

Written instruction that significantly promotes science learning

Yajun Wei^{*}*Institute of Physics Education Research, Guangzhou University, Guangzhou 510006, China*

(Received 1 July 2025; accepted 10 October 2025; published 5 November 2025)

Science textbooks often include text and diagrams to aid learner understanding. Yet traditional written instructions in textbooks often separate them, which can increase cognitive load and hinder effective learning. This research proposes and tests a novel presentational format for written instruction that incorporates spatial contiguity, numbered reading sequences, and color matching, aiming to enhance science learning. The effectiveness of the novel format was assessed through two randomized experimental studies. The first study involved 349 high school students in a popular science context, and the second included 187 students in a complex physics course. The results revealed statistically significant improvements in students' learning outcomes and self-perceived understanding when exposed to the novel instructional format. The improvements were particularly notable in the study focused on complex scientific laws in physics. Furthermore, students reported lower levels of mental effort with the novel format compared to traditional formats, particularly in contexts involving more challenging materials. These results align with cognitive load theory in multimedia learning. The novel presentational format can significantly enhance student learning outcomes and understanding level. It offers a practical and more efficient alternative for instructional material design.

DOI: [10.1103/lhvj-rq13](https://doi.org/10.1103/lhvj-rq13)

I. INTRODUCTION

Multimedia learning, which combines various forms of representations, is effective in helping learners understand content. Traditional science textbooks and other forms of written instruction often separate text and diagrams, leading to the split attention effect and increased cognitive load [1,2]. Integrating text directly into diagrams can bridge this gap, but it presents challenges for complex information, as readers do not know where to start and in what sequence to proceed. This paper proposes a novel format for integrating text into diagrams using color coding and numbering to enhance comprehension and promote learning. Two controlled trials with high school students were conducted to test the effectiveness of this approach.

A. Text and diagram use in textbooks

The combination of more than one representation for conveying information related to a specific topic is referred to as multimedia learning [3], which is widely recognized for its effectiveness in aiding learners to understand the content [4,5]. Textbooks and other forms of written

instruction typically employ both text and diagrams or graphs to elucidate concepts, scientific laws, and processes, particularly in science, technology, engineering, and mathematics (STEM) fields [6]. Diagrams or graphs play a crucial role in visualizing complex information, thereby offering significant support in enhancing comprehension. Diagram content and the form in which they are presented can directly affect readers' interpretation and understanding of the diagrams [7,8].

Most STEM textbooks and other forms of written instruction place text and diagrams separately. When learners attempt to comprehend material presented in this manner, they frequently shift their attention between the text and the diagram, a phenomenon known as the split attention effect [1,2,9,10]. This effect is believed to increase cognitive load [11,12], specifically extraneous cognitive load (ECL). Increased ECL can hinder the effectiveness of learning.

B. Challenges of integrated presentation

To bridge this spatial gap, one solution is to integrate the text directly into the diagram, thereby eliminating the spatial split, as exemplified in Fig. 1 [13,14]. This involves using short text labels or annotations directly on the diagram instead of describing a process in a separate paragraph, following the famous instructional design principle of spatial contiguity. This principle posits that learners comprehend and retain information more effectively when related texts and graphics are presented near each other in space.

^{*}Contact author: weiyajun@gzhu.edu.cn

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

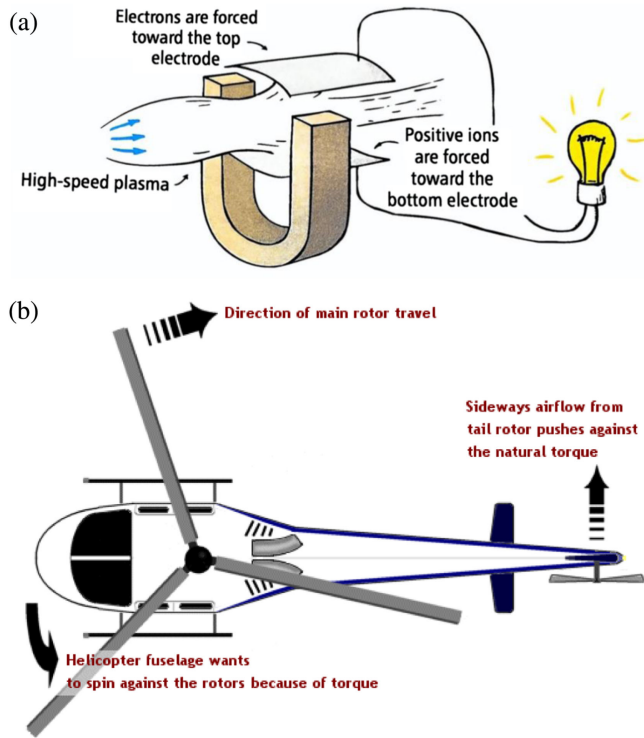


FIG. 1. Examples of written instructions with text integrated onto the diagram. (a) Text-diagram integration for explaining the science of a plasma generator in a science textbook [13] and (b) text-diagram integration for explaining why some helicopters have a tail rotor from a popular science source [14].

However, this approach introduces other challenges, especially for complex scientific processes. If a diagram contains many components and requires multiple text blocks for explanation, it becomes unclear where the reader should start and in what sequence to proceed. Additionally, it is unclear which part of the diagram should be read together with a certain text block. These challenges might explain why presentational formats such as the one in Fig. 1 are not popular and why most written instructions still separate diagrams and text [15]. Previous research on the effect of simply integrating text into diagrams has produced mixed results [16,17].

C. The novel presentational format

In this paper, we propose a novel presentation format for written instruction that is particularly suitable for complex scientific processes. In this format, text is brought onto the diagram, and the order of the text blocks is numbered and color coded. The base part of the diagram is drawn in black and gray, and other parts of the diagram are introduced in a color coded and numbered sequential manner, along with corresponding text blocks. The colors of the diagram components match those of the corresponding text blocks. For instance, to explain “What is the tail rotor of a helicopter for?,” our presentation integrates text onto the

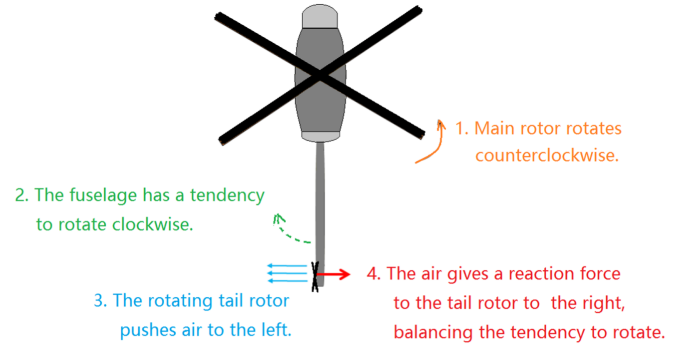


FIG. 2. Example diagram of the proposed presentational format.

diagram, color codes the corresponding text block and diagram component, and numbers them according to the reading sequence, as shown in Fig. 2.

