The impact of annotation on concrete and abstract visual representations in science education: testing the expertise reversal effect

Robert Zheng[1]* , Holly Cordner[1] and Jeffrey Spears[2]

*Correspondence:
robert.zheng@utah.edu

[1] Department of Educational Psychology, University of Utah, 1721 Campus Center Dr. SAEC 3234, Salt Lake City, UT 84112, USA
Full list of author information is available at the end of the article

Abstract

This study investigates the effects of annotation on abstract and concrete visual representations in science education. Two studies were conducted: Study 1 investigated the interaction between annotation and visual representations. The results of this study demonstrated that in science learning, annotation with abstract visual representations was superior to annotation with concrete visual representations. Study 2 tested the expertise reversal effect in a three-factor design where the interaction among annotation, visual representation, and prior knowledge was measured. The findings showed that high-prior-knowledge learners performed better in the annotation-abstract visual condition than in the annotation-concrete visual condition where low-prior-knowledge learners showed the opposite outcome—confirming that the expertise reversal effect is at play in how science learners utilize visual information. The study has clarified the roles of prior knowledge, visual representation, and instructional strategy on learner cognitive processing in science education. This knowledge should prove useful for educators as they engage in the design and development of computer-based science learning.

Keywords: Annotation, Cognitive load, Expertise reversal effect, Problem solving, Science education, Visual representations

Introduction

Research has demonstrated that visual resources, when optimally designed, can significantly improve learners' understanding (Moreno & Mayer, 1999; Sweller & Chandler, 1994). Evidence from empirical research further suggests that different forms of visual representation may differently influence learners' cognitive information processes (Mason et al., 2013). Studies also show that concrete visual representations can influence the way learners utilize their prior knowledge when processing science content (Moreno et al., 2011). Abstract visual representations, on the contrary, are likely to facilitate deep-level processing such as knowledge transfer in learning (Easterday et al., 2009; Kaminski et al., 2008). In addition, visual cues like annotation have been found to significantly affect learners' information processing in visual learning by redirecting their attention, reducing the search path, and alleviating the cognitive load (de Koning et al.,

2010; Madsen et al., 2013). Despite these advances in the knowledge of the relationship between visual representations and visual cues, little research has been conducted to understand how visual cues like annotation may influence learners' information processing in abstract and concrete visual representations. Thus, the first goal of the current study was to examine the relationship between annotation and abstract/concrete visual representations in learning.

Studies have shown that the way visual representation is represented can have different effects on learners, depending on their level of expertise in knowledge domain (Gegenfurtner et al., 2011). In their study on diagrams and expertise, Kalyuga et al. (2012) found that novices performed well in diagrams with text, whereas experienced learners worked better with diagrams only, resulting in an expertise reversal effect. Given this empirical evidence of the functional role of expertise in visual learning, it is evident that research on visual representation and annotation should also consider the mediating effect of expertise. Therefore, the second goal of this study was to investigate the impact of expertise on visual learning by studying a 3-way interaction between visual representations (abstract vs. concrete), annotation (annotation vs. non-annotation), and prior knowledge (high vs. low) in science learning.

To understand the complexity of visual representations and annotation and their impact on learners' cognitive processes, the current study is informed by several theories and framework that include cognitive load theory, research in abstract and concrete visual representations, role of annotation, and expertise reversal effect in visual learning.

Cognitive load theory

Cognitive load theory (CLT), which was conceived in the 1980s by Sweller and colleagues, has become one of the most important theories in learning and psychology (Greenberg et al., 2021; Kalyuga & Plass, 2018; Plass & Kalyuga, 2019). According to CLT, three types of cognitive load can significantly influence learners' cognitive processes in learning (Sweller et al., 1998). They are intrinsic, extraneous, and germane cognitive load. The intrinsic load refers to the difficulty of the learning material as defined by its element interactivity, that is, "the level of interconnectedness between the information elements that need to be processed … to make sense of the learning tasks or materials" (Plass & Kalyuga, 2019, p. 342). For example, memorizing a list of random words requires a lot of effort but does not necessarily pose high intrinsic load. Reading a dense paragraph of text, on the other hand, may induce high intrinsic load, since the learner needs to figure out the semantic structure among the words, the grammatical relationship between parts of speech, and so forth. All of this indicates a high level of element interaction. Thus, the higher the element interactivity is, the more difficult learning the material becomes, and the higher the intrinsic cognitive load will be (Sweller & Chandler, 1994). The extraneous cognitive load is caused by inappropriate design in instruction, such as redundancy or split attention in learning. For instance, teachers may unwittingly increase learners' extraneous load by presenting materials that "require students to mentally integrate mutually referring, disparate sources of information" (Sweller & Chandler, 1991, p. 353). Therefore, the extraneous cognitive load is considered a hindrance to learning and should be minimized. Finally, the germane cognitive

load is induced by learners' efforts to process and comprehend material in order to construct new knowledge. This type of cognitive load is relevant to learning and reflects learners' cognitive engagement.

There has recently been a growing interest in the affective aspects in cognitive load research (Moreno & Mayer, 2007; Plass & Kalyuga, 2019; Plass & Kaplan, 2016). In an early study, Moreno (2010) stressed the importance of motivation in determining the cognitive resources allocated to learning tasks (the resources that are associated with germane load). Plass and Kalyuga (2019) argued that cognitive load, and germane load in particular, may be related to, or involve, affective processes. They contended that activating motivation would lead to effortful cognitive engagement during the learning process. In a study on cognitive load, motivation and prior knowledge in mathematics problem solving, Gupta and Zheng (2020) found germane load was significantly positively correlated with the learner's interest ($r = 0.221$, $p < 0.01$), confirming Plass and Kalyuga's argument about the relationship between motivation and germane cognitive load in learning.

Abstract and concrete visual representations

Visual resources play an important role in the formation and development of mental representations (Crisp & Sweiry, 2006). According to Crisp and Sweiry, there are a number of possible reasons for the apparent superiority of visual over non-visual representations in mental representation formation (Also see Greenberg et al., 2021, Greenberg & Zheng, 2022). First, processing visual material may require less cognitive effort. The learner can grasp the general meaning of an image in as little as milliseconds. Second, visual and non-visual materials may be processed in different cognitive systems. The dual-coding theory (Paivio, 1986) explains why memory for images may be better than memory for non-image information—it is a result of them being encoded both as images and verbal labels, while text is only encoded verbally.

Due to the above reasons, visual representations have been widely applied to learning across domains, especially in STEM education (Frey et al., 2016; Fuchsova & Korenova, 2019; Rau, 2017; Soong et al., 2020). Fuchsova and Korenova (2019) investigated the effects of visual representation on elementary school teachers' human biology training. They noticed that by using augmented reality, the learners demonstrated deeper understanding, greater motivation, and more creativity in learning.

In addition, visual representations can be classified into abstract and concrete visual representations. Abstract visual representations refer to visuals that use conventional symbols (e.g., lines, nodes, boxes, etc.) to represent the relevant elements of a problem where concrete visual representations are those that depict the real-life objects corresponding to a problem (Moreno et al., 2011). Figures 1 and 2 are examples of abstract and concrete visual representations, respectively. In the abstract visual representations, lines and symbols are used to represent the electric circuitry. In contrast, in the concrete visual representations, actual images (e.g., bulb, battery, wire, etc.) are used to represent the electric circuitry.

