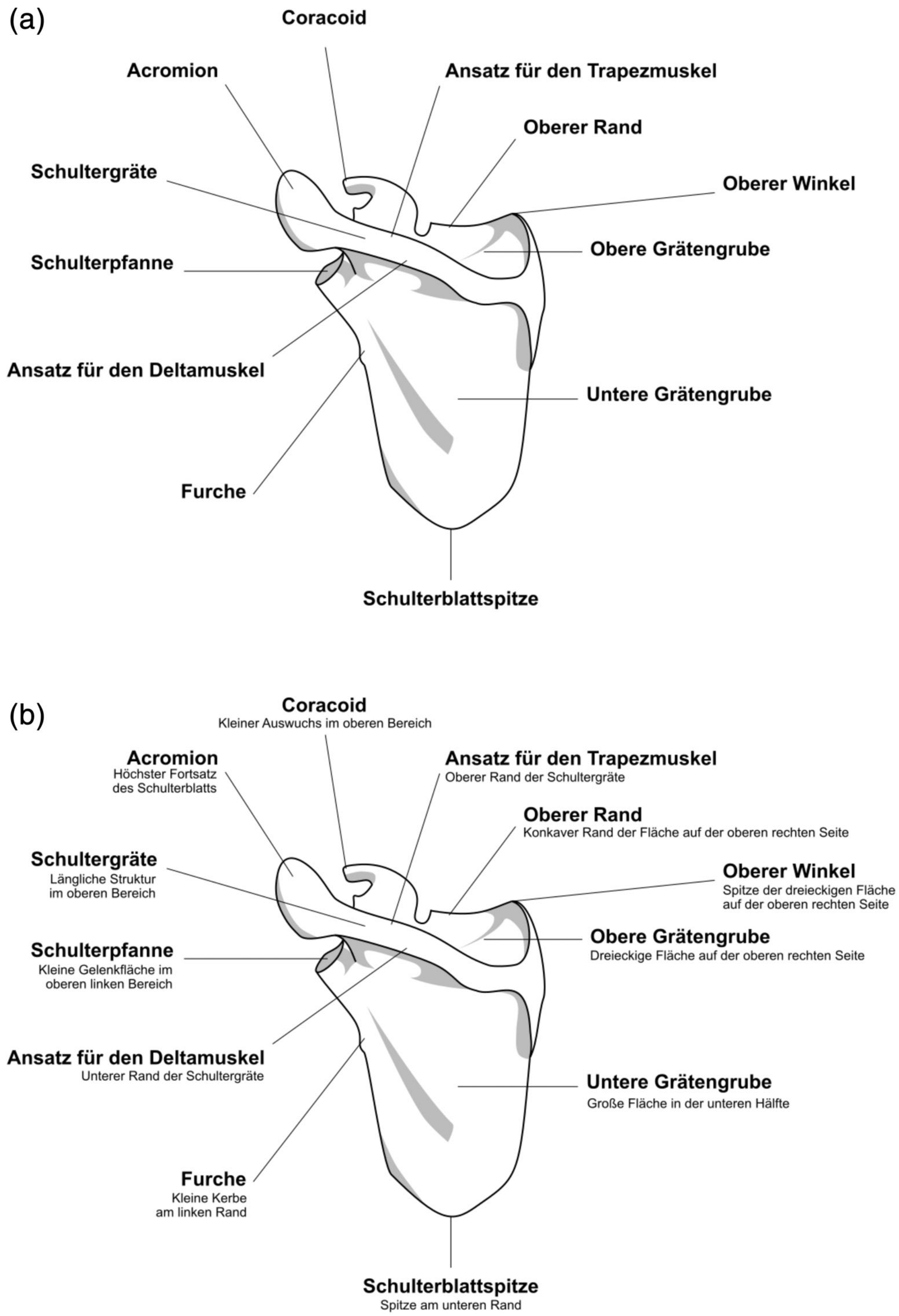


FIGURE 4 The learning materials in Experiment 2 show the human scapula with (b) or without (a) verbal descriptions (based on Gray, 1918, with additional information from Bammes, 2009; Cole, 2015)