This novel integrated presentation starts with the reader examining the gray and black part of the diagram, followed by reading the orange text labeled with number 1 and its corresponding orange diagram component. Then, the reader moves to the green text labeled with number 2 and its corresponding green diagram component, continuing this sequence until the end.

D. Rationale and hypotheses

Compared to traditional presentations that separate text and diagram, our novel format for written presentation not only integrates text onto the diagram, maintaining spatial continuity and minimizing the split attention effect, but also features numbering and color matching, which resembles a quasisquential and progressive disclosure, a technique originally used in animation videos [18–20]. We believe these features aid in managing cognitive load, allowing learners to build their understanding progressively. The numbering explicitly indicates the reading order, reducing visual search effort and potential confusion. Color matching clarifies what to read at each step (text and diagram components that have the same color), preventing confusion from unnecessary information.

Crucially, we propose that these elements—spatial contiguity, numbering, and color matching—function synergistically. While prior research has often examined such principles in isolation, their true potential may only be realized when they are integrated. For instance, numbering alone on a complex diagram would be insufficient, as learners would still struggle to identify which diagram part corresponds to a specific number. Similarly, color coding without a numbered sequence would fail to guide the learner through a logical progression of steps. It is the combination of these elements that provides a clear, unambiguous, and low-effort path for processing information. Therefore, our primary hypothesis is that this

integrated format, functioning as a holistic package, will significantly improve learning outcomes and reduce cognitive effort compared to traditional formats where these supports are absent or disconnected.

Based on these rationales, we hypothesize that the proposed integrated diagram with color coding and numbering can significantly reduce the learner's effort in reading and enhance learning outcomes. To test these hypotheses, we conducted two controlled randomized trials with high school students. The first experiment studied 10th graders' understanding of "What is the tail rotor of a helicopter for?" using traditional and our novel presentations. The second study compared the effectiveness of teaching the complex but important Lenz's law in introductory physics to 11th graders through written instruction using traditional and our novel formats.

II. THEORY AND LITERATURE REVIEW

A. Cognitive load theory

Cognitive load theory (CLT) states that learning requires resources from humans' working memory to process new information [21–23]. However, working memory is limited in its capacity regarding the amount of information that can be processed simultaneously and the time during which information can be retained. Exceeding these limitations results in cognitive overload situations.

The amount of mental resources required to perform a learning task can be described as three categories of cognitive load: intrinsic, extraneous, and germane cognitive load [24–26]. Intrinsic cognitive load (ICL) refers to all mental effort needed to process the inherent complexity of the learning task. It depends on the internal structure of the learning task and the learners' task-relevant prior knowledge. High prior knowledge means that learners can apply previously developed structures to the new information, resulting in lower cognitive load. Extraneous cognitive load (ECL) refers to the mental effort required to process the way the instructional material is presented. Therefore, it is not part of essential learning. Germane cognitive load (GCL) refers to the core processing of information to build new knowledge structures. In recent years, GCL has been assumed to have a redistributive function toward ICL and ECL and therefore not to be an independent type of load, resulting in the reclassification of cognitive load toward a 2-factor ICL and ECL model [26,27].

The split-attention effect occurs when learners have to perform additional mental integration processes because related nonredundant learning components are split up in space or time [1,9,28,29]. Consequently, they must split their attention between single components to construct a coherent mental model. This increases ECL and consumes mental resources that are then no longer available for essential learning processes [9,30].

B. Spatial contiguity in instructional design

Mayer and Fiorella outlined 12 instructional techniques to reduce extraneous processing and avoid cognitive overload situations in the context of multimedia learning [2]. In the present study, we focus on the principles of spatial contiguity [2,31,32], which address ECL by avoiding the split-attention effect [1,9,29].

Spatial contiguity is a fundamental principle in instructional design, rooted in understanding how learners process information. To address spatial separation, physical distances between related information sources should be reduced by presenting them in local proximity. Otherwise, the presentation format causes visual search processes that do not contribute to intrinsic processing but increase ECL (spatial contiguity principle; [2,29,32]).

By minimizing the need for visual scanning and reducing cognitive load, spatial contiguity ensures a more seamless and efficient transfer of knowledge. This approach is especially crucial in multimedia learning environments, where the strategic arrangement of elements on a screen can significantly influence the learner's engagement and comprehension. Implementing spatial contiguity in educational content design is therefore not just about aesthetics, but a strategic choice to optimize learning outcomes.

Empirical evidence for positive effects on learning of spatial contiguity has already been accumulated by many researchers [33,34]. These findings have been consolidated in the review by Schroeder and Cencki [29], who examined the effects of integrated design formats in different multimedia instructional scenarios.

C. Research questions

In summary, there has been consistent and continuing effort in improving instructional design to enhance learning, but no written presentational format like our proposal has been found in textbooks or the literature. Yet cognitive load theory and empirical literature suggest the potential effectiveness of instruction using the proposed presentational format.

In this research, besides proposing the novel approach, we aim to answer several research questions (RQ) through two randomized experiments. The two experiments were designed to test the robustness of our proposed format across different learning contexts: a popular science topic and a complex, curriculum-based physics concept. We hypothesized that the benefits of the novel format would be more pronounced for the more complex material. According to cognitive load theory, the inherent difficulty of a topic like Lenz's law imposes a high intrinsic cognitive load. In such situations, minimizing any additional extraneous cognitive load becomes critical. Therefore, we anticipated that the reduction in extraneous load offered by our integrated format would lead to a greater improvement in learning outcomes for the complex physics topic compared to the relatively simpler

popular science topic. These experiments were guided by the following research questions:

RQ1: Does the proposed novel written instructional format increase students' self-perceived understanding level of the content?

RQ2: Does the proposed novel written instructional format reduce students' self-perceived mental effort in reading?

RQ3: Does the proposed novel written instructional format boost students' learning?

III. STUDY 1

Study 1 aimed to compare the effectiveness of the traditional and novel instructional formats in teaching a popular science topic. Specifically, the study focused on assessing students' mental effort, self-perceived understanding level, and learning outcomes when learning about the function of a helicopter tail rotor.

A. Sample and materials

This study was conducted at a large public high school in an urban setting in China, with all instructional materials and tests administered in Chinese. The purpose of the education at this school is to teach students natural science, technology, social science, literature, and the English language, preparing them to pursue higher education after graduation. Permission to conduct the experiment was obtained from the authority, and 382 grade 10 students enrolled in an introductory engineering course participated in the study during their first formal lesson. In the end, 6 students did not hand in the questionnaire and performance test paper, resulting in an initial sample of 376 students. The ages of the students were not surveyed to save time, but due to the country's strict age requirements for K-12 education, nearly all students are aged 16 years old. The few students who fall outside this age range are most likely to be either 15 or 17 years old.

