

Is there an optimum of realism in computer-generated instructional visualizations?

Alexander Skulmowski¹

Received: 22 February 2022 / Accepted: 4 April 2022 / Published online: 18 April 2022 © The Author(s) 2022

Abstract

Realistic visualizations are abundantly used in digital education. However, the use of realism is still thought to risk a cognitive overload due to excessive details. Moreover, it is still not precisely known whether there is an optimal level of realism that benefits learners the most. In two experimental studies, different versions of anatomical visualizations were compared regarding their effects on retention performance and the subjective cognitive load experienced during learning. In Experiment 1 (n=73), four visualizations with minor variations in the geometry and shading of the model featured in the visualizations were used. Although neither the level of detail in the geometry nor the realism of the shading resulted in significant differences, a detailed model with simplified shading elicited the highest retention scores descriptively. In Experiment 2 (n=156), a schematic visualization was compared with an "idealized" model featuring only simplified shading and a highly realistic rendering. There most realistic version elicited the highest retention scores, but also the highest cognitive load ratings. Taken together, the results suggest that the optimal level of realism might lie on the more realistic end of the spectrum for learning tasks focused on the memorization of shapes that are assessed using image-based tests.

Keywords Visualizations \cdot 3D \cdot Realism \cdot Retention \cdot Cognitive load

1 Introduction

How realistic and detailed should visualizations be to enable effective learning? This question has been a rather controversial issue for the field of digital education. While some claim that realistic computer-generated visualizations contribute little towards performance or can even turn out to be a burden (Scheiter et al., 2009; Smallman

Alexander Skulmowski alexander.skulmowski@ph-karlsruhe.de

¹ Digital Education, Institute for Informatics and Digital Education, Karlsruhe University of Education, Bismarckstr. 10, 76133 Karlsruhe, Germany

& St. John, 2005), some recent results suggest that at least under certain circumstances, realism can be of value to learners (e.g., Skulmowski, 2022; Skulmowski & Rey, 2021). However, current technology provides us with the necessary tools to create a visualizations with a wide array of options concerning the level of realism. Only a few investigations into the effects of different realism degrees have been published (e.g., Brucker et al., 2014; Huk et al., 2010; Imhof et al., 2011; Skulmowski, 2022; Skulmowski & Rey, 2021). This lack of research poses a problem for instructors wishing to use computer-generated visualizations, as it is hard to estimate the effectiveness of (different) realistic visualizations beforehand and knowledge regarding the optimal level of realism could facilitate instructional design using such material. In order to provide insights into the effects of realism levels in visualizations on learning and cognitive load, two experimental studies were conducted on the basis of a fine-grained model of realism in computer graphics described in the following section.

1.1 Geometry, shading, and realism as the basis of computer-generated imagery

Although different definitions and classifications of pictorial realism have been proposed over the years, many of them do not allow a comprehensive description of the possibilities that modern software used in the creation of three-dimensional (3D) computer-generated imagery offers (Skulmowski et al., 2021). Such software typically requires users to follow a certain series of steps when creating digital visualizations. The main components of content creation have been summarized in the geometry, shading, rendering (GSR) model (Skulmowski et al., 2021) described in the rest of this paragraph. First, models need to be created as virtual counterparts of real objects to populate the 3D scene. These models are created from polygonal structures either by drawing the individual polygons one at a time, from simple primitives that can be refined, or using automated methods such as 3D scanning or photogrammetry (Nebel et al., 2020; Skulmowski et al., 2021). Either way, the end result is a geometrical mesh with more or less polygonal detail. Thus, geometry is the first factor to consider as a contributor to realism (Skulmowski et al., 2021). This geometry can then receive material properties such as color, highlights, and bumps in the shading stage (Skulmowski et al., 2021). Lastly, virtual lights and a camera need to be positioned in the scene that determine how the models will be rendered. In addition, there are several rendering options, such as generating drawing-like contours. Different realism degrees that can be achieved using the individual dimensions of the GSR model are shown in Fig. 1. Naturally, the question arises whether these differences in realism will have an impact on learners. In the following section, previous research on the effects of varying these dimensions to create more or less realistic instructional visualizations is presented.

1.2 Research on the effects of different levels of realism

The most comprehensive comparisons between visualizations featuring different levels of realism have been conducted by Dwyer (e.g., Dwyer, 1968a, b) using