realism of the learning phase. The study revealed that realistic details presented during testing can enhance retrieval compared to a rendering primarily consisting of an outline with rudimentary shading. This finding is independent of the degree of realism during the learning stage. Thus, we assume that realism can act as retrieval cues for cued-recall, image-based retention tests, even when a simplified visualization had been studied. One reason for the retention enhancement through realistic visualizations during testing could be the nature of our learning content. The parotid gland has a rather amorphous shape, and thus, participants may have relied upon surface detail to generate and retrieve a mental representation of the object. We assume that the higher amount of details in the realistic version reduced ambiguities during the labeling task and thus helped both groups regardless of the degree of realism during the learning phase. Perhaps, the realistic rendering in the test was also easier to match to participants' mental model based on their general prior knowledge of what a gland looks like.

As Experiment 1 showed an advantage for realistic visualizations in technology-enhanced learning, we sought to assess whether learning with less detailed visualizations could be aided by verbal descriptions in Experiment 2. The basic idea was that written descriptions of

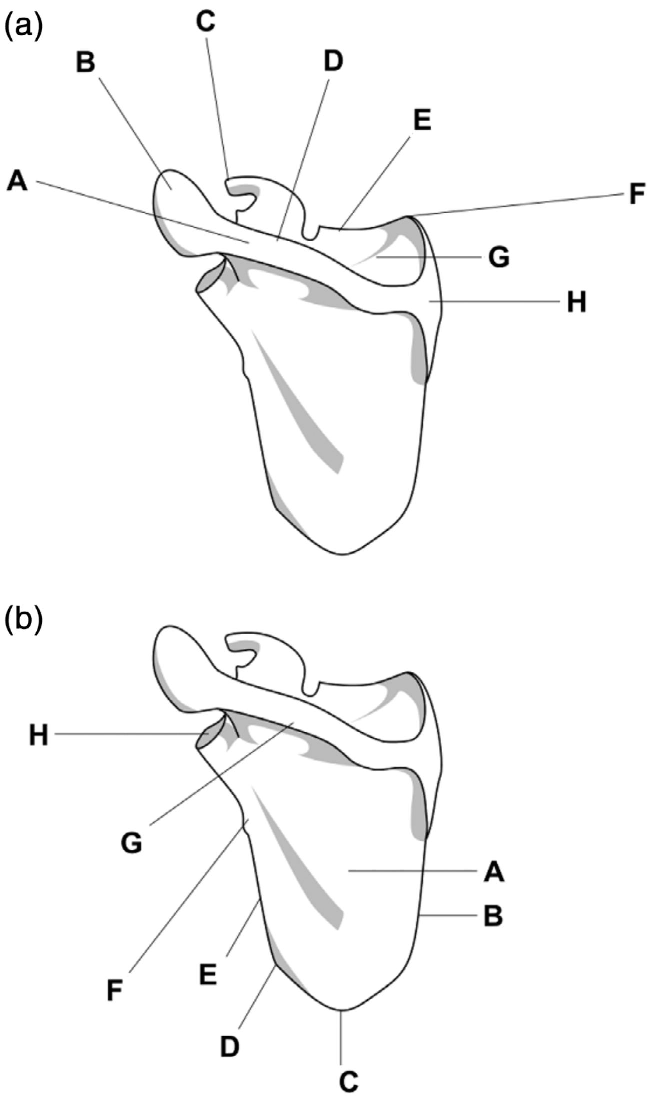


FIGURE 5 The visual learning tests used in Experiment 2. Panels (a) and (b) were presented on the same page