Mason et al. (2013) differentiated the cognitive functioning between abstract and concrete visual representations by showing that concrete visuals "may make information more available in long-term memory," whereas abstract visuals "may lead to more efficient processing of learning material" (p. 377). In a separate study, Moreno et al. (2011)
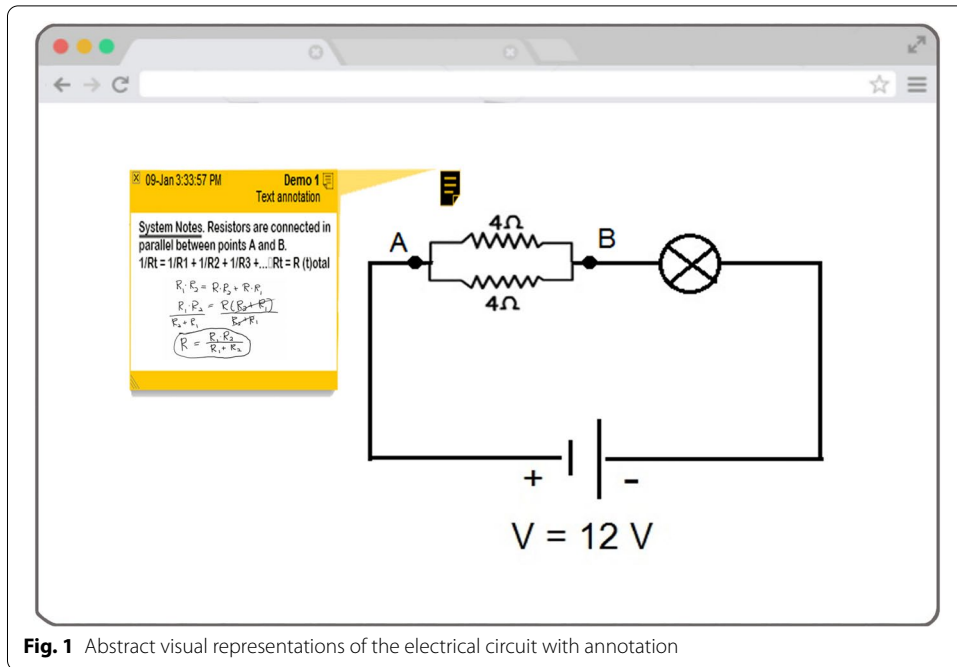
Fig. 1 Abstract visual representations of the electrical circuit with annotation
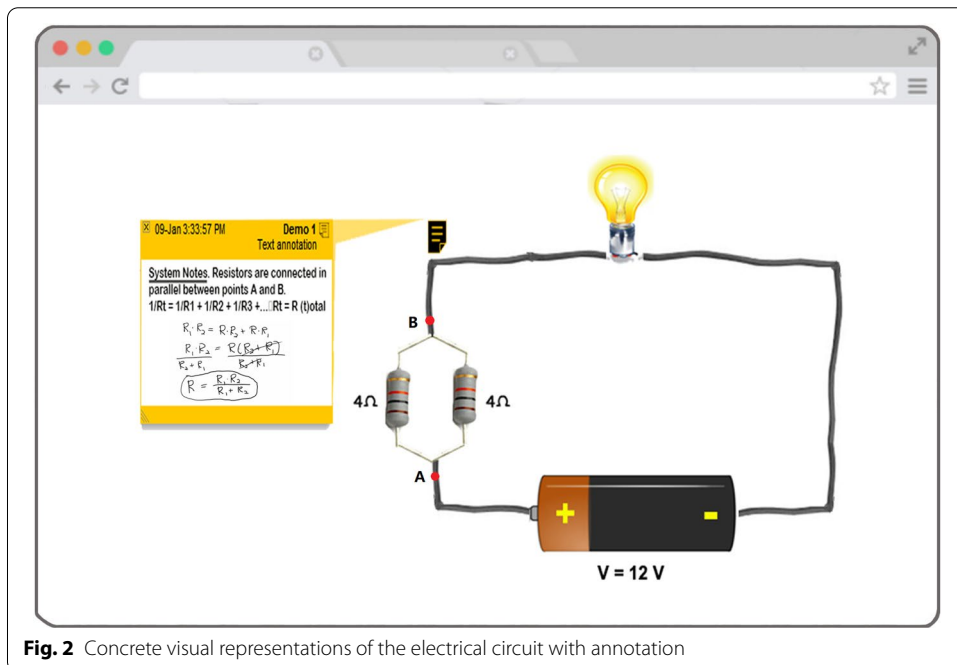

Fig. 2 Concrete visual representations of the electrical circuit with annotation

looked at the differences between abstract and concrete visual representations in electric circuitry. They found that concrete visual representations promoted better comprehension and problem solving by depicting a close correspondence between the representations and the concrete objects that they intended to represent, which made concrete visual representations rely less on knowledge conventions for their interpretations. They claimed that concrete visual representations help build the connection between learners'

prior knowledge and the information to be learned. Abstract visual representations were found to promote problem solving by focusing learners' attention on structural rather than superficial problem information. In other words, abstract visual representations generalize the common underlying structure of the problems that are superficially dissimilar. This knowledge of the general underlying structure of the problems is important in supporting transfer in learning. Moreno et al. therefore concluded that abstract visuals facilitate knowledge transfer while concrete visuals connect with schemata. However, de Bock et al. (2011) compared abstract and concrete visual representations in mathematics learning and came to the conclusion that there were no differences between abstract and concrete visual representations. They found both abstract and concrete visual representation groups performed equally well in knowledge transfer. They speculated that the non-significance in different visuals may be accounted for by learners' abilities to derive the underlying structure in the mathematics problems in both groups regardless of the forms of visual representation. Given the equivocal findings from abstract and concrete visual representation research, further investigation on abstract and concrete visual representations is warranted.

Role of annotation in visual learning

Research in neuroscience has demonstrated that visual stimuli like visual cues can significantly influence brain functions in the domains of attention and executive functioning (Heinrich et al., 2007). Madsen et al. (2013) examined the effects of visually cued and uncued diagrams on learners' cognitive processes in aerodynamics learning. They found that compared to the visually uncued condition, learners in the visually cued condition spent less time looking at "novicelike" areas of the diagram and more time at the "expertlike" areas of the diagram in transfer problem solving.

As a visual cueing strategy, annotation has recently received significant attention among researchers and educators who view annotation as a way to augment the learning process—particularly in science education. Studies have shown that annotation provokes individuals to consider and weigh new perspectives. It supports critical thinking and augments deep processing in learning (Samuel et al., 2011; Wallen et al., 2005). Wallen et al. (2005) note that annotation aids the process of selecting relevant information, organizing the information in memory, and integrating new information with prior knowledge. Moreover, it redirects learners' attention and reduces their visual search, resulting in a reduced extraneous cognitive load in learning.

There are many ways to instantiate annotation in learning. The current study chose to provide an annotation tool with which learners could, for instance, review information about electric circuit design and the formula for calculating electric resistance. They could then make annotations as he/she solves the problem. Figures 1 and 2 illustrate the annotations made by the learners as they calculate the electric resistance in a parallel electric circuit in both abstract and concrete visual representations.