Fig. 1 The geometry, shading, and rendering model as described by Skulmowski et al. (2021). The images in the three rows of the figure exemplify how the different components of computer-generated visualizations can be varied to achieve different looks and effects in the example of visualizing the patella. The geometry dimension determines how high the resolution of the mesh of the model is, and thus how many details can be included. The shading component can be used to apply image textures to the model and also to set other material properties, such as the glossiness. The rendering dimension determines how realistic the visualization will appear overall. Rendering styles can include a drawing-like output and a photorealistic rendering mode. From "Is a Preference for Realism Really Naive After All? A Cognitive Model of Learning with Realistic Visualizations," by A. Skulmowski, S. Nebel, M. Remmele, and G. D. Rey, 2021, Educational Psychology Review, p. 8 (https://doi.org/10.1007/s10648-021-09638-1). © 2021 Skulmowski, Nebel, Remmele, and Rey (licensed under the Creative Commons Attribution License, http://creativecommons.org/licenses/by/4.0)

material such as photographs of organs, of plastic models, drawings featuring a low or high level of detail, and other forms of visualization. Several of these studies were designed to test the notion of a "realism continuum," i.e., a hypothesized link between the degree of realism and learning performance (e.g., Dwyer, 1967, 1969). Dwyer (1969) concludes that this supposed correlation between realism and learning performance cannot be found empirically. However, Dwyer (1969) found some benefits for more realistic visualizations in image-based tests while other tests did not reveal advantages of realism (for discussions of this aspect, see Nebel et al., 2020; Skulmowski et al., 2021).

Surprisingly, only a few studies investigating the effects of controlled variations of realism on learning using computer-generated imagery are currently available. Most studies only contrast two levels of realism, usually a "schematic" version featuring a

contour outline filled with solid colors, often including minimal shading, and a "realistic" version with details, accurate materials, and believable rendering (e.g., Huk et al., 2010; Menendez et al., 2020, 2022; Scheiter et al., 2009; Skulmowski, 2022; Skulmowski & Rey, 2020, 2021). As summarized by Skulmowski et al. (2021), we can broadly distinguish between different learning objectives in instructional realism research: (1) knowledge regarding surfaces, (2) understanding processes, and (3) applying abstract knowledge. From a brief overview of selected studies, Skulmowski et al. (2021) conclude that, as a general rule, realism may be most useful for gaining knowledge of shapes (i.e., in tasks such as anatomy learning), without a clear effect pattern when learning about processes, and often with negative effects on acquiring abstract knowledge. As the focus of this paper lies on the learning of anatomical shapes, the remainder of this section will be dealing with this aspect.

One effect that was found in a study on the usefulness of realistic visualizations in relation to the level of the realism used in retention tests is that learning with a detailed visualization compared to a schematic drawing-like image appears to be particularly beneficial when an equally realistic image is used in the test (Skulmowski & Rey, 2021). The results of that study were interpreted to indicate that learning with a realistic visualization may only pay off if learners need to apply their knowledge to a task involving realistic visuals (Skulmowski & Rey, 2021).

One of the few studies using more than two levels of realism was conducted by King (1986). In that study, participants of various age groups used visualizations consisting of abstract shapes, simplified drawings, or a mixture of photographs and more detailed drawings. The stimuli utilized for the learning tasks either showed a person, an animal, or an object. While retention performance was higher for the stimuli presented in the styles at the ends of the realism spectrum (abstract and realistic) in an immediate test, performance was highest for the images presented in a medium level of realism in a delayed test after one week.

A systematic investigation of the effects of 3D visualizations on credibility was published by Zanola et al. (2009). In their study, a sketch-like rendering of a city, a realistic rendering with plain geometry and simple shading, and a view of that city with more detailed geometry and shading (including small shapes such as windows in houses) was presented. The study revealed that the more realistic visualizations elicited higher credibility ratings. However, learning performance was not a part of that investigation.

In sum, the existing research on the effects of different levels of realism does not seem to provide a clear answer to the question whether there is an optimal degree of realism. While a higher degree of realism does not appear to be detrimental from the reviewed literature, certain conditions need to be met in order for realistic visualizations to unleash their full potential for learning. Furthermore, the existing literature does not allow us to pin-point a specific level of realism as the optimum.

1.3 Cognitive load and the realism paradox

Negative effects of realism have been explained in reference to an assumed or measured cognitive load that the demands of realistic details are thought to entail (e.g., Scheiter et al., 2009; for a discussion, see Skulmowski et al., 2021). Many studies