The traditional format and novel format of the instructional material are shown in Fig. 3. The traditional format of the presentation was adapted from the lecture notes of the school's enrichment course in introductory engineering (selective module on the basics of aeronautic engineering), which has been used in the school for nearly ten years. The novel format did not change the wording but integrated the text into the diagram, numbered the text blocks, and color matched them to the corresponding parts of the diagram, as shown in Fig. 3(b).

B. Instruments and research procedure

The flowchart in Fig. 4 illustrates the procedure of the study. Sheets of paper with instructions presented in either our proposed novel format or the traditional format were

(a) Helicopter Tail Rotor

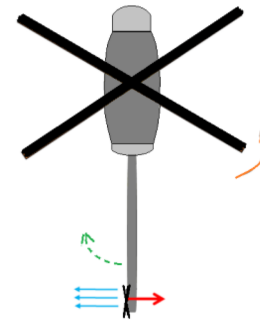
Many helicopters have a small propeller at the tail, called a tail rotor, as shown in the figure below.



Then, what is the tail rotor for?

For a helicopter to hover in air, the main rotor must keep rotating to generate lift. Due to the recoil effect, the fuselage will tend to rotate in the opposite direction. The tail rotor is used to solve this problem. The principle is shown in the figure below:

When the main rotor rotates counterclockwise, the fuselage has a tendency to rotate clockwise. The rotating tail rotor pushes the air to the left, and the air gives the tail rotor a reaction force to the right, balancing the fuselage's tendency to rotate.



(b) Helicopter Tail Rotor

Many helicopters have a small propeller at the tail, called a tail rotor, as shown in the figure below.



Then, what is the tail rotor for?

For a helicopter to hover in air, the main rotor must keep rotating to generate lift. Due to the recoil effect, the fuselage will tend to rotate in the opposite direction. The tail rotor is used to solve this problem.

The principle is shown in the figure below (read in numerical order).

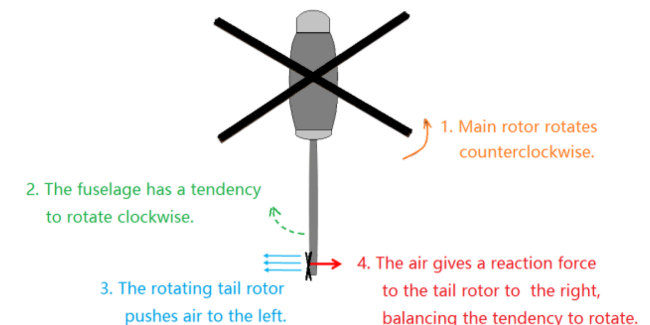


FIG. 3. Comparison between the traditional (a) and novel (b) presentational format.

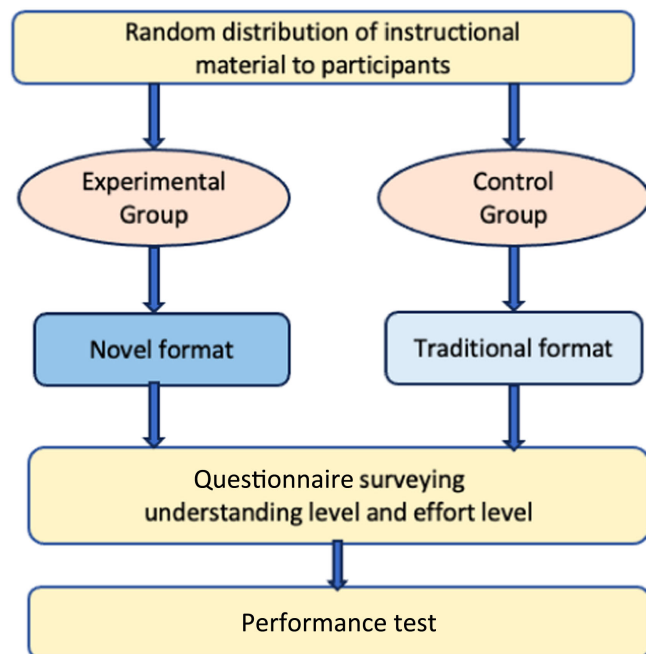


FIG. 4. Experimental procedure for study 1.

randomly distributed to the students. Instructors were trained to follow a standardized procedure for distributing materials and administering tests, and the research team monitored the process to ensure consistency. Students were asked to read the material without seeking help from the teacher or fellow students in the classroom. No electronic devices such as computers or mobile phones were allowed. Next, they were asked to fill out a questionnaire consisting of two questions, question 1 (Q1) and question 2 (Q2). Q1 asked about their self-perceived understanding level of the material, and Q2 asked about their effort in trying to understand the material. Additionally, the questionnaire asked if the participant already knew how a helicopter tail rotor works from another source before reading the provided material. Finally, the students were asked to solve a single item performance test that assessed their actual learning outcomes.

The amount of reading time given to both groups was the same (5 min). The time for completing the questionnaire and the performance test was also kept the same for the two groups (2 min for the questionnaire and 3 min for the performance test). The study protocol was approved by the school authorities and informed consent was obtained from the students and their legal guardians.

The performance-test problem was composed by a high school physics and engineering teacher with a Ph.D. It was reviewed by a team of two STEM education experts (university professors) and three experienced high school engineering teachers to ensure that it directly tested students' understanding of the science in the reading material. The performance-test problem was a single multiple-choice question, objectively rated on a binary

scale: either correct or incorrect (see Appendix A). Although this single item may limit the depth of performance assessment, the large sample size ($N = 349$) and rigorous expert review contribute to the reliability of the findings by providing a robust sample size. The percentage of correct answers on the performance test problem was calculated as the measure of students' learning outcomes, that is, the effectiveness of the written instruction.

Q1, which asked about the students' understanding level, offered the following options: A. I fully understand the content; B. I basically understand the content; C. Intermediate; D. I mostly don't understand; E. I completely don't understand the content. The answers were coded as follows: "I completely don't understand the content" was given an understanding level score of 1, and "I fully understand the content" was given a score of 5. The options "I basically understand the content," "Intermediate," and "I mostly don't understand" were given scores of 4, 3, and 2, respectively. Q2, which asked about their effort in trying to understand the material, offered the following options: A. I put full effort; B. I put most effort; C. Intermediate; D. I put little effort; E. I put almost no effort. The answers were coded as follows: "I put full effort" was given an effort level score of 5, and "I put almost no effort" was given an effort level score of 1. The options "I put most effort," "Intermediate," and "I put little effort" were given scores of 4, 3, and 2, respectively.

C. Results

In the end, 193 students in the traditional condition and 183 students in the novel condition handed in their questionnaire and performance test papers. In the novel condition, 19 students indicated that they had already learned how a helicopter tail rotor works before reading the instruction, and in the traditional condition, the number was 8. These participants were excluded from the data analysis, making the final valid response count 164 in the novel condition and 185 in the traditional condition.