Despite what is known about annotation in learning, little research has been conducted to understand how it supports learners' cognitive processes when different visual representations (e.g., abstract vs. concrete) are used. The current study aimed to

understand the differing roles of annotation in the context of abstract and concrete visual representations in science learning.

Prior knowledge and expertise reversal effect

The expertise reversal effect pertinent to visuals in learning has gained attraction among the researchers over the past decade (Zheng & Greenberg, 2018, Zheng & Gardner, 2020; Gupta & Zheng, 2020; Kalyuga et al., 2012), largely because of the fascinating observations about how expertise may influence the outcome of visual learning (Gegenfurtner et al., 2011). Kalyuga (2009) pointed out that learning procedures and techniques that are beneficial for low-prior-knowledge learners may become relatively inefficient for high-prior-knowledge learners. In a longitudinal study, Kalyuga et al. (1998) investigated the redundancy effect between high- and low-prior-knowledge learners. They found that when diagrams were embedded in the text, novices learned the content well. After the learners underwent intensive training, however, a reversal effect was observed: Diagram-alone materials generated a higher level of performance on the subsequent tests. Kalyuga et al. explained that at the beginning, novices may not have constructed schemata to understand the complex content. Therefore, the text with the diagram helped the novices comprehend the content. However, as the learners gained more knowledge, their learning actually became hindered when additional text was added, since it was then unnecessary and redundant in learning. Kalyuga et al. (2012) argued that when this redundant information cannot be ignored, interference with learning occurs resulting in high extraneous cognitive load as well as a misalignment between learners' effort and task difficulty. Lee et al. (2006) conducted a study on task complexity (high vs. low), visual representation (symbol vs. icon), and prior knowledge (high vs. low). An expertise reversal effect was observed. Low-prior-knowledge learners performed better with symbolic and iconic visual representations, whereas high-prior-knowledge learners performed better with symbolic visual representations only. Lee et al. explained that the multiple visuals were necessary scaffolds for low-prior-knowledge learners who lacked adequate schemata. However, these same scaffolds became redundant to high-prior-knowledge learners and consequently hindered their learning. The current study therefore investigated the role of expertise in visual learning.

Three predictions were made to guide the present study.

Prediction 1: Learners with annotation will outperform these without annotation as measured by learners' comprehension, problem solving, and three types of cognitive load.

Prediction 2: There will be an interaction between annotation and abstract/concrete visual representations as measured by learners' comprehension, problem solving, and three types of cognitive load.

Prediction 3: Learners' performance in annotation and abstract/concrete visual representations will be affected by their expertise in the domain area. Specifically, high-prior-knowledge learners will perform better in the abstract visual representation with annotation (AA) condition, since abstract visuals facilitate the understanding of the underlying structure of the problems. In contrast, low-prior-knowledge learners will perform better in the concrete visual representation with annotation (CA) condition,

because the extra visual support from both concrete visuals and annotation will facilitate schema development for the novices.

Two studies were conducted to test the above predictions. Study 1 explored a two-way interaction between annotation and visual representation with a focus on the difference between annotation and non-annotation in abstract and concrete visual representations. Study 2 tested a three-way interaction between visual representation (abstract vs. concrete), annotation (annotation vs. non-annotation), and prior knowledge (high vs. low). The purpose was to understand how visual representation and annotation may be influenced by learners' prior knowledge.

Study 1

To test predictions 1 and 2, Study 1 investigated (a) the differences between annotation and non-annotation and (b) the relationship between annotation (annotation vs. non-annotation) and visual representations (abstract vs. concrete) in science learning.

Methodology

Subjects and design   Participants ($N=108$) were recruited from a Research I university in the western USA. The average age of the subjects was 21.5 (SD$=1.60$). Of 108 subjects, 49% ($n=53$) were males and 51% ($n=55$) were females. About 57.4% ($n=62$) were white, 4.6% ($n=5$) were African American, 20.4% ($n=22$) were Hispanic, 14.8% ($n=16$) were Asian, and 2.8% ($n=3$) were other. A 2×2 between-subjects factorial design was employed with visual representation (abstract vs. concrete) and annotation (annotation vs. non-annotation) as the independent variables, and comprehension, problem solving, and cognitive load scores (intrinsic, extraneous, and germane) as dependent variables. The pretest scores on learners' prior knowledge of electrical circuitry were entered as a covariate. Four conditions were created. They included Abstract visual+Annotation (AA), Abstract visual+Non-annotation (ANA), Concrete visual+Annotation (CA), and Concrete visual+Non-annotation (CNA). Subjects were randomly assigned to one of the four conditions. A family-wise alpha level of 0.05 was adopted for all analyses.

Learning materials   The learning materials were created with Adobe DX to support interactive computer-based learning in electrical circuitry. The content was adapted from a textbook by Herman (2016). The learning materials covered the concepts of electric circuit (e.g., parallel and multiple resistance in an electric circuit) with electric circuit problems requiring the learners to solve them with Ohm's law. The learning materials had built-in annotation support where the participants were able to click the annotation button to get the resources (e.g., formula for calculating the resistance) and enter their own notes if needed.

Measurement   The measurement for Study 1 included prior knowledge test (PKT), the posttest, and cognitive load questionnaire (CLQ), the details of which are described as follows.
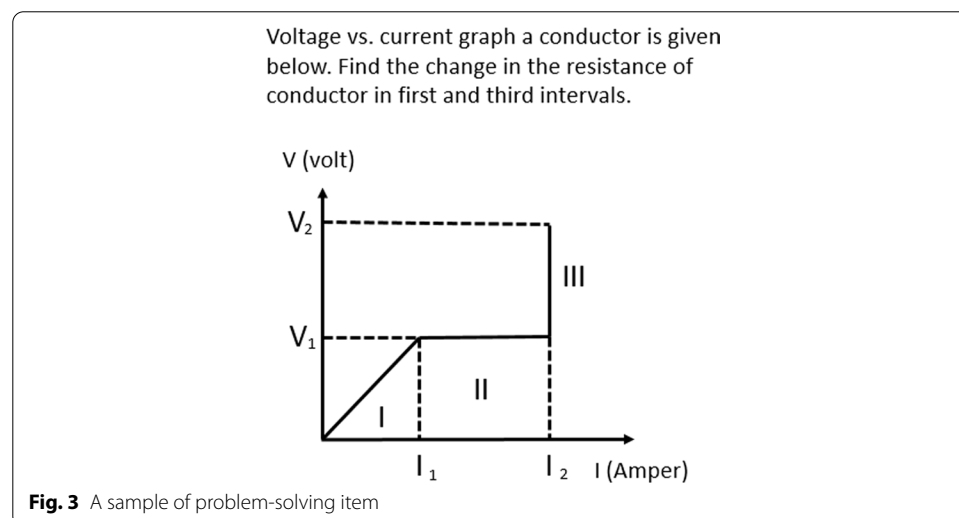
Prior knowledge test (PKT) The PKT consisted of 10 items aiming to test learners' prior knowledge in electric circuitry. The test included basic concepts from elements of

electric circuit (e.g., current, voltage) to the types of electric circuits (e.g., series, parallel). The maximum score one could obtain was 10 points. The PKT was adapted from a screen test in Herman's (2016) textbook. The items were reviewed by a panel of experts whose feedback was incorporated in the finalization of the instrument. The internal consistency for the current study for the PKT was $\alpha = 0.806$, suggesting good reliability for measuring the prior knowledge on the subject.