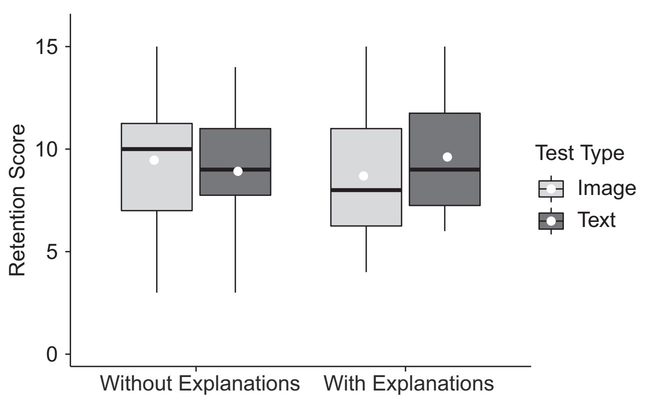


FIGURE 6 Boxplot of the retention scores of Experiment 2 with a minimum of 0 and a maximum of 16 per test. The white points denote the means

the abstract shapes in a diagram of the scapula would help learners categorize and identify the parts of that bone. Furthermore, the description should provide learners with a verbal cue that helps them generate a more memorable association for the rather abstract shapes found in schematic visualizations. As we expected another specificity effect, we hypothesized that the effects of verbal descriptions will be dependent on the test format. In accordance with the hypothesis, verbal descriptions helped participants concerning their performance on a text-based test, while participants learning without the additional descriptions scored higher on an image-based test.

Taken together, the results indicate that learning with visualizations is concreteness-specific. In particular, realistic visualizations appear to prepare learners best for situations in which detailed knowledge concerning the appearance of an object is required (confirming some earlier results of Dwyer Jr, 1967), while verbal descriptions are a benefit for tests focusing on written questions.

4.1 | Implications for the design of retention tests

The most important practical implication of our studies is that the degree of concreteness used in learning tests can affect learners' test performance. As we have shown, there could be a considerable difference in retention performance if, for example, students in two anatomy classes are compared using tests differing in their level of realism, potentially leading to biased test scores. The same conclusion needs to be considered when comparing the results of different studies of the effects of realism.

Most importantly, a practical implication is that testing using realistic visualizations is particularly helpful for learners who have learned with realistic renderings. Further studies should explore whether the cause of this finding is a byproduct of realistic visualizations being less ambiguous and therefore having a facilitating effect during testing.

On a more general note, our result should not be misunderstood to imply that more realistic pictorial representations necessarily lead to better test scores under all circumstances. Excessive amounts of detail may hinder the recognition of shapes (see, e.g., the discussion of perception theories in Skulmowski & Rey, 2018) and could thus be counterproductive (see, e.g., Scheiter et al., 2009). Importantly, the design of a task may have an enormous impact on performance issues when using visualizations (Hegarty, 2011). Consequently, more research with clearly defined levels of realism in various learning tasks with different tests will be needed before arriving at more precise guidelines.

Our results need to be considered in the context of previous research on learning with realistic visualizations that has shown differences based on spatial ability (e.g., Brucker et al., 2014; Huk, 2006) and prior knowledge (Joseph & Dwyer, 1984). Based on the result that learners with a higher spatial ability benefit more from realistic visualizations than individuals with lower spatial abilities (Huk, 2006; see Höffler, 2010, for a meta-analysis), the results of Experiment 1 should be replicated with learners differing in their spatial ability. Furthermore, it has been found that preferences regarding pictorial and textual information can affect learning performance (Koć-Januchta, Höffler, Thoma, Prechtl, & Leutner, 2017). Since learners preferring visual information over texts learn better when they use pictures instead of verbal materials (Koć-Januchta et al., 2017), the result we found concerning the matching of test types to the information format found in visualizations may be affected by learner preferences as well.