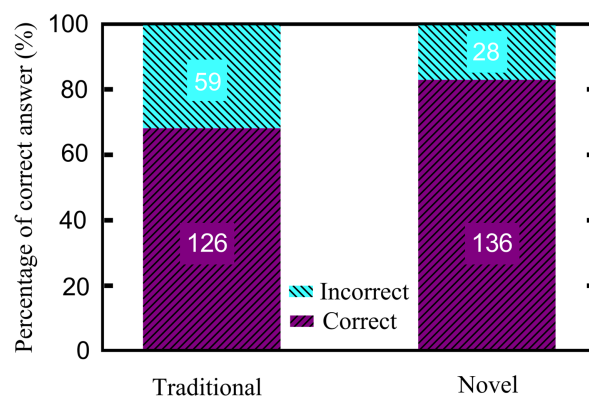


FIG. 5. Comparison of learning outcomes between the novel and traditional conditions.

TABLE I. Descriptive statistics and comparison of self-reported understanding and effort levels for study 1.

Measure	Group	<i>N</i>	Mean	SD	Median	Comparison
Q1 (understanding level)	Control group	185	4.11	0.74	4	Significant difference**, <i>U</i> test, $p < 0.05$
	Experimental group	164	4.27	0.72	4	
Q2 (effort level)	Control group	185	2.11	0.85	2	No significant difference, <i>U</i> test, $p > 0.05$
	Experimental group	164	2.01	0.87	2	

1. Learning outcomes

Figure 5 presents the percentage of correct answers to the performance test problem for the experimental group and control group. The experimental group achieved an 82.9% correction rate, which was significantly higher than the control group's 68.1% correction rate. The chi-square test with these values resulted in a chi-square statistic of 10.201 with a p value of 0.0014, indicating a statistically significant difference between the performance of the two groups.

2. Understanding level and effort level

Table I presents the descriptive statistics and comparison of self-reported understanding and effort levels coded to a 1–5 scale. A Shapiro-Wilk test was run to assess whether the distributions of responses for Q1, which measures the understanding level, and Q2, which measures the effort level, were normally distributed in both groups. The low p values in the Shapiro-Wilk tests (< 0.001 for Q1 and < 0.001 for Q2) indicate that the distributions for understanding level and effort level in both groups were not normally distributed.

Given the non-normality of the data, the Mann-Whitney *U* test, a nonparametric test, was used for comparing the means of Q1 and Q2 responses between the traditional and novel groups. The results of the Mann-Whitney *U* test indicate that there is a significant difference between the two groups in terms of the students' self-reported level of understanding ($p = 0.025$). However, their self-perceived effort does not differ significantly between the two formats ($p = 0.223$). These results indicate that students achieved a higher level of understanding with the proposed novel presentation format compared to the traditional format with the same amount of mental effort put into the learning.

Additionally, a correlation analysis was conducted to examine the relationships between performance, understanding level, and effort level. The results are presented in Table II. In study 1, no significant correlation was observed between the performance test and the understanding level (Kendall's $\tau = 0.076$, $p > 0.05$). Similarly, there was no significant correlation between the performance test and the effort level (Kendall's $\tau = -0.094$, $p > 0.05$). However, a significant negative correlation was found between the understanding level and the effort level (Kendall's $\tau = -0.658$, $p < 0.01$), indicating that higher self-reported understanding was associated with lower self-reported effort.

IV. STUDY 2

Whereas study one compares the mental effort, understanding, and actual learning outcomes between the traditional and our novel presentation of a popular science question, in this study we examine these quantities when students encounter more challenging and complex learning material.

A. Materials and sample

Lenz's law in physics makes a good exemplar material for this study. Lenz's law states that for a closed conducting loop, the induced electromotive force created by a changing magnetic flux sets up a current in the loop that tends to oppose the change in flux. The law is introduced in physics courses in high schools and universities. Previous research and surveys have found the topic to be confusing, difficult, and complex for students [35–37].

The traditional format and novel format of the instructional material are shown in Fig. 6. The traditional format of instruction is adapted from the national high school physics textbook [38]. Many classic textbooks from different parts of the world present the topic in a similar way [38–40]. The novel format of instruction does not change the wording of the text but integrates it into the diagram, with color coding and numbered sequences.

This study was conducted during the third semester of a two-year-long elementary physics course at a large public high school in an urban setting in China. The course, which is noncalculus-based, covers classical mechanics, electromagnetism, optics, and thermal physics, thereby preparing students for further studies in STEM subjects at university. The course had covered classical mechanics, static electricity, and electrical current before introducing Lenz's law and Faraday's law of electromagnetic induction. Therefore, students had similar physics education experiences.

TABLE II. Correlations between measures in study 1. * $p < 0.05$, ** $p < 0.01$.

Measure	Performance test	Q1 (understanding level)	Q2 (effort level)
Performance test	1.000	0.076	−0.094
Q1 (understanding level)	0.076	1.000	−0.658**
Q2 (effort level)	−0.094	−0.658**	1.000

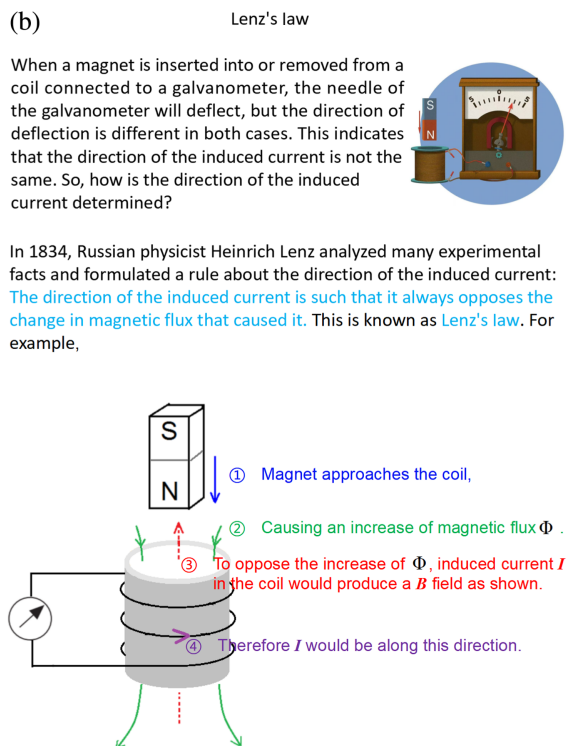
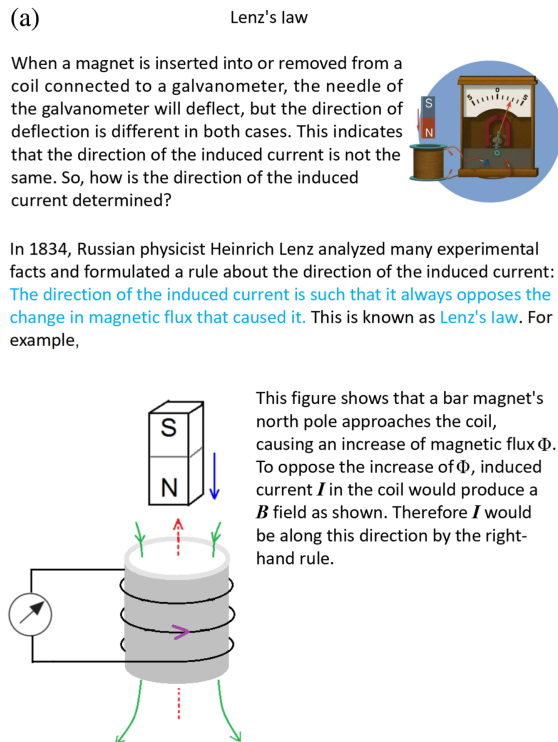


FIG. 6. Comparison between the traditional (a) and novel (b) presentational format.

