

Research Article

Comparison of Written and Spoken Instruction to Foster Coordination between Diagram and Equation in Undergraduate Physics Education

Verena Ruf ¹, Stefan Küchemann ¹, Jochen Kuhn ¹ and Pascal Klein ²

¹Technische Universität Kaiserslautern, Fachbereich Physik, Didaktik der Physik, Erwin-Schrödinger Straße Bau 46, 67663 Kaiserslautern, Germany

²Georg-August-Universität Göttingen, Fakultät für Physik, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

Correspondence should be addressed to Verena Ruf; vruf@physik.uni-kl.de

Received 28 October 2021; Revised 21 January 2022; Accepted 7 February 2022; Published 16 March 2022

Academic Editor: Zheng Yan

Copyright © 2022 Verena Ruf et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Visual-graphical representations are used to visualise information and are therefore key components of learning materials. An important type of convention-based representation in everyday contexts as well as in science, technology, engineering, and math (STEM) disciplines are vector field plots. Based on the cognitive theory of multimedia learning, we aim to optimize an instruction with symbolical-mathematical and visual-graphical representations in undergraduate physics education through spoken instruction combined with dynamic visual cues. For this purpose, we conduct a pre-post study with 38 natural science students who are divided into two groups and instructed via different modalities and with visual cues on the graphical interpretation of vector field plots. Afterward, the students rate their cognitive load. During the computer-based experiment, we record the participants' eye movements. Our results indicate that students with spoken instruction perform better than students with written instruction. This suggests that the modality effect is also applicable to mathematical-symbolical and convention-based visual-graphical representations. The differences in visual strategies imply that spoken instruction might lead to increased effort in organising and integrating information. The finding of the modality effect with higher performance during spoken instruction could be explained by deeper cognitive processing of the material.

1. Introduction

Visual-graphical representations are used in everyday contexts and especially in science, technology, engineering, and math (STEM) disciplines. An important type of representation is vector-field plots. They are representations of vector fields, such as electromagnetic fields, the flow of fluid, or force fields, consisting of a set of arrows that indicate the direction and magnitude of a field [1]. Besides the visualization as vector field plots, vector fields can be represented by mathematical equations. Whereas equations can be used to calculate parameters analytically, vector field plots can visualise a lot of information at a glance, such as directions and velocities of flow processes, intensities of force fields, or special characteristics, e.g., divergence (changes of field components in the respective direction) and curl (changes of field

components perpendicular to the respective direction). Previous studies have found that students have difficulties with graphically interpreting vector field plots [1, 2]. Consequently, it is important to support students in interpreting this representation.

To help students interpret visual-graphical representations, van Gog [3] pointed out that it is beneficial to direct students' attention to relevant areas using cues (cueing principle). Indeed, results by Klein, Viiri, and Kuhn [4] showed that visual cues helped students interpret divergence and curl in vector fields. In their work, the authors taught students a multistep visual strategy using a written text with static visualizations. The use of videos and spoken text might be better suited to explain these strategies. According to the modality effect, multimedia instruction is more conducive to learning if pictures are combined with narration rather than

written text [5, 6]. Further improvement of the instructional material by Klein, Viiri, and Kuhn [4] based on the above-mentioned principles might therefore be possible.

In summary, we aim to optimize the method of teaching the graphical interpretation of vector field plots by comparing written and spoken instructions with static or dynamic visual cues. A pre- and posttest were administered before and after learning and participants were asked to rate their cognitive load immediately after seeing the material. We analysed test scores, as well as ratings of confidence and cognitive load. In addition, eye-tracking measures were used to identify learners' attention distribution and may provide insights into learners' cognitive processes [7, 8]. To our knowledge, there has not yet been a comparison of such measures regarding the differences between spoken and written instruction combined with symbolic-mathematical and convention-based graphical representation with visual cues. Furthermore, the validation of the modality effect for other types of instruction besides text-picture combinations has practical implications for the design of multimedia learning material in other contexts.

2. Literature Review

2.1. Vector Fields. Several studies have investigated students' visual strategies when analysing vector fields. Singh and Maries [2] found that physics students have difficulty understanding changes in vector field plots, namely, divergence (flux or flow) and curl (circulation or rotation), in their graphical representations (see also [1, 9]). Vector field plots are represented by arrows containing quantitative information in the form of their lengths in two perpendicular directions, the x and y components (see Figure 1). To determine divergence and curl, students have to compare arrows in terms of their length and direction. In contrast, research showed that students had less problems determining these aspects mathematically via equations [2, 10]. Bollen et al. [10] investigated students' difficulties in understanding divergence and curl using vector fields in the context of electrodynamics and electromagnetism (see also [9]). They found that students struggled with interpreting graphical representations of vector fields, indicating a lack of conceptual understanding [9, 10].