were conducted in the theoretical framework of cognitive load theory (Sweller et al., 1998, 2019). This theory divides the information that learners are presented with into an intrinsic and an extraneous component (Sweller et al., 2019). For effective learning to take place, the aim is to enable learners to devote as much of their cognitive capacity to the intrinsic cognitive load, the actual content they need to learn (Sweller et al., 1998). Poor instructional design can hinder learning through filling learners' cognitive capacity with *extraneous cognitive load* (Sweller et al., 1998) and the design of instructional visualizations is particularly prone to such problems (for an overview, see Renkl & Scheiter, 2017). In the cognitive model of learning with realistic visualizations, the GSR components of realistic visualizations are thought to contribute towards a *perceptual load*, consisting of demands on learners to work with very detailed visualizations containing many visual elements (Skulmowski et al., 2021). Does the solution to the aforementioned issues arising from realism lie in simply removing details that are deemed "unnecessary"? The answer to this question is not as straightforward as one would hope. Recent studies revealed that the relationship between realism and cognitive load cannot be adequately described as a simple negative correlation. For instance, it was found that when schematic and realistic components are combined in a display, some realistic visualizations can raise the overall extraneous load, but the retention performance for the realistic parts can turn out to be higher as well (Skulmowski & Rey, 2020; see also Koc & Topu, 2022, for a related finding on high cognitive load during learning with 3D visualizations). This counterintuitive finding has been named the *realism paradox* (Skulmowski & Rey, 2020) and has raised concerns over the explanatory power of (extraneous) cognitive load in the context of digital learning (Skulmowski & Xu, 2022). Another recent study found that in a virtual reality learning task, extraneous cognitive load was positively correlated with learning results (Tugtekin & Odabasi, 2022), offering additional evidence for the claim that in virtual and realistic environments, high perceptual demands may be perceived as demanding, but that these environments can still be beneficial for learning. It has been argued that the effects of realism need to be analyzed with the desired impact on cognitive processing and the mode of assessment in mind (Skulmowski & Xu, 2022). In some cases, letting learners invest more effort may actually be a better preparation for a later test than oversimplifying the learning task (Skulmowski, 2021; Skulmowski & Xu, 2022).

As an interim summary, it is generally assumed that unnecessary cognitive load that may be introduced by irrelevant details in realistic visualizations can be an obstacle for learners. However, some studies imply that lowering cognitive load may not be the ideal strategy for the design of realistic visualizations, as it may be the case that the higher cognitive demands of realistic visualizations actually have a positive effect on retention performance.

1.4 Idealization as the best of both worlds?

One aspect of realism that is in need of a closer examination is the perceptual load associated with it. In a number of papers, realism has been analyzed in terms of *geon theory* (Biederman, 1985, 1987; for discussions in the context of realistic

visualizations, see Nebel et al., 2020; Skulmowski et al., 2021; Skulmowski & Rey, 2018). Geon theory holds that one of the steps in human visual perception is to mentally simplify the visual information in the field of view by treating objects not as the highly complex structures they may be, but rather as (a combination of) geometric primitives, such as boxes and cylinders (Biederman, 1985). Thus, a tree can be visually processed as a brown cylinder with a (slightly deformed) green sphere on top when seen from afar. In this stage, details such as the ridges in the tree bark or discolorations of individual leafs are not (yet) focused on. Several educational fields attempt to use this aspect of visual perception for their advantage. For instance, several books on learning to draw, in particular those concerned with (artistic) anatomy, use geometric primitives as proxies for the complex shapes of the human body. A particularly noteworthy example for this strategy are the works of George Bridgman (e.g., Bridgman, 1973), in whose instructional books parts of the body such as the arms, legs, and the neck are presented as simple geometric shapes. These basic shapes can be rather easily drawn in correct perspective and then serve as the base for further refinement, for example by subdividing the cylindrical shapes of the arms into smaller primitives approximating the forms of their muscles. Based on the popularity of this approach in art instruction, segmenting the shapes of objects into more *idealized* and prototypical shapes may be a promising approach for the design of computer-generated instructional visualizations.

The use of idealized 3D forms as a potential optimum between too simplified and too detailed instructional visualizations has not been thoroughly studied extensively yet using computer-generated visualizations. Somewhat related comparisons have been undertaken by Dwyer (1968a) when drawings and photographs of the heart were compared with photographs of a plastic model, featuring an idealized, smooth shape without irregularities that may confuse novice learners. The results of that study do not indicate a clear advantage of the different visualization types over a mere oral presentation across all tests. However, in an identification test, a detailed shaded drawing elicited higher scores than the photographs of the model. By contrast, Dwyer (1969) found in a related study that learning using photographs of a plastic model was more effective in an identification test than learning with drawings or photographs of real anatomical structures. As these results do not offer clear guidance on the issue of idealization in visualizations and since the studies could not make use of computer-generated imagery yet, there is a gap concerning this aspect in the literature.

In sum, an idealized mode of visualization that uses simplified 3D models while avoiding the use of too many irregular details on the one side and a too abstract, drawing-like presentation could prove to be the optimal way of designing 3D instructional visualizations and should be investigated empirically.

1.5 The present studies

Based on the described theoretical models and empirical results, two experiments were conducted. The first study was designed to assess whether variations in the geometry and shading of a model lead to differences in retention performance and

subjective (extraneous) cognitive load. The second study compares an idealized rendering and a more detailed realistic version with a schematic visualization.

2 Experiment 1

Previous studies suggest that learners can be overburdened by irrelevant details (e.g., Scheiter et al., 2009), but since other studies demonstrated that an oversimplification may not be a good alternative (e.g., Skulmowski, 2021; Skulmowski & Rey, 2018), an experiment was conducted to assess whether a medium level of realism can prove effective for learning an anatomical structure. In this study, the two components geometry and shading of the GSR model are experimentally manipulated. The visualization used in the study featured combinations of an idealized or a detailed geometry, combined with either simplified or realistic shading. An interaction effect was hypothesized in which the two "medium" combinations (idealized geometry and realistic shading; detailed geometry and simplified shading) should lead to better retention scores than the two extreme conditions featuring a strongly simplified or a highly detailed rendering. Furthermore, the study assessed whether an idealized geometry and a simplified shading can enhance learning. Concerning (extraneous) cognitive load, based on the realism paradox presented above (Skulmowski & Rey, 2020), it was hypothesized that a higher level of realism in both factors leads to higher cognitive load scores for each factor.