An invitation to attend the experiment was sent to 202 students, and 192 consented to participate in the study. In the end, 5 students did not hand in the questionnaire and performance test paper, resulting in a final sample of

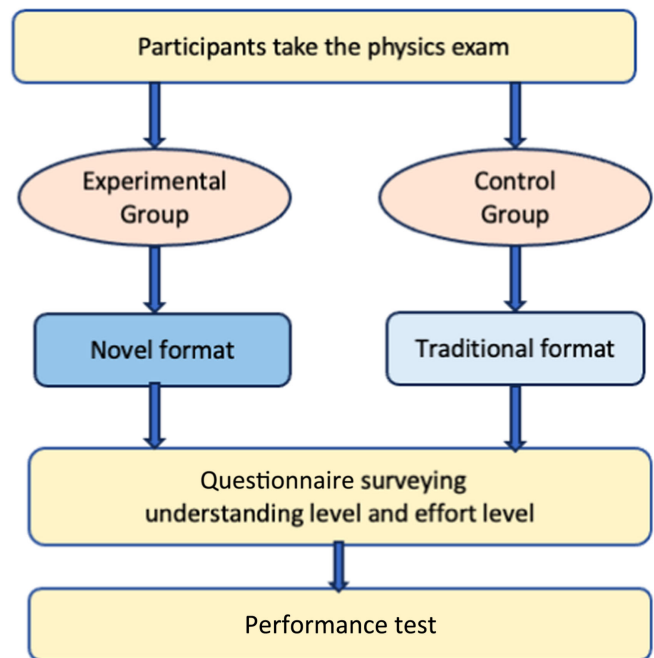


FIG. 7. Experimental procedure for study 2.

187 students—comprising 120 male and 67 female students. The participants were nearly all aged 17 years, in accordance with the country's strict age requirements for enrollment in K-12 education.

B. Instruments and research procedure

Figure 7 shows the experimental procedure of this study. First, students participated in a school wide assessment exam with physics being one of the subjects accessed. Next, participants were randomly assigned to the experimental group, which received a sheet of paper with written instructions on Lenz's law presented in our proposed novel format, and the control group, which received the traditional format. The instructor is the same for both groups, who is a member of the research team. He was trained to maintain consistency and ensured adherence to the protocol. Then, the participants were asked to fill out a questionnaire consisting of two questions. Question 1 (Q1) asked the participants to indicate their self-perceived understanding level of the material. Question 2 (Q2) asked them to report their effort in reading the material. Finally, the students were asked to solve one performance test problem that assessed their actual learning outcomes. The amount of time given to both groups was the same, being 8 min for the reading and 2 min for the questionnaire and 3 min for the performance test. The study protocol was approved by the school authorities and informed consent was obtained from the students and their legal guardians.

The questionnaire responses in study 2 were coded the same way as in study 1, using a 1–5 scale for both understanding level (Q1) and effort level (Q2).

The performance test problem was a multiple-choice question adapted from the national textbook that directly tested students' understanding of Lenz's law [38]. In addition, a team of six science education specialists, including three university physics education researchers (all with Ph.D. degrees) and three expert-level high school physics teachers (all with M.Sc. degrees), checked the content construct of the performance test problem for validity assurance. The members of the expert team unanimously concluded that the performance test problem selected from the national high school physics textbook directly tests students' mastery of Lenz's law (see Appendix B). Although using a single multiple-choice question may raise concerns about reliability and validity, this question was rigorously developed and reviewed to ensure it effectively measured the intended learning outcomes. The problem was rated purely objectively on a binary scale: either correct or incorrect. The percentage of correct answers to the performance test problem was calculated as the measure of student learning outcomes, or the effectiveness of the written instruction.

C. Results

1. Invariance of the sample groups

Among the 192 initial consenting participants, 93 in the control group and 94 in the experimental group handed in valid questionnaires and performance test papers, making the final sample size 187, with 67 girls and 120 boys. The p values in the Shapiro-Wilk tests, both above 0.05 ($p = 0.164$ of the control group, $p = 0.329$ of the experimental group), indicate that the distributions for physics exam scores in both groups are normally distributed. Figure 8 shows the distribution of their physics exam scores, which are used as an indicator of the students' prior knowledge. A t test was conducted on the physics exam scores to verify the absence of statistically significant differences in participants' prior physics knowledge. The t test results indicate good invariance in the physics exam scores. Not only are the mean values very similar, but the p values are also well above 0.05 ($p = 0.697$), and the

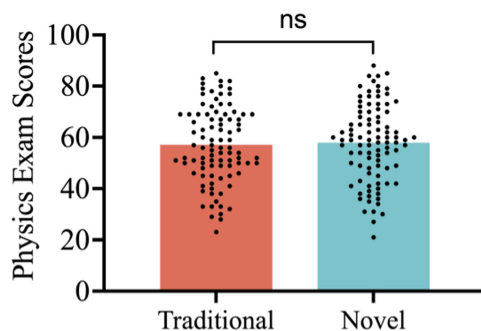


FIG. 8. Distribution of physics exam scores. ns = no significant difference.

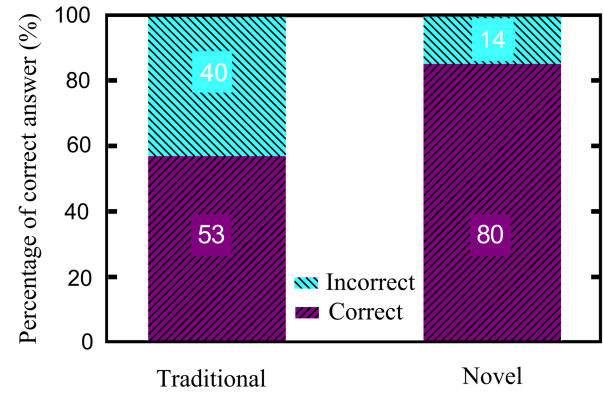


FIG. 9. Comparison of learning outcomes between the novel and traditional conditions.

distributions of the midterm physics exam scores for the different groups are very similar, as shown in Fig. 8.