4.2 | Limitations and outlook

The stability and generalizability of using realistic details as retrieval cues need to be assessed in follow-up studies that could use variations of our study design. For instance, it would be informative to see whether the advantages of realistic visualizations in the testing phase can be found in cases in which the visualizations in the learning and testing phases differ from each other, for example, by using a slightly different camera angle. Since we only had participants with low prior knowledge, the effect should be replicated with groups of learners with different levels of expertise as this factor can affect learning with realistic visualizations (Joseph & Dwyer, 1984). Furthermore, as Belenky and Schalk (2014) describe that transfer tasks may benefit more from learning with abstract visualizations (contrary to previous results of Dwyer Jr, 1967), this question would be interesting to assess using learning materials such as those used in Experiment 1.

While the duration of our learning tasks was rather short, the results apply to several real-world situations. For instance, university lectures often consist of a serious of diagrams that are shown in rapid succession with only a few minutes in-between presentation slides. Some lectures already incorporate quick (technology-enhanced) testing sessions which can take on a similar format as our studies. However, additional research is necessary to see whether the effect is generalizable to longer phases of studying. A further potential extension of our findings could be to assess whether our results transfer to learning in a virtual reality environment; our results should be replicated in the context of virtual reality instruction.

4.3 | Conclusion

From our results, we can conclude that the degree of realism and concreteness used in learning tests can have a substantial effect on retention performance. Learners who had learned with a realistic visualization achieved considerably higher scores when being tested using a realistic rather than a schematic visualization. Despite multiple theoretical accounts generally advising to avoid realistic instructional visualizations, our study revealed that realism may have the potential to positively affect retrieval performance under certain circumstances. Moreover, we have provided evidence that there are specificity effects in learning with schematic visualization and text-based cues. In short, learning with realistic renderings leads to the highest learning scores when the test is equally realistic, while written descriptions particularly help learners to respond to a text-based test. These results should be considered when choosing tests in practical settings as well as in research.

ACKNOWLEDGMENT

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

A.S. designed both studies in cooperation with G.D.R. A.S. designed the learning materials and collected the data. A.S. conducted the statistical analyses. A.S. created the first draft of the manuscript, G.D.R. provided revisions. A.S. and G.D.R. have read and approved the manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

The research project was reviewed by the ethics commission of the Faculty of Humanities of Chemnitz University of Technology. No individual review of the studies in the project was required.

ORCID

Alexander Skulmowski https://orcid.org/0000-0002-1682-021X

REFERENCES

Anglin, G. J., Vaez, H., & Cunningham, K. L. (2004). Visual representations and learning: The role of static and animated graphics. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 865–916). Mahwah, NJ: Lawrence Erlbaum.

Bammes, G. (2009). *Die Gestalt des Menschen*. Freiburg, Germany: Christophorus Verlag.

Beck, C. R. (1984). Visual cueing strategies: Pictorial, textual and combinational effects. *ECTJ*, *32*, 207–216.

Belenky, D. M., & Schalk, L. (2014). The effects of idealized and grounded materials on learning, transfer, and interest: An organizing framework for categorizing external knowledge representations. *Educational Psychology Review*, *26*, 27–50.

Brucker, B., Scheiter, K., & Gerjets, P. (2014). Learning with dynamic and static visualizations: Realistic details only benefit learners with high visuospatial abilities. *Computers in Human Behavior*, *36*, 330–339.

Castro-Alonso, J. C., Ayres, P., & Paas, F. (2016). Comparing apples and oranges? A critical look at research on learning from statics versus animations. *Computers & Education*, *102*, 234–243.

Castro-Alonso, J. C., Ayres, P., & Sweller, J. (2019). Instructional visualizations, cognitive load theory, and visuospatial processing. In J. C. Castro-Alonso (Ed.), *Visuospatial processing for education in health and natural sciences* (pp. 111–143). Cham, Switzerland: Springer.

Cole, T. C. H. (2015). *Wörterbuch der Biologie/dictionary of biology*. Berlin, Germany: Springer Spektrum.

Dwyer, F. M. (1971). Color as an instructional variable. *AV Communication Review*, *19*, 399–416.