2.1 Methods

Participants and design Based on an assumed effect size of $\eta_p^2 = 0.10$ (in line with previous research finding even larger effects, e.g., Skulmowski & Rey, 2021) and a power of 0.80, a target sample size of 73 participants was computed using G*Power (Version 3.1.9.2; Faul et al., 2009). The 2×2 between-subjects design consists of the factors shape (idealized vs. detailed) and shading (simplified vs. realistic).

In order to be included in the final analysis, participants needed to fulfill certain criteria. As described below in detail, participants were asked regarding the participation requirements at the beginning of the web-based study. They were eligible to enter the regular data collection if they passed these requirements. At the end of the study, they were asked two quality control questions that needed to be answered appropriately (see below for an explanation) in order for their participation to be counted as completed and they also needed to reach the last page of the study. A total of 75 eligible participants completed the study before data collection was stopped. Only the data of the originally planned 73 participants was considered in the analyses, although the pattern does not change with the additional two participants included.

In the study, 62 female and 11 male students within the age range of 18 and 30 years participated for partial course credit in a lecture on Digital Education held at a university of education in Germany. This lecture was open to students in

the teacher training courses at the undergraduate and graduate level. Randomization to one of the four experimental groups was achieved through block randomization and resulted in nearly equal group sizes $(n_{\text{Group 1}} \text{ [idealized shape+simplified shading]} = 17$, $n_{\text{Group 2}}$ [idealized shape+realistic shading] = 18, $n_{\text{Group 3}}$ [detailed shape+simplified shading] = 19, $n_{\text{Group 4}}$ [detailed shape+realistic shading] = 19).

Materials The study utilized four different visualizations of the lung of which one was presented to each participant during the learning stage (see Fig. 2). Blender (Version 3.0.0) was used for the creation of all renderings in both experiments. The model featured a smooth and idealized geometry without smaller details in two of the conditions (see Fig. 2a, b) that was contrasted with a bumpy and irregular, more detailed geometry (see Fig. 2c, d). This model was either presented with a solid and simplified shading with a plastic-like surface (see Fig. 2a, c) or a realistic shading involving several material layers (such as a more irregularly colored texture, glossiness, and bump mapping) that can be seen in Fig. 2b and d. Based on the idea that a schematic visualization was the option that would lead to the least biased result, a drawing-like visualization with minimal shading and a contour outline was produced for the learning test (see Fig. 2e). The idea behind this choice was that this style was sufficiently different from all four versions shown during the learning phase that it would not give a specific advantage to one of the four groups based solely on the realism level of the visualization used for testing (see also Skulmowski



Fig. 2 The figure displays the four different versions of the lung rendering used for the learning phase (**a-d**) and the learning tests (**e**) that were created in Blender 3.0.0 based on various references (Gray, 1918, pp. 1095–1096; Rauber, 1892, p. 628; Cole, 2015, and other reference material). In all versions, it was indicated below the lung renderings whether the left or right lung was displayed with a label in all uppercase letters. The labels consist of a mixture of Latin terms (such as "Bronchus") and German names (e.g., "Lungenarterie" [pulmonary artery])

& Rey, 2018). For each correct answer, participants were awarded one point for their total score, with a maximum of 18. The retention test had a reliability of McDonald's $\omega = 0.77$.

During the study, the extraneous cognitive load survey items developed by Klepsch et al. (2017) were used in a modified form (as in Skulmowski & Rey, 2020) that asked participants regarding their learning experience using the visualizations rather than the entire learning task as in the original items. The survey had a reliability of $\omega = 0.87$.

Procedure The procedure of both experiments in this paper was similar to the one described by Skulmowski and Rey (2020). After providing informed consent, a page was presented on which participants were asked to respond to questions regarding their age range, their prior knowledge, their native language, their currently used device, and whether they had participated in the study before. Only if they were native speakers of German, had no or little knowledge of lung anatomy, used a PC or laptop rather than a device with a small screen to participate, and had not already participated, they could enter the regular study. They then received the instruction that they would be presented with visualizations of the left and right lung and had 60 s to memorize the names, shapes, and locations of the parts. Then, they were directed to a filler task in which they were asked to sort the 16 German federal states according to their number of day schools within the time limit of 60 s. The following retention test consisted of a page in which the left and right lung were displayed, labeled with letters (see Fig. 2e). Below the two images, participants were asked to select the correct part name for each lettered component from drop-down menus. As there were labeled parts in the test images that had not been presented with a label in the learning phase, participants were asked to select the option "NOT LEARNED" for these components. There was no time limit for this task and participants were reminded not to use additional resources for answering the tests. As the test was presented relatively soon after the learning phase, the test is mainly focused on shortterm memory. Following the test page, they were asked to indicate their gender and course of study along with two quality control questions. These questions asked participants whether they were strongly distracted during the learning task and whether they experienced a major technical difficulty (as in Skulmowski & Rey, 2020). Both of these questions needed to be answered with a negative response in order to proceed to the final pages and in order for datasets to count as complete. Both studies in this article were conducted using SoSci Survey (Leiner, 2021).