2. Learning outcomes

Figure 9 presents the percentage of correct answers to the performance test problem for the experimental group and control group. The experimental group achieved an 85.1% correction rate, significantly higher than the control group's 57.0% correction rate. The chi-square test with these values results in a chi-square statistic of 17.995 with a p value of <0.001 , indicating a strong statistically significant difference between the performance of the two groups. This suggests that the proposed novel presentational format leads to significant improvement in student learning compared to the traditional presentational format.

3. Understanding level and effort level

Table III shows the comparison of the self-reported understanding level (Q1) and self-perceived effort level (Q2) of the students. Shapiro-Wilk tests were run to assess whether the responses for Q1 and Q2 are normally distributed in both groups. The low p values in the Shapiro-Wilk tests ($p < 0.001$ for Q1 of the control group, $p < 0.001$ for Q1 of the experimental group, $p < 0.001$ for Q2 of the control group, and $p < 0.001$ for Q2 of the experimental group) indicate that the distributions for understanding level and effort level in neither group are normally distributed.

Given the non-normality of the data, the nonparametric Mann-Whitney U test was conducted to compare the understanding levels between the control and experimental groups. The results of the Mann-Whitney U test indicate that there is a statistically significant difference between the two groups in terms of the students' self-reported level of understanding ($p = 0.004$).

In Table III, it can also be seen that students on average report a lower level of effort put into learning in the novel group. A Mann-Whitney U test yields a significant difference in the distribution of effort levels between the two

TABLE III. Descriptive statistics and comparison of self-reported understanding and effort levels for study 2.

Measure	Group	<i>N</i>	Mean	SD	Median	Comparison
Q1 (understanding level)	Control group	93	3.45	1.07	4	Significant difference**, <i>U</i> test, $p < 0.01$
	Experimental group	94	3.91	0.88	4	
Q2 (effort level)	Control group	93	2.70	1.18	3	Significant difference *, <i>U</i> test, $p < 0.05$
	Experimental group	94	2.33	1.05	2	

TABLE IV. Correlations between measures in study 2. * $p < 0.05$, ** $p < 0.01$.

Measure	Performance test	Q1 (understanding level)	Q2 (effort level)
Performance test	1.000	0.139*	-0.051
Q1 (understanding level)	0.139*	1.000	-0.457**
Q2 (effort level)	-0.051	-0.457**	1.000

formats ($p = 0.027$). These results indicate that using the novel format can significantly reduce the effort learners put into reading and enhance their understanding of the material.

A correlation analysis was conducted to examine the relationships between performance, understanding level, and effort level. The results are presented in Table IV, which shows a positive correlation between the performance test and the understanding level (Kendall's $\tau = 0.139$, $p < 0.05$), indicating that higher performance was associated with higher self-reported understanding. No significant correlation was observed between the performance test and the effort level (Kendall's $\tau = -0.051$, $p > 0.05$). However, a significant negative correlation was found between the understanding level and the effort level (Kendall's $\tau = -0.457$, $p < 0.01$), indicating that higher self-reported understanding was associated with lower self-reported effort.

V. DISCUSSION

This paper proposes a novel presentational format for written instruction that features spatial contiguity, a numbered reading sequence, and color matching. It also presents the results of two studies exploring student learning, self-reported understanding levels, and the effort put into learning when using the proposed format to teach in both popular science contexts and complex physics courses.

A. Interpreting the results

The results of both studies provide strong support for our hypothesis that an integrated instructional format featuring spatial contiguity, numbered sequencing, and color matching significantly improves learning. We argue that this effectiveness does not stem from any single design element, but from their powerful synergistic interaction. As proposed in our rationale, these components are

mutually dependent. The numbered sequence directs the learner's path, but it is the color coding that makes this path easy to follow by instantly linking text to the corresponding visual element, thereby minimizing the cognitive search that plagues traditional formats. The dramatic improvement in learning outcomes, especially for the complex topic of Lenz's law, underscores the value of this holistic approach. The improvement in student self-reported understanding levels might suggest it can also enhance students' confidence, which positively impacts their interest, motivation, and attitude toward science [41]. Our findings thus offer a novel contribution by demonstrating that combining these principles creates an effect greater than the sum of its parts, offering practitioners a practical and powerful template for designing effective instructional materials.

Notably, as indicated by their respective p values, the gap between the traditional format and the novel format appears more pronounced in the second study, which involves more challenging learning material. This indicates the potential superiority of the novel format, particularly in instructing complex scientific processes such as Lenz's law. The percentage of correct answers to the performance test problem in the traditional group for the popular science helicopter tail rotor question is 68% (126/185), much higher than that in the Lenz's law question, which is 57% (53/93). This also serves as indirect evidence that Lenz's law in study 2 is more challenging than the helicopter tail rotor in study 1. We define the ratio between the percentage of correct answers in the novel group versus that in the traditional group as the improvement power of the novel format. In study 2, this value is 1.5, and in study 1, it is 1.2. The improvement power of the novel format is much higher in study 2 than in study 1.

Regarding the students' self-reported level of effort put into learning, it is not statistically different in the first study, which examines student learning about the popular science question "What is the tail rotor of a helicopter for?" but significantly different in the second study, which examines student learning about Lenz's law. As mentioned before, the learning material in the second study is much more challenging and complex than that in the first study. We interpret the difference in the results of the effort level between the two studies as indicating how effective the novel format is in terms of saving the reader's effort, which is related to the nature of the learning material. For more complex scientific concepts and processes, the novel format

saves more learning effort, and for relatively easy concepts and processes, the advantage of the novel format in saving readers' effort is not obvious.

This explanation is also in line with cognitive load theory. The cognitive load theory states that the amount of mental resources required to perform a learning task can be described as two independent categories of cognitive load: ICL and ECL [24,26]. ICL refers to all mental effort needed to process the inherent complexity of the learning task. ECL refers to the mental effort required to process the way the instructional material is presented, which is not part of essential learning. For effective learning of new knowledge to take place, the total cognitive load should not surpass the working memory capacity of a learner. In the helicopter tail rotor learning case, the ICL is relatively low, so even though the traditional presentational format poses higher ECL, the total cognitive load would only surpass the working memory capacity of a small portion of the students. But in the case of Lenz's law learning, the ICL is already high, and adding high ECL through the traditional format would make the total cognitive load exceed the working memory capacity of a large portion of the students. However, the novel format adds little ECL, making the total cognitive load still manageable for most students. Therefore, the gap experienced between the experimental and control groups is much bigger for the more complex Lenz's law learning.

B. Quality control

The study design featured several controls to ensure consistency and avoid bias. First, the variables being compared were well controlled. The text of the novel presentational format and the traditional presentation format are worded identically. Also, the diagrams of the two formats are kept the same. These measures rule out all other influences on the outcome of interest and make the proposed features of the novel presentational format, namely, color coding of the text, the numbering of the text annotations, and spatial contiguity, the variables that differed between the comparisons.

Second, the performance test problem was directly related to the problem in the instruction to ensure that what was tested was what the students should have learned. Students needed to use the same scientific principles and procedures in the instruction to solve the corresponding performance test problems.