Based on previous findings, Klein et al. [11] developed materials with multiple representations, including an equation, a graphical vector field representation, and a written instruction, to promote a step-by-step procedure for visually assessing the divergence of a field. They examined students' visual strategies for interpreting divergence in graphical representations of vector fields and discovered differences in visual processing between two strategies: the first was a graphical representation of partial derivatives visualised by comparing horizontal and vertical depictions of arrows, and the other was based on the flux concept, which involved measuring arrows arranged around an imaginary rectangle. A more detailed analysis of eye movements, especially saccadic direction, showed that students had difficulty interpreting partial derivatives [11]. Mozaffari et al. [12] were able to identify the strategy students used when graph-

ically interpreting vector field plots based on their eye movements. The scores of students taught with only one strategy peaked at 64%, indicating the need for improved instructional material [11].

Klein, Viiri, and Kuhn [4] realised this by adding visual cues to help students interpret divergence and curl in vector fields plots. The visual cues proved beneficial for learning [4]. In addition to increased performance, students who saw the cues reported lower mental effort and higher confidence regarding their performance compared to students learning with the material without visual cues [4]. This supported research findings describing that cues guide attention to relevant information [7, 13]. However, there is still room for improvement by using additional methods to facilitate learning. Research has shown that the use of multimedia learning principles based on the cognitive theory of multimedia learning has positive effects on learning (e.g., [5, 14–16]).

2.2. Cognitive Theory of and Cognitive Processes during Multimedia Learning. The cognitive theory of multimedia learning is based on three assumptions about information processing that are relevant when using multiple representations in learning material: that there are separate channels for processing visual and auditory information (dual-channel assumption), that the processing capacity of each channel is limited (limited-capacity assumption), and that active cognitive processing is required to form mental models [14, 17]. Words and images of the instructional material are received in sensory memory and then processed in verbal and visual working memory subsystems, respectively [14]. On this basis, a verbal and a pictorial mental model are created, which have to be combined and merged with prior knowledge from long-term memory in limited-capacity working memory [14].

In his cognitive theory of multimedia learning, Mayer [14] distinguished between three basic cognitive processes that learners have to perform during instruction with text and images: selecting relevant information from words and images, organising the selected words and images in a verbal and a pictorial model, respectively, and integrating this information with each other and with prior knowledge. The selection can be interpreted as a visual search or as a distribution of attention to an area of particular interest [7]. Since the goal is to build a coherent mental model of one representation, this process is also called "local coherence formation" [18]. Organisation describes the logical structuring of information within the visual and auditory channel, respectively; and integration means building cross-channel connections between various information elements [7]. For example, if the task is to determine whether the curl of a vector field is zero, the first step is to select a part of a row and a column that are orthogonal to each other in the vector field plot (selection). Then, the lengths of the arrows in x and y directions have to be compared (organisation). Finally, this information must be matched with prior knowledge to determine whether the curl of a vector field is zero or not (integration). This process is similar when an equation is given as well, with each step performed separately for the equation and integrated with graphical information and

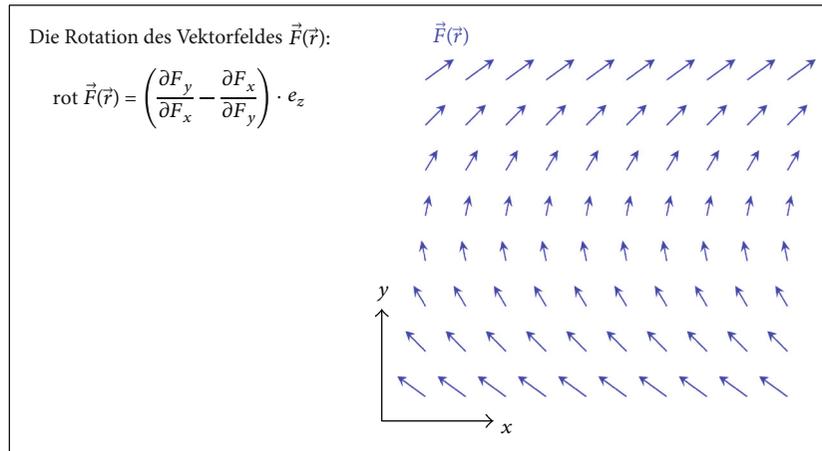


FIGURE 1: Example of a vector field on the right with a mathematical equation about determining its curl on the left.

prior knowledge. These processes do not take place once at the end of learning, but continually during learning [14].