2.2 Results

In the analyses of both studies, the assumptions of analysis of variance (ANOVA) procedures were tested using the Shapiro–Wilk test (calculated on the model residuals) and Levene's test. If one or more of these assumptions were violated as indicated by a significant deviation, a nonparametric ANOVA was conducted using aligned rank transformation (Fawcett & Salter, 1984) instead of a parametric ANOVA.

Extraneous load As a Shapiro–Wilk test assessing the normality of the residuals of the parametric ANOVA of the extraneous cognitive load data indicated a violation of this assumption, a nonparametric ANOVA was conducted. No significant main effect or interaction was found, all $ps \ge 0.269$ (see Fig. 3a for the untransformed data).

Retention An ANOVA computed using the retention scores did not result in significant main effects or an interaction, all $ps \ge 0.390$ (see Fig. 3b).

3 Experiment 2

The first study did not reveal a clear advantage for a medium level of realism that had been hypothesized. However, a few tendencies could be observed on the descriptive level of the data. In line with previous research (Skulmowski & Rey, 2020), a higher level of realism in the shape and shading dimensions resulted in slight increases in extraneous cognitive load. Concerning retention performance, the more detailed shape combined with simplified shading elicited the highest average retention score among all four combinations. This descriptive result could be taken as an indication that certain combinations of the GSR dimensions in the middle of the realism may have become clearer if a realistic visualization would have been used (Skulmowski & Rey, 2021). In order to achieve more definitive results, a second study was conducted using visualizations with even stronger differences concerning their level of realism. Furthermore, the learning test in this study utilized a realistic visualization rather than a schematic drawing as in the first experiment.



Fig. 3 The descriptive data of Experiment 1 shown as boxplots with violinplot elements. The group means are indicated by white dots. Figure 3a shows the extraneous cognitive load data on a scale from 1 to 7, Fig. 3b contains the retention scores with a minimum of 0 and a maximum of 18 points awarded for each correct response

It was hypothesized that an idealized realistic visualization and a detailed realistic rendering result in higher extraneous cognitive load ratings than a schematic visualization. Concerning retention performance, the idealized and realistic version should lead to higher scores than a schematic version.

3.1 Methods

Participants and design As several changes were made compared to the first study to ensure a higher probability for clear differences between the groups, a medium effect size of $\eta_p^2 = 0.06$ (power = 0.80) was chosen and resulted in a sample size of 156 participants. The block-randomized experiment used a design with three between-subjects groups that either learned with a schematic visualization (n = 52), an idealized realistic rendering (n = 51), or a detailed realistic rendering (n = 53). Eligibility criteria and the conditions for the completeness of datasets were adapted from Experiment 1. The participants were 136 female and 20 male students within the age range of 18 and 30 years taking part in the same lecture in the same institution as described for Experiment 1.

Materials The study compared three different versions of a diagram of kidney anatomy regarding their impact on learning (see Fig. 4). One group was presented with a schematic version with a simplified geometry and a drawing-like rendering style (see Fig. 4a), the second group learned using an idealized version with a slightly more organically-formed geometry than the one used in the schematic version and simplified shading (see Fig. 4b), and a more detailed geometry with realistic shading in the third group (see Fig. 4c).

Two retention tests using the detailed and realistically shaded version were used. The first retention test (see Fig. 4d) included all 12 items that were learned in the learning phase, but most of the labels were moved to a different position on the image to prevent participants from mentally retrieving the labels simply by their order and position, rather than their visual properties. The second retention (see Fig. 4e) test included the subset of the eight items that were moved to a different location in the image of the first retention test. This second test was included to assess whether there is a difference between a more difficult test that requires learners to understand the shapes and structures and an easier one that could in principle be solved by relying on the order of the items. The two retention tests had a reliability of $\omega = 0.65$ (computed using principal factor analysis) and $\omega = 0.78$, respectively. As in Experiment 1, the learning tests were presented shortly after the learning phase, thus implying that short-term memory was tested.

Experiment 2 used the items for extraneous cognitive load from Klepsch et al. (2017) in the modified version of Skulmowski and Rey (2020). The reliability of the extraneous cognitive load survey items was $\omega = 0.87$.