The posttest The posttests consisted of a comprehension test and problem-solving test. The comprehension test had ten questions assessing learners' understanding of the concepts and principles related to electrical circuit such as explaining the difference between parallel and multiple resistance in electric circuit. The maximum score one could obtain on the comprehension test was 10 points. The problem-solving test included five near transfer problems with a maximum of 10 points possible for the entire test. In the problem-solving test, learners were asked to solve a problem based on a given condition. The learner would calculate, for example, the level of resistance in the electric current using Ohm's law and then find a solution for the proper functioning of the electric circuit. The inter-item reliabilities for comprehension and problem solving were $\alpha = 0.812$ and $\alpha = 0.721$, respectively, showing medium to high reliabilities. Figure 3 provides an example of a problem-solving test item.

Cognitive load questionnaire (CLQ) A self-report questionnaire that evaluates the cognitive load in learning was used. The 11-point Likert-scale questionnaire ($N = 10$) was adapted from Leppink et al. (2013) to measure three types of cognitive load: intrinsic load (IL) items 1–3, extraneous load (EL) items 4–6, and germane load (GL) items 7–10. Examples of the questions include "The topic covered in the electric circuit material was very complex" (IL), "The instruction and explanation during the learning were very ineffective" (EL), and "The annotation with visuals really enhanced my understanding of the content covered" (GL) (see Appendix for the whole questionnaire).

The instrument reports medium to high reliabilities with intrinsic load $\alpha = 0.81$, extraneous load $\alpha = 0.75$, and germane load $\alpha = 0.82$. The current study reported

Fig. 3 A sample of problem-solving item

similar reliabilities with intrinsic load $\alpha = 0.81$, extraneous load $\alpha = 0.76$, and germane load $\alpha = 0.84$, all indicating good reliability.

Procedure   Participants were informed of the nature of the study and completed the consenting process before participating in the study. They were then randomly assigned to one of the four learning conditions: AA, ANA, CA, and CNA. The participants completed a demographic survey and a prior knowledge test. They were then told to log onto the computer to start the learning session that included the electrical circuitry materials followed by a practice session. At the end of practice session, the participants were asked to complete a posttest that consisted of comprehension and problem-solving subtests. Finally, the CLQ was administered. The entire study took about one and a half hours. The data were aggregated for final analyses.

Results

Statistical assumptions were evaluated and met. The MANCOVA was performed using SPSS v. 26 with annotation (annotation vs. non-annotation) and visualization (abstract vs. concrete) as independent variables and comprehension, problem solving, and three cognitive load scores as dependent variables. As the raw scores for the three cognitive load measures varied due to the differing number of questions in each load category (IL = 3, EL = 3, GL = 4), Z-scores were calculated to allow for meaningful comparisons among the outcomes. Prior knowledge scores were entered as a covariant in the final analyses. Table 1 presents the descriptive statistics with means and standard deviations.
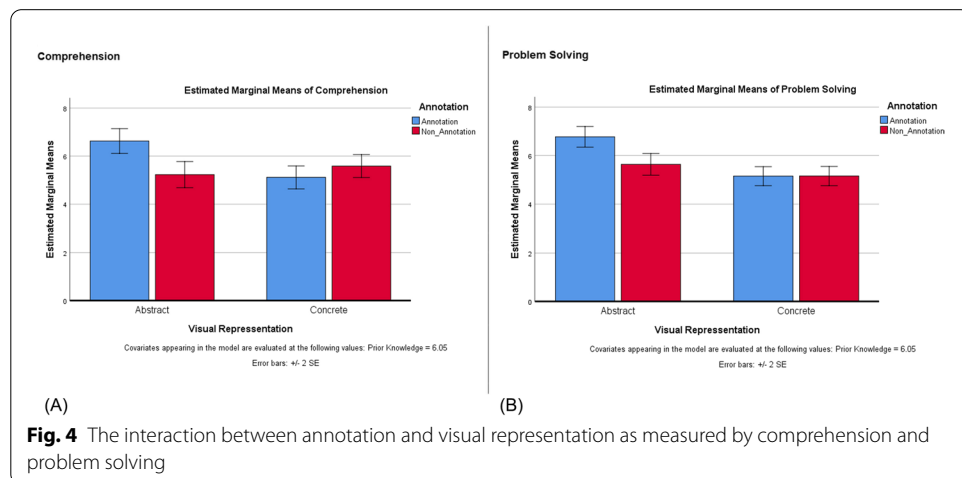
The results of multivariate tests show prior knowledge as a covariant was significant $\lambda = 0.59$, $p < 0.001$, $\eta^2 = 0.40$. Main effects were observed for annotation $\lambda = 0.69$, $p < 0.001$, $\eta^2 = 0.30$ and visual representation $\lambda = 0.70$, $p < 0.001$, $\eta^2 = 0.29$. The follow-up between-subjects tests revealed that there was a significant difference in annotation measured by problem solving $F(1, 107) = 7.38$, $p < 0.01$, $\eta^2 = 0.06$, but not by comprehension $F(1, 107) = 3.33$, $p = 0.07$.

It was found that learners who studied in the AA condition generally outperformed these in the CA condition (Fig. 4). There was a significant interaction between visual representation and annotation $\lambda = 0.76$, $p < 0.001$, $\eta^2 = 0.23$ as measured by comprehension $F(1, 107) = 11.73$, $p < 0.001$, $\eta^2 = 0.10$, problem solving $F(1, 107) = 6.41$, $p < 0.05$, $\eta^2 = 0.05$ and germane cognitive load $F(1, 107) = 8.89$, $p < 0.01$, $\eta^2 = 0.07$, but not by intrinsic cognitive load $F(1, 107) = 1.56$, $p = 0.213$ and extraneous cognitive load $F(1, 107) = 3.83$, $p = 0.053$, suggesting a connection between germane cognitive load and performance.

As expected, learners experienced higher extraneous cognitive load without annotation than with annotation $F(1, 107) = 39.11$, $p < 0.001$, $\eta^2 = 0.27$. A significant difference in intrinsic cognitive load was observed for visual representation $F(1, 107) = 5.48$, $p < 0.05$, $\eta^2 = 0.05$ where learners experienced higher intrinsic load in concrete visual representation than in abstract visual representation conditions. Finally, germane cognitive load was significant for annotation $F(1, 107) = 8.17$, $p < 0.01$, $\eta^2 = 0.07$. The interaction between annotation and visual representation was significant $F(1, 107) = 8.89$, $p < 0.01$, $\eta^2 = 0.07$.