2.3. Cognitive Load Theory and Principles of Multimedia Learning. When designing multimedia learning material, it is important to consider the expected mental resources and cognitive capacity of the learners to ensure the best possible learning opportunity. Laptops and computers are ideally suited for multimedia material, especially as Dontre [19] found no explicit disadvantages of laptop use for educational functions. Many studies examine cognitive load in multimedia learning environments (e.g., presented on a computer screen; see review by [20]). Sweller [6] described implications for learning with multimedia based on the cognitive load to better manage working memory resources, such as using different modalities (see also [5]). These assumptions are based on cognitive load theory: cognitive (mental) load consist of extraneous, intrinsic, and germane cognitive load, each based on particular aspects of learning from instructions, and unfavourable conditions can lead to mental overload and therefore insufficient working memory capacity [21].

Extraneous cognitive load depends on the design of the learning material, intrinsic cognitive load on the learning task, and its complexity, and germane cognitive load is caused by the actual learning [21]. If the extraneous or intrinsic cognitive load is too high, the remaining working memory capacity for learning is impaired [22, 23]. Considering principles of multimedia learning can reduce unnecessary cognitive load. As current challenges in online teaching include the development of new learning materials and the use of multimedia tools such as videos [24], the principles of multimedia learning might become more important.

The use of multiple representations, such as graphical representations and text, is referred to as learning with multimedia [14]. Using the cognitive theory of multimedia learning and the cognitive load theory, design principles can be developed that ensure optimal support for the three processes of information processing without impeding a limited-capacity working memory. These principles are not only applicable in conventional paper and pencil learning environments but also in computer-based online learn-

ing [15]. During COIVD-19, teachers and students were the second-largest group of digital technology users (after healthcare users, [25]). As online learning therefore becomes more common and blended learning with online and face-to-face instruction seems to be similarly effective [26], such considerations will be increasingly relevant for educators. Hughes, Costley, and Lange [27], for example, found that students reported lower extraneous load with increasing media diversity in video lectures as part of online courses. Costley and Lange [28] found an analogous relationship for germane cognitive load. Lee and List [29] compared learning strategies in text- and video-based instruction for material in the context of biology. They found that students with text-based instruction used higher-level strategies, but video-based instruction was associated with better comprehension, as indicated by the modality principle. Brünken, Plass, and Leutner [30] also reported a modality effect and found evidence that this was due to differences in cognitive capacity requirements.

Lee and List [29] looked at different modalities, such as seeing and hearing. The modality principle states that, under certain conditions, using both modalities can increase working memory capacity by reducing extraneous cognitive load [5, 16, 31]. It is important that the visual material is shown at the same time as a corresponding narration to ensure temporal contiguity [32]. For example, if part of the learning material, such as a graphic, is presented visually and the instruction is presented aurally, this can lead to higher learning gains than purely visual material. The requirement is that representations cannot be understood alone, but have to be linked together to be fully comprehended [5]. In a study with primary school children, Herrlinger et al. [33] concluded that the modality effect seemed to be a prerequisite for the multimedia effect.

Wong et al. [34] investigated the effect of the segment duration of animations as well as visual graphics in combination with speech on working memory. Their results indicated a reversal effect, i.e., segments that were too long increased working memory load (see also [35]). This denoted a constraint of the modality effect. Soicher and Becker-Blease [36] found no difference in recall or transfer

performance when they compared a learner-segmented to a nonsegmented instruction about kidney function. However, both groups viewed the presentation twice [36]. Stiller and Zinnbauer [37] found that both watching the video twice and segmentation were better than watching a continuous video in terms of procedural knowledge and transfer. In a recent article on principles for designing instructional videos, Mayer [32] recommended segmentation to help students deal with complex material. Bao [38] also proposed teaching small modules to ensure students' continuous attention. Luzón and Letón [39] found that animated hand-writing of a step-by-step solution to probability facilitated cognitive processes, such as information selection.