Procedure The general procedure was identical to the one used in Experiment 1. However, the prior knowledge question asked whether participants had little or no



Fig.4 The figure contains the three versions of the kidney utilized in the learning task (a-c) and the learning tests (d-e) created using references (Gray, 1918, p. 1221; Cole, 2015, Stock, 2018), and additional reference material). As in Experiment 1, the labels are a mixture of German (e.g., "Harnleiter" [urethra]) and Latin (e.g., "Papilla") labels

knowledge concerning kidney anatomy and the learning phase lasted 50 s with an adapted instruction stating that kidney anatomy was to be learned in the task.

3.2 Results

As between-subjects designs have lower statistical power than 2×2 designs, it was decided to bypass omnibus tests and to use directional Dunnet's tests to compare the group using the schematic version (as the control group) to the two groups who learned with more realistic renderings.

Extraneous load As the extraneous cognitive load data violated both the assumptions of normality of residuals and variance homogeneity, Dunn's post-hoc test was used as a nonparametric alternative. As expected, there was a significant difference in extraneous cognitive load between the schematic and the detailed rendering, p=0.024 (without alpha error correction), with the latter version receiving higher cognitive load ratings (see Fig. 5a). This significant result was confirmed with a Games-Howell post-hoc test using Tukey correction, t(96.98)=-2.41, p=0.046. Comparisons between the schematic and idealized version (p=0.067) and between the idealized and detailed version (p=0.321) did not result in significant differences using Dunn's test (without alpha error correction).

Retention For the first retention test, the assumption of homoscedasticity was violated and thus, a one-sided directional Tamhane-Dunnett's test (single-step *p*-value adjustment method) was used to compare the schematic version with the two realistic renderings. The directional assumption of the test was that the realistic renderings would lead to



Fig. 5 The descriptive data of Experiment 2 shown as boxplots with violinplot elements. The group means are indicated by white dots. Figure 5a displays the extraneous cognitive load data on a scale from 1 to 7, Fig. 5b presents the retention scores of the first retention test with a minimum of 0 and a maximum of 12 points

higher scores than the schematic version. The detailed rendering elicited significantly higher retention scores than the schematic version, p=0.035, with no significant difference between the schematic and idealized version, p=0.699 (see Fig. 5b).

A similar pattern was found in the second retention test, but without significant differences between the schematic (M=5.38, SD=2.3), the idealized (M=5.47, SD=2.38), and the detailed (M=6.04, SD=1.98) versions (all $ps \ge 0.080$) when tested using Dunn's post-hoc test (without alpha error correction).

4 General discussion

Two studies were conducted to assess whether differences in the level of realism in the geometry, shading, and rendering of computer-generated visualizations can lead to an instructional advantage. In particular, an optimal level of these factors was sought. The first study did not reveal a significant effect of the geometry and shading components of realism. However, the descriptive results indicate that a detailed shape with simplified shading led to the highest retention scores. Extraneous cognitive load ratings increased on the descriptive level with each additional step of realism, with the lowest cognitive load ratings for the idealized model with simplified shading and the highest average rating for the detailed model combined with the realistic shading.

In the second study, a strongly simplified, drawing-like visualization was compared to an idealized model with simplified shading and to a detailed model with realistic shading. Only the most realistic model offered participants a significant advantage over the schematic visualization. The visualization featuring a "medium" level of realism only raised retention performance slightly on the descriptive level. However, the most realistic model also received significantly higher extraneous cognitive load ratings than the schematic version. Taken together, the results of the two studies suggest that there may not be a generalizable optimal level of realism that leads to the highest retention performance. In particular, an idealization of the shape that avoids fine-grained details did not prove to be the most successful in both studies. However, on the descriptive level, Experiment 1 revealed an optimum for a detailed model with simplified shading. However, the highest level of realism featuring details in the geometry and a realistic material resulted in the lowest median retention score. Thus, this pattern could be interpreted as an indication of a potential cognitive overload that hampered performance. Further studies with even stronger differences between the visualizations used are needed to come to a definite conclusion regarding this aspect. In Experiment 2, the highest level of realism was found to result in the best retention performance despite causing the highest level of extraneous load. This result is a confirmation of the realism paradox and shows that realism is a form of digital learning that does not seem to consistently follow the rule that a minimization of extraneous cognitive load leads to higher learning outcomes (Skulmowski & Xu, 2022).

One of the causes behind the more substantial positive effect of realism found in Experiment 2 may be that a realistic visualization was used in the retention test. As previous research indicated, the matching of a high level of realism in the learning and testing stage can lead to the highest retention performance (Skulmowski & Rey, 2021). While Experiment 1 used a schematic visualization in the retention test and did not result in a clear difference between various types of realistic visualizations, Experiment 2 included a realistic retention test and revealed a positive effect of the most realistic instructional visualization.