Third, study 2 was conducted with minimal disturbance of the participants' routine learning. The experiment was carried out when the students reached the corresponding topics according to their syllabus. We argue that because the students barely noticed that they were participating in an educational experiment, their performance and reactions during the written instruction would be the same as in an everyday learning context. Therefore, our results were obtained from a context that is

close to a real, everyday classroom setting rather than a contrived experimental setting.

In the first study, due to the popular science nature of the material, a prior knowledge test was less practical and quite time-consuming; therefore, we included more participants to ensure reliability. With over 350 participants purely randomly assigned to the control and experimental conditions, it gives reliable results. In the second experiment, the reading material is much more subject-specific; a physics exam score would serve as a reliable pretest to assess the participants' prior knowledge and ensure invariance between the two groups.

C. Limitations and future research

While this research contributes valuable insight into the design of instructional material, it is important to acknowledge several limitations that may impact the interpretation of the findings. A primary limitation is the use of a single multiple-choice question to assess learning outcomes and single-item scales for self-reported understanding and effort. We recognize that a single item may not fully capture the multifaceted nature of student knowledge and cognitive load. This approach was chosen due to the practical constraints of conducting the experiments within a limited classroom timeframe, but we strongly agree that future research should employ more robust, validated multi-item performance tests and established cognitive load questionnaires to ensure greater reliability and validity. This would allow for a more nuanced understanding of the format's effects.

Second, the research exclusively focuses on high school students as participants. The results may not extend to other populations, such as adult students at universities or younger learners at junior middle schools. Examining participants from different age groups and educational backgrounds would provide a more comprehensive understanding of the proposed novel written instructional format.

Another major limitation of this study is that it did not incorporate assessments for long-term retention. This means that the study did not evaluate how well the students retained the information over an extended period after the initial learning phase. Without such assessments, it is difficult to determine the durability of the learning outcomes and whether the instructional format has a sustained impact on students' understanding and retention of the material. Including long-term retention assessments would provide valuable insights into how effective the written instructional format is over time and whether it continues to benefit students' learning in the long run.

It could also be suggested that a limitation of our study is that its holistic design does not allow for the disentanglement of the individual effects of spatial contiguity, numbering, and color coding. However, we contend that

this was a deliberate and necessary design choice. The central thesis of this research is that these elements are synergistic and interdependent. To test them in isolation would be to test a different, and likely weaker, phenomenon. Our goal was to test the effectiveness of a practical, integrated instructional package, which reflects how such principles would be best applied in real-world educational materials. Therefore, this study's contribution lies in validating the combined power of the proposed format.

Future research should address these limitations by incorporating a wider range of topics and diverse learner populations to investigate the generalizability of the results. The findings of this study have the potential to catalyze future work on applying this novel written instructional format in other subjects such as engineering as well as research focusing on evaluating its effectiveness. Future work may also explore students' attention, visual search, and reading time through technologies such as eye tracking to further our understanding of the impact of the novel presentational format in enhancing learning.

VI. CONCLUSIONS

In conclusion, this paper introduces a novel presentational format for written instruction, incorporating elements such as spatial contiguity, numbered reading sequences, and color matching. Through two studies examining student learning in both popular science contexts and complex physics courses, it has been demonstrated that this novel format yields statistically significant improvements in students' learning outcomes and self-reported understanding levels.

Notably, the effectiveness of the novel format appears to be particularly pronounced when teaching more challenging and complex scientific concepts, as evidenced by the significant difference in performance between the traditional and novel formats in the study examining Lenz's law. Additionally, while there was no statistically significant difference in effort level reported by students in the study relatively easier topic in popular science context, a significant difference was observed in the study on the complex topic Lenz's law, suggesting that the novel format may be more effective in reducing learners' effort when tackling complex topics.

The findings of this study provide valuable insights into the design of instructional materials and underscore the potential of the novel presentational format to enhance student learning across various educational contexts. This research has the potential to inform future efforts aimed at improving instructional design and maximizing student engagement, comprehension, and efficiency in science learning.

The study obtained ethics approval from the school authorities. Informed consent was obtained from each participant prior to the research procedure.

ACKNOWLEDGMENTS

Guangdong Office of Philosophy and Social Sciences supported this project (GD24-CJY43). The author wishes to thank X. Chen for her help with the typesetting of the manuscript and X. Zhang for his help with the administration of the experimental studies.

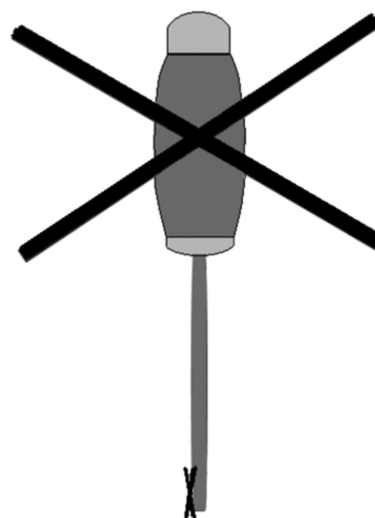
The authors declare no competing interests.

DATA AVAILABILITY

The data that support the findings of this article are openly available [42], embargo periods may apply.

APPENDIX A: POST TEST PROBLEM FOR STUDY 1

The diagram below shows a top view of a helicopter hovering in the air. If the main rotor rotates clockwise, on which side of the tail rotor would a bird feel the wind? ()

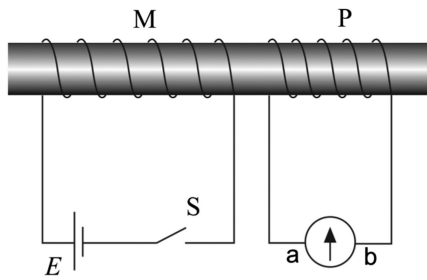


- A. Front
- B. Back
- C. Left
- D. Right
- E. Above
- F. Below

Answer Key: D

APPENDIX B: POST TEST PROBLEM FOR STUDY 2

In the diagram shown below, coils M and P are wound on the same iron core. At the moment switch S is closed, the direction of the induced current in coil P is ()



- A. From terminal **a** to terminal **b** of the galvanometer
 B. From terminal **b** to terminal **a** of the galvanometer
 C. Cannot determine the direction based on the given information
 D. No induced current will occur
 Answer Key: A