The ability to link different representations is also important, for example, when translating between symbolic representations, such as equations, and graphical representations. This can be facilitated by drawing attention to the connection between two representations, e.g., by using the same colours in both representations or highlighting relevant features to increase the processing of relevant information and decrease the processing of irrelevant information [3]. Signaling connections in this way is thought to stimulate germane load via organisation and integration [3]. Therefore, it can be helpful to accentuate the structure of a material (signaling/cueing principle, [14, 15]). Mayer [32] recommended this principle to help students focus on relevant information when watching instructional videos. Crooks et al. [40] investigated the interaction of the modality and cueing effect. Their results showed that participants who received written explanations of vocal articulation performed better on free and spatial recall, comprehension, and matching tests than those with spoken instructions, indicating a reverse modality effect. However, students with written instructions learned in self-paced systems that allowed them sufficient time to study the written text thoroughly in addition to the graphics. Crooks et al. [40] did not find an effect of cueing on performance in any condition. Visual cues could nonetheless be helpful under other circumstances, as Klein, Viiri, and Kuhn [4] indicated in the context of teaching a strategy for determining divergence and curl of vector fields. Schneider et al. [41] also reported a positive effect of organisational highlighting when learning with concept maps. They investigated this in combination with spatial contiguity and segmentation and recommended using either cues or segments; however, this could depend on prior knowledge [41]. Richter and Scheiter [42] noticed that students in secondary education had a better recall when they learned with multimedia signals, but only when their prior knowledge was low. The signals had no effect on recall for students with high prior knowledge. However, Richter and Scheiter [42] also found an increased subjective germane cognitive load for learners with high prior knowledge, suggesting that the signals influenced cognitive processes.

Jeung, Chandler, and Sweller [43] developed geometry learning material with high and low visual search requirements. They presented the material either visually only, with additional audio, or with audio and visual cues. Jeung et al. [43] found that in the high visual search condition, audio instructions were useful solely when cues were also shown.

Performance in the low visual search condition increased when spoken instructions were given via audio; cues were not necessary to benefit from transient information [43]. Since cognitive load influenced the effectiveness of a combination of aural and visual instruction, the results suggested that cognitive capacity should be taken into account when designing learning materials [43], implying that cueing contributed to the modality effect only when the task is sufficiently demanding. In conclusion, both the modality principle, i.e., reducing cognitive load, and the cueing principle, i.e., directing attention to relevant information, are beneficial for information processing. A combination of the two principles could be especially helpful in cognitively demanding tasks to ensure that learners can benefit from the instruction.

2.4. Eye Tracking. To investigate cognitive processes during multimedia learning, for example, to find out which of two designs enables deeper learning or requires less cognitive resources, eye movements during learning can be analysed using eye tracking. Eye tracking is a nonintrusive method for analysing participants' eye movements [44]. Typical eye-tracking measures are fixations or saccades (fast movements between fixations). Eye movements are thought to be indicators of cognitive processes [8].

Especially relevant for this study is that cognitive processes, such as selection, organisation, and integration, can be inferred from various eye-tracking measures [7]. Fixation duration is thought to be indicative of organisation, as deeper processing and longer fixations correspond [8, 45]. For example, Schmidt-Weigand Kohnert, and Glowalla [46] studied animations with spoken or written text and found that the duration students viewed animations was positively correlated with retention and transfer test scores. Schüler and Merkt [47] analysed integrative processes by examining gaze behaviour, including average fixation duration, of university students instructed via consistent or inconsistent videos. They found differences in gaze behaviour during the presentation of inconsistent information, even when students were unaware of the conflict [47]. Integrative strategies have also been researched via gaze switches between corresponding information, as they indicate a link between sources [48]. Gaze switches can, for instance, indicate difficulties in linking multimedia elements [7]. Wang, Tsai, and Tsai [49] found a negative correlation between the number of gaze switches between text and video and retention performance. In contrast, based on a review of science learning in digital environments, Yang et al. [13] suggested that concept learning, as indicated by connecting between sources, can be facilitated by increasing gaze switches between representations (see also [50, 51]). The contradictory results could be explained by the different stimuli: text with simultaneous video compared to a variety of stimuli, such as spoken and written text as well as animations.

Problem-solving processes are assumed to be similar to those in learning [52], which is why eye-tracking measures can be adopted for them. In problem solving, there are differences in visual measures between novices and experts [53]. Harsh et al. [53] found that with increasing experience, patterns became visible in eye-tracking metrics when

looking at graphs and answering multiple-choice questions. Klein, Küchemann, Brückner, Zlatkin-Troitschanskaia, and Kuhn [54] found that higher-performing students focused more attention on relevant areas in a graph than lower-performing students. Teaching patterns of how to look at graphical representations, a visual strategy, could therefore be a viable instructional approach.