4.1 Implications for the use of computer-generated instructional visualizations

In sum, the results of the two studies show that dismissing realistic visualizations due to their potential demands appears to be largely unfounded. Experiment 2 confirms previous results claiming that a more realistic style can be of value to learners in tasks focused on learning visual shapes compared to schematic visualizations. However, both experiments did not reveal a specific level of realism as the optimum. While the highest retention scores were achieved after learning with a visualization featuring a detailed geometry combined with simplified shading in Experiment 1, participants learned the most from a detailed model with realistic shading in Experiment 1 resulted in a performance drop compared to the second-most realistic rendering. This visualization might be an example for the cognitive overload caused by too many fine details and irregularities that make it harder to understand the basic shapes (in line with geon theory) which may also explain previous negative results from the field of realism research (e.g., Scheiter et al., 2009).

The results lead us to three conclusions: (1) As the strongest difference was found between a schematic and a highly realistic version (confirming previous results), a higher level of realism may be the safest choice for visual learning tasks with an emphasis on shapes. (2) As soon as a realistic rendering is used, the exact level of realism might not be a crucial factor to consider. (3) However, when choosing a high

level of realism, the shapes still must be recognizable enough for learners given their respective level of expertise. There may be a point at the very end of the realism spectrum at which visualizations indeed become too detailed to visually "decrypt." Thus, the optimal level of realism for learning shapes is likely to be a realistic rendering mode with a moderate amount of details that are not so excessive as to hinder learners from capturing the basic components portrayed. A simplified diagram may offer too few concrete building blocks for learners' mental representations, while visualizations should not overwhelm learners with too many minor details and irreg-

visualizations should not overwhelm learners with too many minor details and irregularities in shape and shading that prevent learners from distinguishing the relevant from the irrelevant information.

4.2 Limitations and outlook

The main contribution of the two experiments presented in this paper lies in the highly controlled creation of the instructional material that allows comparisons without confounding factors. However, there are many ways to generate computer-generated imagery and 3D models, such as scan data or photogrammetry (Nebel et al., 2020). The output of such methods can have vastly different styles and resolutions that should be considered in future empirical work. Considering the results of the studies, very high resolution models created using such methods may indeed have the potential to be overwhelming, and suitable procedures for the simplification of such models should be evaluated.

One limitation of both studies is that they focused on retention performance, but not on transfer. Furthermore, the learning tasks center around the learning of shapes. It needs to be verified if the result pattern found in the studies can be replicated when the tasks revolve around learning processes or applying abstract knowledge.

In the interpretation of the results, it needs to be considered that the first study used a learning test featuring a very low level of realism, while the second study included the most realistic version that was presented to one of the three groups in the learning phase in the retention tests. Thus, it would be interesting to see whether the four visualizations presented in Experiment 1 can lead to differences in learning if a more realistic learning test is used. Likewise, future studies could be conducted to replicate the positive effect of realism found in Experiment 2 with a less realistic learning test. However, as medical education usually aims at preparing learners to work with real anatomical structures rather than with simplified diagrams, the results of Experiment 2 have a better chance of being transferable to real-world settings using the tests used in the present study.

It needs to be noted that the ratio of female and male participants was unbalanced in the studies. The vast majority of participants in both experiments was female. Lastly, one of the retention tests in Experiment 2 resulted in a reliability of $\omega = 0.65$, which may be a result of differences between the label positions in the learning and testing phases. However, the construction of the test followed the method of several studies cited in this paper that reached high reliability scores of over 0.8 (e.g., Skulmowski, 2021, 2022). Replications should acknowledge this aspect and improve the test by further eliminating ambiguities.

4.3 Conclusion

The two studies presented in this paper underline the complexity and difficulty of using realistic instructional visualizations effectively. The results clearly demonstrate that an approach that is solely fixated on "simplifying" realistic visualizations may not be the most promising strategy. In both studies, the visualizations with the lowest realism levels did not result in the highest retention scores. As shown in Experiment 1, more nuanced combinations of different realism components did not lead to significant differences regarding retention. In addition, Experiment 2 demonstrated that there is a significant advantage for a highly realistic rendering compared to a schematic visualization, while an idealized rendering did not lead to such an increase in retention. Judging from the studies, if there is an optimum for realism in the learning of shapes, it appears to lie on the more realistic end of the spectrum. However, further research is needed to formulate more specific guidelines.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The data of this study are available from the author upon request.

Code availability Not applicable.

Declarations

Ethics approval The research described in this article does not require an ethics review in Germany by law. All relevant national and institutional guidelines have been followed.

Consent to participate All participants provided their informed consent to participate in the studies.

Consent for publication Not applicable.

Conflicts of interest/Competing interests The author is an editorial board member for the journal Educational Psychology Review and served on the editorial board of Human Behavior and Emerging Technologies.