Table 1 Descriptive statistics with means and standard deviations for study 1 ($N = 108$)

	Annotation	Visual representation	Mean	SD	N
Comprehension	Annotation	Abstract	6.92	0.68	26
		Concrete	5.11	1.57	27
		Total	6.00	1.51	53
	Non-annotation	Abstract	4.78	1.60	27
		Concrete	5.75	1.11	28
		Total	5.27	1.44	55
Problem Solving	Annotation	Abstract	7.27	0.72	26
		Concrete	5.15	1.56	27
		Total	6.19	1.61	53
	Non-annotation	Abstract	4.89	1.21	27
		Concrete	5.43	1.26	28
		Total	5.16	1.25	55
Z_Intrinsic	Annotation	Abstract	− 2.86	0.72	26
		Concrete	− 1.31	0.73	27
		Total	− 0.20	0.72	53
	Non-annotation	Abstract	− 0.39	0.92	27
		Concrete	0.21	0.90	28
		Total	− 0.08	0.95	55
Z_Extraneous	Annotation	Abstract	− 0.41	0.35	26
		Concrete	− 0.71	0.82	27
		Total	− 0.57	0.65	53
	Non-annotation	Abstract	0.26	0.81	27
		Concrete	0.75	1.17	28
		Total	0.51	1.03	55
Z_Germane	Annotation	Abstract	0.47	1.04	26
		Concrete	0.23	0.78	27
		Total	0.35	0.92	53
	Non-annotation	Abstract	− 0.21	0.63	27
		Concrete	0.22	0.56	28
		Total	0.01	0.63	55

Fig. 4 The interaction between annotation and visual representation as measured by comprehension and problem solving

As expected, learners experienced higher extraneous cognitive load without annotation than with annotation $F(1, 107) = 39.11$, $p < 0.001$, $\eta^2 = 0.27$. A significant difference in intrinsic cognitive load was observed for visual representation $F(1, 107) = 5.48$, $p < 0.05$, $\eta^2 = 0.05$, indicating learners experienced higher intrinsic load in concrete visual representation than in abstract visual representation conditions. Finally, germane cognitive load was significant for annotation $F(1, 107) = 8.17$, $p < 0.01$, $\eta^2 = 0.07$ revealing the relationship between germane load and annotation. There was a significant interaction between annotation and visual representation $F(1, 107) = 8.89$, $p < 0.01$, $\eta^2 = 0.07$ suggesting that the types of visual representation were related to the presence of annotation in science learning.

Regardless of the significant interaction between visual representation and annotation, the results, however, remained inconclusive. As Kalyuga (2007) noted, the effects of instructional strategies may differ relative to learners' prior knowledge. Given the significance of prior knowledge as a covariant in Study 1, a follow-up study that examined the impact of prior knowledge on visual representation and annotation was called for.

Study 2

Two hundred and twenty-seven participants were recruited from the same university. Of 227 participants, 59% ($n = 135$) were females and 41% ($n = 92$) were males. The average age of the subjects was 22.5 ($SD = 1.72$). About 63% ($n = 143$) were white, 7% ($n = 16$) were African American, 9.7% ($n = 22$) were Hispanic, 15% ($n = 34$) were Asian, and 5.3% ($n = 12$) were other.

Methodology

The design and measurement in Study 2 were similar to these in Study 1. The materials in learning and practice sessions were the same as these in Study 1.

Procedure   The procedure in Study 2 was almost the same as Study 1 except that the participants were divided into high- and low-prior-knowledge groups based on the pretest and then randomly assigned to one of the AA, ANA, CA, and CNA conditions.

Defining high- and low-prior-knowledge learners Two different methods were considered when defining high- and low-prior-knowledge learners. They were: median split method and tri-split method. The median split method finds the median point and splits a continuous variable like prior knowledge into half (Rucker et al., 2015). The drawback of median split method is that it arbitrarily defines the participants who are one position above and below the median point as high- or low-prior-knowledge learners which, as Liu and Reed (1994) point out, may significantly skew the results. McClelland et al. (2015) warn that median-split method is likely to increase Type II error. In contrast to median split method, Liu and Reed (1994) proposed a tri-split method that divided the participants into upper-third quarter, middle-third quarter, and lower-third quarter. It eliminates the middle-third quarter and keeps only the upper and lower third quarters in its final analysis. Since the tri-split method eliminates middle one-third sample, it clearly creates the high and low categories by retaining top and bottom one-third samples, thus avoiding artificially labelling the samples as high or low and minimizing the risk of

Type II error. Based on the results of the pretest ($N = 227$, $M = 5.54$, $\sigma = 1.56$), the participants were divided into high-, low-, and middle-prior-knowledge groups with those who scored one standard deviation above the mean as high-prior-knowledge learners ($n = 81$, $m = 7.35$, $s = 0.50$) and those scored one standard deviation below the mean as low-prior-knowledge learners ($n = 82$, $m = 3.84$, $s = 0.37$). The middle group ($n = 64$, $m = 5.45$, $s = 0.53$) were eliminated from the final analysis.

Results

To test Prediction 3, a three-way ANOVA was performed using SPSS v. 26 with annotation (annotation vs. non-annotation), visual representation (abstract vs. concrete), and prior knowledge (high vs. low) as independent variables and comprehension, problem solving and CL scores as dependent variables. Table 2 presents the descriptive statistics with means and standard deviations for Study 2.

The multivariate tests revealed a main effect for the interaction among prior knowledge, annotation, and visual representation $\lambda = 0.695$, $p < 0.001$, $\eta^2 = 0.31$. The follow-up analysis showed a significant main effect for annotation $\lambda = 0.569$, $p < 0.001$, $\eta^2 = 0.43$, visual representation $\lambda = 0.707$, $p < 0.001$, $\eta^2 = 0.29$, and prior knowledge $\lambda = 0.848$, $p < 0.001$, $\eta^2 = 0.15$. Prior knowledge was significantly interacted with visual representation $\lambda = 0.678$, $p < 0.001$, $\eta^2 = 0.32$ and annotation $\lambda = 0.906$, $p < 0.05$, $\eta^2 = 0.09$. The interaction between annotation and visualization was significant $\lambda = 0.729$, $p < 0.001$, $\eta^2 = 0.23$.

The results of between-subjects tests revealed that high-prior-knowledge learners performed better in the AA condition, whereas low-prior-knowledge learners performed better in the CA condition with a significant 3-way interaction by comprehension $F(1, 162) = 14.77$, $p < 0.001$, $\eta^2 = 0.08$ and problem solving $F(1, 162) = 7.37$, $p < 0.01$, $\eta^2 = 0.04$. Prior knowledge significantly interacted with visual representation by comprehension $F(1, 162) = 24.32$, $p < 0.001$, $\eta^2 = 0.13$ and problem solving $F(1, 162) = 6.57$, $p < 0.05$, $\eta^2 = 0.04$. However, it significantly interacted with annotation by problem solving only $F(1, 162) = 5.02$, $p < 0.05$, $\eta^2 = 0.03$ (Fig. 5).