- [1] P. Chandler and J. Sweller, The split-attention effect as a factor in the design of instruction, *Br. J. Educ. Psychol.* **62**, 233 (1992).
- [2] R. E. Mayer and L. Fiorella, in *The Cambridge Handbook of Multimedia Learning*, edited by R. E. Mayer (Cambridge University Press, New York, 2014), pp. 279–315.
- [3] R. E. Mayer, Multimedia learning, *Psychol. Learn. Motiv.* **41**, 85 (2002).
- [4] A. Eitel, K. Scheiter, A. Schüler, M. Nyström, and K. Holmqvist, How a picture facilitates the process of learning from text: Evidence for scaffolding, *Learn. Instr.* **28**, 48 (2013).
- [5] N. Ott, R. Brünken, M. Vogel, and S. Malone, Multiple symbolic representations: The combination of formula and text supports problem solving in the mathematical field of propositional logic, *Learn. Instr.* **58**, 88 (2018).
- [6] S. Malone, A. Susac, J. Kuhn, and S. Küchemann, Registered reports on the role of representational competencies in multimedia learning and learning with multiple representations, *Front. Psychol.* **14**, 1260833 (2023).
- [7] L. L. Pozzer and W. M. Roth, Prevalence, function, and structure of photographs in high school biology textbooks, *J. Res. Sci. Teach.* **40**, 1089 (2003).
- [8] L. Unsworth, *Teaching Multiliteracies Across the Curriculum: Changing Contexts of Text and Image in Classroom Practice* (Open University Press, Buckingham, 2001).
- [9] P. Ayres and J. Sweller, *The Cambridge Handbook of Multimedia Learning*, edited by R. E. Mayer (Cambridge University Press, Cambridge, 2014), pp. 206–226.
- [10] P. Chandler and J. Sweller, Cognitive load theory and the format of instruction, *Cognit. Instr.* **8**, 293 (1991).
- [11] F. Paas, A. Renkl, and J. Sweller, Cognitive load theory and instructional design: Recent developments, *Educ. Psychol.* **38**, 1 (2003).
- [12] J. Sweller, P. Ayres, and S. Kalyuga, *Cognitive Load Theory* (Springer, New York, 2011).
- [13] P. G. Hewitt, *Conceptual Physics*, 12th ed. (Pearson, Boston, 2015).
- [14] Quora, How does a single-rotor helicopter work, <https://www.quora.com/How-does-a-single-rotor-helicopter-work> (2024) [Accessed 2024-04-01].
- [15] S. W. Slough, E. M. McTigue, S. Kim, and S. K. Jennings, Science textbooks' use of graphical representation: A descriptive analysis of four sixth grade science texts, *Read. Psychol.* **31**, 301 (2010).
- [16] G. Rop, A. Schüler, P. P. Verkoeijen, K. Scheiter, and T. V. Gog, The effect of layout and pacing on learning from diagrams with unnecessary text, *Appl. Cognit. Psychol.* **32**, 610 (2018).
- [17] K. Scheiter and A. Eitel, Signals foster multimedia learning by supporting integration of highlighted text and diagram elements, *Learn. Instr.* **36**, 11 (2015).
- [18] W. G. Cobourn and G. C. Lindauer, A flexible multimedia instructional module for introductory thermodynamics, *J. Eng. Educ.* **83**, 271 (1994).
- [19] J. L. Hilder, The impact of textual display strategies on learning from electronic presentations, Doctoral dissertation, Virginia Tech, 2019.
- [20] J. M. Luzón and E. Letón, Use of animated text to improve the learning of basic mathematics, *Comput. Educ.* **88**, 119 (2015).
- [21] P. A. Kirschner, J. Sweller, F. Kirschner, and J. Zambrano R, From cognitive load theory to collaborative cognitive load theory, *Int. J. Comput.-Support. Collab. Learn.* **13**, 213 (2018).
- [22] J. Sweller, Cognitive load during problem solving: Effects on learning, *Cognit. Sci.* **12**, 257 (1988).
- [23] J. J. Van Merriënboer and J. Sweller, Cognitive load theory and complex learning: Recent developments and future directions, *Educ. Psychol. Rev.* **17**, 147 (2005).
- [24] J. Leppink and A. Van den Heuvel, The evolution of cognitive load theory and its application to medical education, *Perspect. Med. Educ.* **4**, 119 (2015).
- [25] J. Sweller, J. J. Van Merriënboer, and F. G. Paas, Cognitive architecture and instructional design, *Educ. Psychol. Rev.* **10**, 251 (1998).
- [26] J. Sweller, J. J. Van Merriënboer, and F. Paas, Cognitive architecture and instructional design: 20 years later, *Educ. Psychol. Rev.* **31**, 261 (2019).

- [27] M. Thees, S. Kapp, M. P. Strzys, F. Beil, P. Lukowicz, and J. Kuhn, Effects of augmented reality on learning and cognitive load in university physics laboratory courses, *Comput. Hum. Behav.* **108**, 106316 (2020).
- [28] R. Mayer and C. Pilegard, *The Cambridge Handbook of Multimedia Learning*, 2nd ed., edited by R. Mayer (Cambridge University Press, New York, 2014), pp. 316–344.
- [29] N. L. Schroeder and A. T. Cenkci, Spatial contiguity and spatial split-attention effects in multimedia learning environments: A meta-analysis, *Educ. Psychol. Rev.* **30**, 679 (2018).
- [30] R. E. Mayer and R. Moreno, A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory, *J. Educ. Psychol.* **90**, 312 (1998).
- [31] R. E. Mayer, *Multimedia Learning*, 2nd ed. (Cambridge University Press, New York, 2009).
- [32] R. E. Mayer and R. Moreno, Nine ways to reduce cognitive load in multimedia learning, *Educ. Psychol.* **38**, 43 (2003).
- [33] P. Ginns, Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects, *Learn. Instr.* **16**, 511 (2006).
- [34] B. Çeken and N. Taşkın, Multimedia learning principles in different learning environments: A systematic review, *Smart Learn. Environ.* **9**, 19 (2022).
- [35] K. Jelicic, M. Planinic, and G. Planinsic, Analyzing high school students' reasoning about electromagnetic induction, *Phys. Rev. Phys. Educ. Res.* **13**, 010112 (2017).
- [36] A. Duffy, A pictorial approach to Lenz's law, *Phys. Teach.* **56**, 224 (2018).
- [37] A. Maries, M. J. Brundage, and C. Singh, Using the conceptual survey of electricity and magnetism to investigate progression in student understanding from introductory to advanced levels, *Phys. Rev. Phys. Educ. Res.* **18**, 020114 (2022).
- [38] Q. Peng, *Physics Textbook for High School* (People's Education Press, Beijing, 2019).
- [39] P. W. Zitzewitz, T. G. Elliott, D. G. Haase, K. A. Harper, M. R. Herzog, J. B. Nelson *et al.*, *Physics: Principles and Problems* (The McGraw-Hill Companies Inc., Columbus, OH, 2009).
- [40] C. Hamper, *Higher Level Physics: Developed Specifically for the IB Diploma* (Pearson Education Limited, Upper Saddle River, NJ, 2009).
- [41] P. Potvin and A. Hasni, Interest, motivation and attitude towards science and technology at K-12 levels: A systematic review of 12 years of educational research, *Stud. Sci. Educ.* **50**, 85 (2014).
- [42] Y. Wei, Dataset from experiments on written presentation formats, Mendeley Data, V1, 2025, [10.17632/f5kn27rnj.1](https://doi.org/10.17632/f5kn27rnj.1).