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

Biederman, I. (1985). Human image understanding: Recent research and a theory. *Computer Vision, Graphics, and Image Processing, 32, 29–73.*

- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. Psychological Review, 94, 115–147.
- Bridgman, G. B. (1973). Constructive anatomy. Dover Publications.
- 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.
- Cole, T. C. H. (2015). Wörterbuch der Biologie/dictionary of biology. Berlin, Germany: Springer Spektrum.
- Dwyer, F. M., Jr. (1967). The relative effectiveness of varied visual illustrations in complementing programed instruction. *The Journal of Experimental Education*, 36(2), 34–42.
- Dwyer, F. M., Jr. (1968a). An experiment in visual learning at the eleventh-grade level. The Journal of Experimental Education, 37(2), 1–6.
- Dwyer, F. M. (1968b). The effectiveness of selected visual illustrations in teaching science concepts to college freshmen. *The Journal of Educational Research*, 61(8), 343–347.
- 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 Technol*ogy, 6(3), 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(4), 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(2), 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
- Huk, T., Steinke, M., & Floto, C. (2010). The educational value of visual cues and 3D-representational format in a computer animation under restricted and realistic conditions. *Instructional Science*, 38, 455–469.
- Imhof, B., Scheiter, K., & Gerjets, P. (2011). Learning about locomotion patterns from visualizations: Effects of presentation format and realism. *Computers & Education*, 57, 1961–1970.
- King, A. (1986). The role of realism in memory for computergraphics by children and adults. *The Journal of Experimental Education*, 55(1), 43–48.
- Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, 8, 1997. https://doi. org/10.3389/fpsyg.2017.01997
- Koc, T., & Topu, F. B. (2022). Using three-dimensional geospatial technology in primary school: Students' achievements, spatial thinking skills, cognitive load levels, experiences and teachers' opinions. *Education and Information Technologies*. https://doi.org/10.1007/s10639-021-10810-x
- Leiner, D. J. (2021). SoSci Survey [Computer software]. Available at https://www.soscisurvey.de
- 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*, 34(5), 1101–1117.
- Menendez, D., Rosengren, K. S., & Alibali, M. W. (2022). Detailed bugs or bugging details? The influence of perceptual richness across elementary school years. *Journal of Experimental Child Psychology*, 213, 105269.
- Nebel, S., Beege, M., Schneider, S., & Rey, G. D. (2020). A Review of Photogrammetry and Photorealistic 3D Models in Education From a Psychological Perspective. *Frontiers in Education*, 5, 144.
- Rauber, A. (1892). Lehrbuch der Anatomie des Menschen: Allgemeiner Teil, Knochen, Bänder und Muskeln. Erster Band, Allgemeiner Teil, Knochen, Bänder, Muskeln, Eingeweidelehre. Leipzig, Germany: Eduard Besold. Retrieved from https://wellcomecollection.org/works/sp2xb4ma
- Renkl, A., & Scheiter, K. (2017). Studying visual displays: how to instructionally support learning. Educational Psychology Review, 29(3), 599–621. https://doi.org/10.1007/s10648-015-9340-4
- Scheiter, K., Gerjets, P., Huk, T., Imhof, B., & Kammerer, Y. (2009). The effects of realism in learning with dynamic visualizations. *Learning and Instruction*, 19(6), 481–494.
- 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. (2020). 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.

- Skulmowski, A., & Rey, G. D. (2021). Realism as a retrieval cue: Evidence for concreteness-specific effects of realistic, schematic, and verbal components of visualizations on learning and testing. *Human Behavior and Emerging Technologies*, 3, 283–295.
- Skulmowski, A., & Xu, K. M. (2022). Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load. *Educational Psychology Review*, 34, 171–196.
- Skulmowski, A., Nebel, S., Remmele, M., & Rey, G. D. (2021). Is a preference for realism really naive after all? A cognitive model of learning with realistic visualizations. *Educational Psychology Review*. https://doi.org/10.1007/s10648-021-09638-1
- Skulmowski, A. (2021). When color coding backfires: A guidance reversal effect when learning with realistic visualizations. *Education and Information Technologies*, 1–16. https://doi.org/10.1007/ s10639-021-10796-6
- Skulmowski, A. (2022). Realistic visualizations can aid transfer performance: Do distinctive shapes and descriptive labels contribute towards learning? *Journal of Computer Assisted Learning*, 1–11. https://doi.org/10.1111/jcal.12640
- Smallman, H. S., & St. John, M. (2005). Naive realism: Misplaced faith in realistic displays. *Ergonomics in Design*, 13, 6–13.
- Stock, K. (2018). Nephrologische Ultraschalldiagnostik. Der Nephrologe, 13, 236–243.
- Sweller, J., van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10, 251–296.
- Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31, 261–292.
- Tugtekin, U., & Odabasi, H. F. (2022). Do interactive learning environments have an effect on learning outcomes, cognitive load and metacognitive judgments?. *Education and Information Technologies*, 1–40. https://doi.org/10.1007/s10639-022-10912-0
- Zanola, S., Fabrikant, S. I., & Çöltekin, A. (2009). The effect of realism on the confidence in spatial data quality in stereoscopic 3D displays. In *Proceedings of the 24th International Cartography Conference (ICC 2009)*, Santiago, Chile (pp. 15–21).

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