Dwyer, F. M., Jr. (1967). The relative effectiveness of varied visual illustrations in complementing programed instruction. *The Journal of Experimental Education*, *36*, 34–42.

Dwyer, F. M., Jr. (1969). The effect of varying the amount of realistic detail in visual illustrations designed to complement programmed instruction. *Programmed Learning and Educational Technology*, *6*, 147–153.

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.

Fawcett, R. F., & Salter, K. C. (1984). A Monte Carlo study of the F test and three tests based on ranks of treatment effects in randomized block designs. *Communications in Statistics: Simulation and Computation*, *13*, 213–225.

Gray, H. (1918). *Anatomy of the human body* (20th ed.). Philadelphia, PA: Lea & Febiger. Retrieved from https://hdl.handle.net/2027/uc2.ark:/13960/t9m32qk0z

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science*, *3*, 446–474.

Hegarty, M., Smallman, H. S., & Stull, A. T. (2012). Choosing and using geospatial displays: Effects of design on performance and metacognition. *Journal of Experimental Psychology: Applied*, *18*, 1–17.

Hegarty, M., Smallman, H. S., Stull, A. T., & Canham, M. S. (2009). Naïve cartography: How intuitions about display configuration can hurt performance. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *44*, 171–186.

Höffler, T. N. (2010). Spatial ability: Its influence on learning with visualizations—A meta-analytic review. *Educational Psychology Review*, *22*, 245–269.

Huk, T. (2006). Who benefits from learning with 3D models? The case of spatial ability. *Journal of Computer Assisted Learning*, *22*, 392–404.

JASP Team (2019). JASP (Version 0.9.2) [Computer software]. Retrieved from https://jaspstats.org.

Joseph, J. H., & Dwyer, F. M. (1984). The effects of prior knowledge, presentation mode, and visual realism on student achievement. *The Journal of Experimental Education*, *52*, 110–121.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, *13*, 351–371.

Khalil, M. K., Paas, F., Johnson, T. E., & Payer, A. F. (2005). Interactive and dynamic visualizations in teaching and learning of anatomy: A cognitive load perspective. *The Anatomical Record Part B: The New Anatomist*, *286*, 8–14.

Koć-Januchta, M., Höffler, T., Thoma, G. B., Prechtl, H., & Leutner, D. (2017). Visualizers versus verbalizers: Effects of cognitive style on learning with texts and pictures–An eye-tracking study. *Computers in Human Behavior*, *68*, 170–179.

Leiner, D. J. (2019). SoSci Survey (Version 3.2.03-i) [Computer software].

Lenzner, A., Schnotz, W., & Müller, A. (2013). The role of decorative pictures in learning. *Instructional Science*, *41*, 811–831.

Lin, Y. Y., Holmqvist, K., Miyoshi, K., & Ashida, H. (2017). Effects of detailed illustrations on science learning: An eye-tracking study. *Instructional Science*, *45*, 557–581.

Lokka, I. E., & Çöltekin, A. (2019). Toward optimizing the design of virtual environments for route learning: Empirically assessing the effects of changing levels of realism on memory. *International Journal of Digital Earth*, *12*, 137–155.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology*, *88*, 49–63.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Menendez, D., Rosengren K. S., Alibali, M. W. (2020). Do details bug you? Effects of perceptual richness in learning about biological change. *Applied Cognitive Psychology*, 1–17. http://dx.doi.org/10.1002/acp.3698.

Patrick, M. D., Carter, G., & Wiebe, E. N. (2005). Visual representations of DNA replication: Middle grades students' perceptions and interpretations. *Journal of Science Education and Technology*, *14*, 353–365.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology*, *90*, 25–36.

Prisacari, A. A., & Danielson, J. (2017). Computer-based versus paper-based testing: Investigating testing mode with cognitive load and scratch paper use. *Computers in Human Behavior*, *77*, 1–10.