Regarding cognitive load, a significant 3-way interaction was observed as measured by extraneous $F(1, 162) = 39.40$, $p < 0.001$, $\eta^2 = 0.20$ and germane load $F(1, 162) = 6.19$, $p < 0.05$, $\eta^2 = 0.04$. Noticeable differences were found between high- and low-prior-knowledge learners in terms of conditions. The high-prior-knowledge learners showed a lower intrinsic load in the AA condition than in the CA condition. In contrast, the low-prior-knowledge learners had a higher intrinsic load in the AA condition than in the CA condition. In terms of extraneous load, the high-prior-knowledge learners had low extraneous load in both AA and CA conditions, whereas the low-prior-knowledge learners showed a high extraneous load in the AA condition and a low extraneous load in the CA condition. Finally, the high-prior-knowledge learners showed a higher germane load in the CA condition compared to the AA condition. For low-prior-knowledge learners, the germane load was high in the CA condition but very low in the AA condition (Fig. 6).
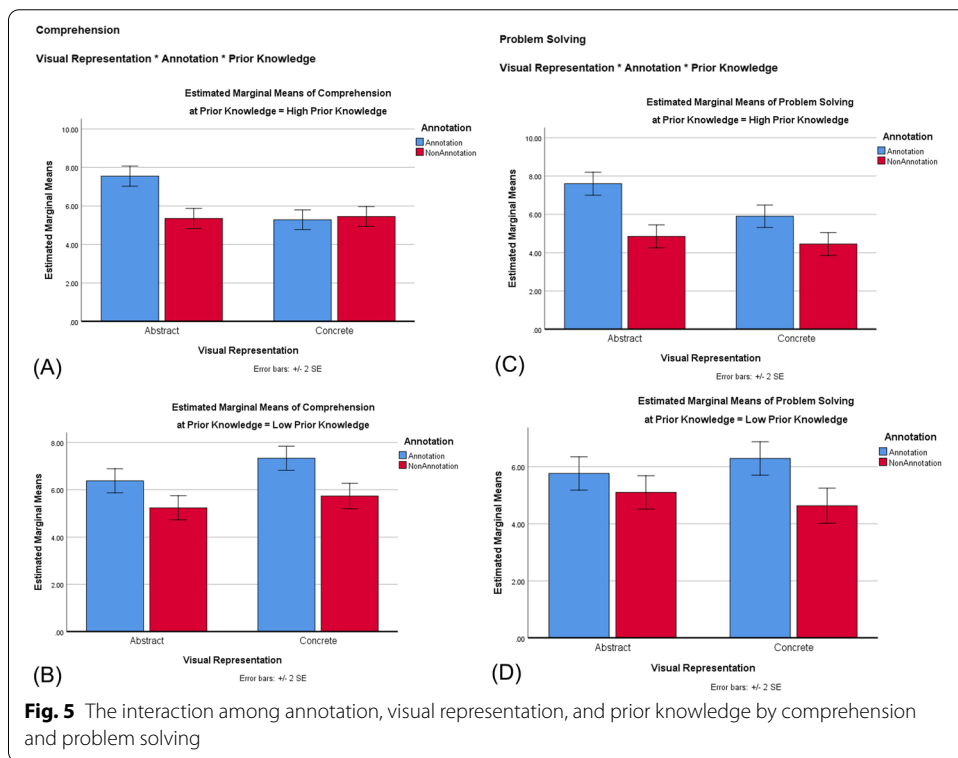
Both intrinsic $F(1, 162) = 12.24$, $p < 0.01$, $\eta^2 = 0.07$ and extraneous load $F(1, 162) = 31.86$, $p < 0.001$, $\eta^2 = 0.17$ was significant for the interaction between prior

Table 2 Descriptive statistics with means and standard deviations for study 2 ($N = 163$)

	Prior knowledge	Visual representation	Annotation	Mean	SD	N
Comprehension	High	Abstract	Annotation	7.55	0.88	20
			Non-annotation	5.35	1.13	20
			Total	6.45	1.50	40
		Concrete	Annotation	5.28	1.38	21
			Non-annotation	5.45	1.57	20
			Total	5.36	1.46	41
	Low	Abstract	Annotation	6.38	1.43	21
			Non-annotation	5.23	0.99	21
			Total	5.80	1.34	42
		Concrete	Annotation	7.33	0.85	21
			Non-annotation	5.73	0.80	19
			Total	6.57	1.15	40
Problem Solving	High	Abstract	Annotation	7.60	1.14	20
			Non-annotation	4.85	0.87	20
			Total	6.22	1.71	40
		Concrete	Annotation	5.90	1.89	21
			Non-annotation	4.45	1.46	20
			Total	5.19	1.83	41
	Low	Abstract	Annotation	5.76	1.04	21
			Non-annotation	5.09	1.17	21
			Total	5.42	1.15	42
		Concrete	Annotation	6.28	1.48	21
			Non-annotation	4.63	1.34	19
			Total	5.50	1.63	40
Z_Intrinsic	High	Abstract	Annotation	−0.45	0.66	20
			Non-annotation	−0.26	1.20	20
			Total	−0.35	0.96	40
		Concrete	Annotation	0.27	0.92	21
			Non-annotation	0.20	0.84	20
			Total	0.24	0.87	41
	Low	Abstract	Annotation	0.42	1.05	21
			Non-annotation	0.22	0.89	21
			Total	0.32	0.96	42
		Concrete	Annotation	−0.36	0.70	21
			Non-annotation	0.25	0.67	19
			Total	−0.07	0.75	40
Z_Extraneous	High	Abstract	Annotation	−0.19	0.87	20
			Non-annotation	0.08	0.38	20
			Total	−0.05	0.67	40
		Concrete	Annotation	−0.21	0.75	21
			Non-annotation	0.11	0.66	20
			Total	−0.05	0.72	41
	Low	Abstract	Annotation	1.80	0.34	21
			Non-annotation	0.17	0.52	21
			Total	0.99	0.93	42
		Concrete	Annotation	−0.60	0.86	21
			Non-annotation	0.34	0.41	19
			Total	−0.15	0.83	40
Z_Germane	High	Abstract	Annotation	0.04	0.22	20

Table 2 (continued)

	Prior knowledge	Visual representation	Annotation	Mean	SD	N
			Non-annotation	−0.18	0.20	20
			Total	−0.07	0.23	40
		Concrete	Annotation	0.09	0.13	21
			Non-annotation	0.03	0.15	20
			Total	0.06	0.15	41
	Low	Abstract	Annotation	−0.04	0.22	21
			Non-annotation	−0.02	0.18	21
			Total	−0.03	0.20	42
		Concrete	Annotation	0.33	0.15	21
			Non-annotation	0.04	0.17	19
			Total	0.04	0.16	40

Fig. 5 The interaction among annotation, visual representation, and prior knowledge by comprehension and problem solving

knowledge and visual representation. Finally, extraneous load was significant for the interaction between prior knowledge and annotation $F(1, 162) = 10.60, p < 0.01, \eta^2 = 0.06$.

Discussion

The goal of the present study was to assess the impact of annotation on abstract and concrete visual learning in the context of learners' expertise. Two studies were conducted to understand (a) the role of annotation, (b) the relationship between

Fig. 6 The interaction among annotation, visual representation, and prior knowledge by intrinsic, extraneous, and germane cognitive load

annotation and visual representation, and (c) the effects of prior knowledge on annotation and visual representation in science learning.

The role of annotation and the relationship between annotation and visual representation

The results confirmed, in general, Predictions 1 and 2 that (a) participants who learned with annotation outperformed these without annotation in science learning and (b) there was a significant interaction between annotation and visual representation. Additionally, learners who studied with annotation did better on comprehension and problem solving than these who studied without annotation, suggesting that annotation provides additional cognitive support to learning (Fig. 4). Also, learners in the AA condition performed better than these in the CA condition, which seems to contradict the literature, since concrete visuals are generally more favored in learning given that they

"make information more available in long-term memory" (Mason et al., 2013, p. 377). However, when examining the abstract and concrete visuals from the lens of annotation, it makes sense to see why abstract visuals with annotation are more effective than concrete visuals with annotation. The former supports the understanding of the underlying structure of the problem, where the latter directs learners' attention to superficial details (Moreno et al., 2011). In situations where learners did not have adequate schemata of the content domain, learning via concrete visual with annotation helped build the learners' schemata, but did not aid their ability to solve problems. Given the potential influence of schema in concrete- and abstract-annotation processing, a follow-up study (Study 2) was called for to further investigate the effects of prior knowledge on annotation and visual representations.

The effects of prior knowledge on annotation and visual representations in science learning

The results of Study 2 revealed a significant 3-way interaction among prior knowledge, annotation, and visual representation, showing an expertise reversal effect. It was found that the learners' performance in annotation and abstract/concrete visual representations were affected by their expertise in the domain area. For example, high-prior-knowledge learners performed better in the AA condition than in the CA condition in terms of comprehension and problem solving (Fig. 5). They tended to have lower intrinsic and extraneous load when learning in the AA condition than they did in the CA condition (Fig. 6a–d). The opposite was true about low-prior-knowledge learners who performed better in the CA condition than in the AA condition as measured by comprehension and problem solving. The low-prior-knowledge learners tended to have lower intrinsic and extraneous cognitive loads in the CA condition than in the AA condition. The above findings confirmed Prediction 3 that high-prior-knowledge learners would perform better in the abstract visual representation with annotation (AA), since abstract visuals facilitated the understanding of the underlying structure of the problem and that low-prior-knowledge learners would perform better in the concrete visual representation with annotation (CA) because the extra visual support from both concrete visuals and annotation would facilitate the schema development for the novices.

There are a couple of points from this study that warrant further attention. First, it was found that high-prior-knowledge learners performed better in the AA condition compared to the ANA condition. However, this difference was washed out in the CA and CNA conditions as measured by comprehension (Fig. 5a). This is probably due to the redundant effect of concrete visual representation with annotation. High-prior-knowledge learners who already have the schemata for the subject might find annotation with concrete visual representations redundant. On the other hand, low-prior-knowledge learners found the CA condition beneficial (Fig. 5b) and experienced low intrinsic and extraneous load in learning (Fig. 6b, d). The finding suggests that teachers and professional trainers should be aware of the critical role of prior knowledge in affecting learning outcomes when implementing instructional strategies like visual representations and annotation in learning.

Second, there was a shift of germane cognitive load from the AA condition to the CA condition by high-prior-knowledge learners, which is contradictory to the literature.

High-prior-knowledge learners performed well in the AA condition in both comprehension and problem solving (Fig. 5a, c) with lower intrinsic and extraneous load (Fig. 6a, c). However, their germane load in the AA condition was lower than in the CA condition (Fig. 6e). While it is not quite certain what exactly caused this shift, it is speculated that since the high-prior-knowledge learners had already developed relevant schemata as shown by their low intrinsic load (Fig. 6a), the abstract visual representation with annotation may not incentivize them enough to engage in further learning. Rather, their attention was diverted to the surface features of the learning material in the CA condition, resulting in a failure of substantial learning gain as measured by comprehension and problem solving (Fig. 5a, c). The take-away message from this finding is that teachers and professional trainers should pay attention to individualization of learning content. Experienced learners may become distracted if the content is not challenging enough to meet their Zone of Proximal Development (Vygotsky, 1978).

Conclusion

The findings of this study have demonstrated that learners' prior knowledge can significantly influence the effectiveness of visual and annotation strategies in science learning. High-prior-knowledge learners learn better when annotation is used along with abstract visual representations than with concrete visual representations. This is because annotation with abstract visual representations may better support the structuring of problems (Moreno et al., 2011) for high-prior-knowledge learners. In contrast, low-prior-knowledge learners performed better with annotation and concrete visual representations. This is because concrete visuals with annotation provide the necessary cognitive support for developing their schemata in learning.

The present study has significant theoretical and instructional implications. At the theoretical level, the study has contributed to the understanding of relationship between annotation and visual representations, particularly the differences between abstract and concrete visual representations in the context of prior knowledge. Previous research was limited to the individual roles of abstract vs. concrete visuals or annotation vs. non-annotation. The present study extends previous work by revealing the interaction between annotation and visual representations. Moreover, the study has demonstrated the learning benefits for high- and low-prior-knowledge learners in terms of the interaction between annotation and type of visual representations. This contributes to the understanding of the expertise reversal effect framework through the lenses of annotation and visual representations. At the instructional level, the study reveals the critical role of learners' prior knowledge in the design of visual representations with annotation. High-prior-knowledge learners learn better with annotation and abstract visuals, whereas low-prior-knowledge learners learn better with annotation and concrete visuals, since annotation with concrete visuals provides necessary support to novices for schema construction.

As with any empirical research, this study is not without limitations. Firstly, the study may suffer from an imbalance in gender (e.g., Study 2) that could introduce some noise to the results. Secondly, the self-report questionnaire for cognitive load measurement may suffer from response bias and/or lack of introspection on the part of participants. One such problem may have been that learners may not be able to tell "a learning task

was imposing a heavy cognitive load because it was intrinsically difficult or because the instructional procedures used were deficient" (Sweller, 2018, p. 6). Future research should explore the affective factor like motivation, self-efficacy, etc., along with cognitive factors to better understand the roles of annotation and visual representation in learning. Following that line, the study of germane load needs to be placed in the context of motivation, as germane load reflects the learners' desire to learn. Thus, a knowledge of learner motivation will add explanatory power to how germane load is related to performance (Gupta & Zheng, 2020; Moreno & Mayer, 2007; Plass & Kalyuga, 2019; Plass & Kaplan, 2016). Future studies may consider a "convergent approach" (Zheng & Cook, 2012) in cognitive load measurement by combining self-report measure with psych-physiological measures like eye-tracking and bodily sensors, as well as motivational measures like facial expression detection to better gauge various types of cognitive load in learning. Finally, future research should look into the relationship between learner working memory capacity (WMC) and cognitive load when considering annotations and visual representations, because WMC can significantly influence learners' level of cognitive load in science learning (Greenberg et al., 2021, Greenberg & Zheng, 2022).

Appendix

CLQ Questionnaire. Intrinsic Load (Items 1, 2, and 3), Extraneous Load (Items 4, 5, and 6), and Germane Load (Items 7, 8, 9, and 10). Liker Scale: 0 1 2 3 4 5 6 7 8 9 10 (0 meaning *not at all the case* and 10 meaning *completely the case*) (Adapted from Leppink et al., 2013).

1. The topic covered in the electric circuit material was very complex.
2. The topic covered the electric circuit material that I perceived as very complex.
3. The topic covered the electric circuit concepts and definitions that I perceived as very complex.
4. The instruction and explanation during the learning were very unclear.
5. The instruction and explanation during the learning were very ineffective.
6. The instruction and explanation were full of unclear verbal and visual information.
7. The annotation with visuals really enhanced my understanding of the content covered.
8. The annotation with visuals really enhanced my knowledge and understanding of the content.
9. The annotation with visuals really enhanced my understanding of the principles in electric circuits.
10. The annotation with visuals really enhanced my understanding of the concepts and definitions in electric circuits.

Author information
Robert Zheng is professor of Educational Psychology at the University of Utah. His research interests include multimedia and cognition and cognitive load theory.