R Core Team. (2019). *R: A language and environment for statistical computing. R foundation for statistical computing (Version 3.6.0) [Computer software]*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Reinwein, J., & Huberdeau, L. (1997). A second look at Dwyer's studies by means of meta-analysis: The effects of pictorial realism on text comprehension and vocabulary. (ERIC Document Reproduction Service No. ED 407 671).

Roediger, H. L., & Adelson, B. (1980). Semantic specificity in cued recall. *Memory & Cognition*, *8*, 65–74.

Scheiter, K., Gerjets, P., Huk, T., Imhof, B., & Kammerer, Y. (2009). The effects of realism in learning with dynamic visualizations. *Learning and Instruction*, *19*, 481–494.

Schneider, S., Nebel, S., Beege, M., & Rey, G. D. (2019). The retrieval-enhancing effects of decorative pictures as memory cues in multimedia learning videos and subsequent performance tests. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000432

Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 49–70). Cambridge, England: Cambridge University Press.

Schwartz, L., & Yovel, G. (2019). Learning faces as concepts rather than percepts improves face recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 1733–1747.

Skulmowski, A., & Rey, G. D. (2018). Realistic details in visualizations require color cues to foster retention. *Computers & Education*, *122*, 23–31.

Skulmowski, A., & Rey, G. D. (2020a). Subjective cognitive load surveys lead to divergent results for interactive learning media. *Human Behavior and Emerging Technologies*, *2*, 149–157.

Skulmowski, A., & Rey, G. D. (2020b). The realism paradox: Realism can act as a form of signaling despite being associated with cognitive load. *Human Behavior and Emerging Technologies*, *2*, 251–258. https://doi.org/10.1002/hbe2.190

Smallman, H. S., & St. John, M. (2005). Naive realism: Misplaced faith in realistic displays. *Ergonomics in Design*, *13*, 6–13.

St. John, M., Cowen, M. B., Smallman, H. S., & Oonk, H. M. (2001). The use of 2D and 3D displays for shape-understanding versus relative-position tasks. *Human Factors*, *43*, 79–98.

Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*, 261–292.

Sweller, J., van Merrienboer, J. J., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–296.

Syring, M., Bohl, T., Kleinknecht, M., Kuntze, S., Rehm, M., & Schneider, J. (2015). Videos oder Texte in der Lehrerbildung? Effekte unterschiedlicher Medien auf die kognitive Belastung und die motivational-emotionalen Prozesse beim Lernen mit Fällen. *Zeitschrift für Erziehungswissenschaft*, *18*, 667–685.

Tillmann, B. N. (2016). *Atlas der Anatomie des Menschen* (3rd ed.). Berlin, Germany: Springer-Verlag.

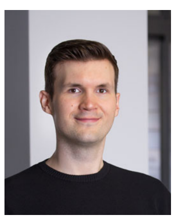
Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology*, *77*, 593–601.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer-Verlag.

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). *The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures*. Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '11). Vancouver, British Columbia (May 7–12, 2011) (pp. 143–146). New York, NY: ACM Press.

AUTHOR BIOGRAPHIES



Dr. Alexander Skulmowski works as a postdoc (Psychology of Learning with Digital Media) at Chemnitz University of Technology in Germany. His research is focused on current aspects of digital education including virtual reality, interactive learning media, and the design of visualizations. His areas of specialization include realism in the design of visualizations and embodied learning.

Prof. Dr. Günter D. Rey is a professor of Psychology of Learning with Digital Media at Chemnitz University of Technology in Germany. His research spans a wide range of psychological effects of digital learning media. This includes the cognitive and motivational basis of learning as well as emotional and social issues in the design of multimedia learning environments.

How to cite this article: Skulmowski A, Rey GD. Realism as a retrieval cue: Evidence for concreteness-specific effects of realistic, schematic, and verbal components of visualizations on learning and testing. *Hum Behav & Emerg Tech*. 2021;3:283–295. https://doi.org/10.1002/hbe2.209