Holly Cordner is a PhD student of Educational Psychology at the University of Utah. Her research interests include human memory, cognition, and educational technology.

Jeffrey Spears is an assistant professor at Utah State University. His research interests include digital technology and cognitive process in old adults.

Declarations

Author details
[1]Department of Educational Psychology, University of Utah, 1721 Campus Center Dr. SAEC 3234, Salt Lake City, UT 84112, USA. [2]Department of Social Work, Utah State University Eastern, Price, UT, USA.

References
Crisp, V., & Sweiry, E. (2006). Can a picture ruin a thousand words? The effects of visual resources in exam questions. *Educational Research, 48*(2), 139–154.
de Bock, D., Deprez, J., van Dooren, W., Roelens, M., & Verschaffel, L. (2011). Learning mathematics? A replication and elaboration of Kaminski, Sloutsky, and Heckler's study. *Journal of Research in Mathematics Education, 42*(2), 109–126.
de Koning, B. B., Tabbers, H. K., & Rikers, R. M. J. P. (2010). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction, 20*(2), 111–122.
Easterday, M. W., Aleven, V., & Scheines, R. (2009). Constructing causal diagrams to learn deliberation. *International Journal of Artificial Intelligence in Education, 19*(4), 425–445.
Frey, R., Fisher, B., & Solomon, E. (2016). A visual approach to helping instructor integrate, document, and refine active learning. *Journal of College Science Teaching, 45*(5), 20–26.
Fuchsova, M., & Korenova, L. (2019). Visualization in basic science and engineering education of fugure primary school teachers in human biology education using augmented reality. *European Journal of Contemporary Education, 8*(1), 92–102.
Gegenfurtner, A., Lehtinen, E., & Saljo, B. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review, 23*, 523–552.
Greenberg, K., & Zheng, R. (2022). Cognitive load theory and its measurement: A study of secondary tasks in relation to working memory. *Journal of Cognitive Psychology*. https://www.tandfonline.com/doi/full/10.1080/20445911.2022.2026052.
Greenberg, K., Zheng, R., Gardner, M., & Orr, M. (2021). Individual differences in visuospatial working memory capacity influence the modality effect. *Journal of Computer Assisted Learning*, *37*(3), 735–744. https://doi.org/10.1111/jcal.12519.
Gupta, U., & Zheng, R. (2020). Cognitive load in solving mathematics problems: Validating the role of motivation and the interaction among worked examples, task difficulty, and prior knowledge. *European Journal of STEM Education*, *5*(1), 05. https://doi.org/10.20897/ejsteme/9252.
Heinrich, H., Gevensleben, H., & Strehl, U. (2007). Neurofeedback—Train your brain to train behavior. *Journal of Child Psychology and Psychiatry, 48*(1), 3–16.
Herman, S. (2016). *Delmar's standard textbook of electricity* (6th ed.). Cengage Learning.
Kalyuga, S. (2007). Expertise reversal effect and its implication for learner-tailored instruction. *Educational Psychological Review, 19*, 509–539.
Kalyuga, S. (2009). *Managing cognitive load in adaptive multimedia learning*. Information Science Reference.
Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*, 1–17.
Kalyuga, S., & Plass, J. (2018). Cognitive load as a local characteristic of cognitive processes. In R. Zheng (Ed.), *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (pp. 59–74). Routledge.
Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. *Educational Psychology Review, 24*(2), 313–337.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008). The advantage of abstract examples in learning math. *Science, 320,* 454–455.

Lee, H., Plass, J., & Homer, B. (2006). Optimizing cognitive load for learning from computer-based science simulations. *Journal of Educational Psychology, 98*(4), 902–913.

Leppink, J., Paas, F., van der Vleuten, P. M., van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*(4), 1058–1072.

Liu, M., & Reed, W. M. (1994). The relationship between the learning strategies and learning styles in a hypermedia environment. *Computers in Human Behavior, 10*(4), 419–434.

Madsen, A., Rouinfar, A., Larson, A. M., Loschky, L. C., & Rebello, N. S. (2013). Can short duration visual cues influence students' reasoning and eye movements in physics problems? *Physics Education Research, 9*(2), 020104-1-020104–16.

Mason, L., Pluchino, P., Tornatora, M. C., & Ariasi, N. (2013). An eye-tracking study of learning from science text with concrete and abstract illustrations. *The Journal of Experimental Education, 81*(3), 356–384.

McClelland, G., Lynch, J., Irwin, J., Spiller, S., & Fitzsimons, G. (2015). Median splits, type II errors, and false positive consumer psychology: Don't fight the power. *Journal of Consumer Psychology, 25*(4), 679–689. https://doi.org/10.1016/j.jcps.2015.05.006

Moreno, R. (2010). Cognitive load theory: More food for thought. *Instructional Science, 38*(2), 135–141.

Moreno, R., & Mayer, R. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358–368.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*(3), 309–326.

Moreno, R., Ozogul, G., & Reisslein, M. (2011). Teaching with concrete and abstract visual representations: Effects on students' problem solving, problem representations, and learning perceptions. *Journal of Educational Psychology, 103*(1), 32–47.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.

Plass, J., & Kalyuga, S. (2019). Four ways of considering emotion in cognitive load theory. *Educational Psychology Review, 31*(2), 339–359.

Plass, J., & Kaplan, U. (2016). Emotional design in digital media for learning. In S. Tettegah & M. Gartmeier (Eds.), *Emotions, technology, design, and learning* (pp. 131–162). Elservier.

Rau, M. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychological Review, 29*(4), 717–761.

Rucker, D., McShane, B., & Preacher, K. (2015). A researcher's guide to regression, discretization, and median splits of continuous variables. *Journal of Consumer Psychology, 25*(4), 666–678. https://doi.org/10.1016/j.jcps.2015.04.004

Samuel, R. D., Kim, C., & Johnson, T. E. (2011). A study of a social annotation modeling learning system. *Journal of Educational Computing Research, 45*(1), 117–137.

Soong, R., Jenne, A., & Biswas, R. (2020). Exploring the maker culture in chemistry: Making an affordable thermal imaging system for reaction visualization. *Journal of Chemical Education, 97*(10), 3887–3891.

Sweller, J. (2018). The role of independent measures of load in cognitive load theory. In R. Zheng (Ed.), *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (pp. 3–8). Routledge.

Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction, 8*(4), 351–362.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185–233.

Sweller, J., Van Merriënboer, J. J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wallen, E., Plass, J., & Brunken, R. (2005). The function of annotations in the comprehension of scientific texts: Cognitive load effects and the impact of verbal ability. *Educational Technology Research and Development, 53*(3), 59–72.

Zheng, R., & Cook, A. (2012). Solving complex problems: A convergent approach to cognitive load measurement. *British Journal of Educational Technology, 43*(2), 233–246.

Zheng, R., & Gardner, M. (2020). *Memory in education*. New York: Routledge.

Zheng, R., & Greenberg, K. (2018). The boundary of different approaches in cognitive load measurement: Strengths and limitations. In R. Zheng (Ed.), *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (pp. 45–56). New York, NY: Routledge.

Zheng, R., Yang, W., Garcia, D., & McCadden, B. P. (2008). Effects of multimedia on schema induced analogical reasoning in science learning. *Journal of Computer Assisted Learning*, *24*, 474–482.

Publisher's Note