2.5. Purpose of the Study. Building on previous research, instructions with cues for interpreting curl based on graphical representations of vector fields in a physical context were used. To address cognitive demands, and considering research on design principles of multimedia learning material, we investigated whether auditory rather than visual presentation of text could further support learning with visual cues. Text-based instructions with visual cues were developed by Klein, Viiri, and Kuhn [4]. The spoken instruction was developed by Dr. Küchemann and consisted of the same graphics as in the written version, with the text being spoken. The cues appeared as soon as they were mentioned in the text (temporal contiguity principle) and were assumed to act as system-paced segmentation (see segmentation principle of [15, 32]).

We postulated three hypotheses based on the cognitive theory of multimedia learning [17], the cognitive load theory (e.g., [21, 55]), and findings of Schmidt-Weigand et al. [46]:

- (1) Participants with spoken instruction perform better in the posttest due to increased active cognitive processing
- (2) The extraneous cognitive load is lower for spoken than for written instruction, as it is assumed that more working memory capacity is available due to a lower load on the visual channel
- (3) Participants with spoken instruction examine the graphic more closely during instruction

3. Methods

3.1. Participants. Thirty-eight students from the Technische Universität Kaiserslautern and the Georg-August University Göttingen voluntarily participated in the study. The students were enrolled in either physics or physics education, except for one participant who was studying mathematics. Twenty students were undergraduates, and 18 were enrolled in a graduate degree or diploma programs. Participants were on average in their fourth semester ($sd = 2.32$), ranging from second to twelfth semester, with a mean age of 21.31 years ($sd = 2.19$). The average final high school grade (Abitur) was 1.66 ($sd = 0.63$), with 1.00 being the best grade.

3.2. Procedure and Study Design. The study took place in the laboratory in the spring of 2019. After participants arrived, they gave informed consent and read informational materials about physical definitions of vector fields. They were then assigned to one of two groups: written instruction ($N = 18$) or spoken instruction ($N = 20$). Participants first completed a pretest, followed by the instruction and a cogni-

tive load test. Last, they answered the posttest. The experiment lasted approximately 45 minutes. Instruction by written or spoken text was the independent variable, which was a one-factor between-subjects design to investigate differences between types of instruction. The dependent variables consisted of:

- (i) Pre- and posttest responses regarding interpretation of vector fields (pretest: Cronbach's $\alpha = 0.34$, posttest: $\alpha = 0.74$)
- (ii) Cognitive load ratings of the instructions on a scale from 0 ("totally agree") to 5 ("totally disagree") for statements about the instruction's difficulty, language, and representations, as well as its comprehension (based on [55]). Internal consistency was comparable to that of Leppink et al. [55] (ECL: Cronbach's $\alpha = 0.93$, ICL: $\alpha = 0.81$, GCL: $\alpha = 0.89$)
- (iii) Time participants spent on learning with the instruction as a control variable
- (iv) Gaze switches between relevant areas of the instructional material, indicating integrative processes (see, e.g., [7, 48])
- (v) Proportional fixation duration on relevant areas of the instructional material as a measure of proportional attention

As an additional variable, we surveyed confidence for each response, ranging from "very confident" (1) to "very uncertain" (4). Confidence scores were used to account for possible guessing during the test.

3.3. Materials. The material consisted of a test of learning material used as a pre- and posttest, the learning material, and a cognitive load test (available under [56]).

3.3.1. Learning Material. Dr. Küchemann developed the instruction based on the material by Klein, Viiri, and Kuhn [4] in German with visual cues. It consisted of an exemplary vector field plot on the right, and the equation on the left, with the instructional text in the written instruction displayed above and below the equation. In both conditions, the instruction consisted of a step-by-step explanation in German of how to interpret graphically curl, the magnitude of rotation, of a vector field. The graphical representation of the vector field contained coloured rectangles as visual cues on how to apply the explained strategy, taking into account the direction and length of the arrows. The cues were visible throughout the written instruction and appeared as soon as they were mentioned in the audio during the spoken instruction. Otherwise, the written and spoken explanations were identical.

The learning material taught how to determine whether a vector field's curl is zero or not. First, the equation was explained (on the left-hand side of the learning material, in the written instruction preceded and followed by the instructional text). For visual interpretation, one has to select a part of a row and a column in the vector field plot that are

orthogonal to each other. These were indicated by visual cues in the x and y directions. An example of a vector field plot as shown in the learning material can be seen in Figure 2. Next, text or audio described how to compare the length of the arrows in the x and y directions to determine whether the curl is zero or not. This requires looking at the length of the arrows on the y -axis (red frame) in the x -direction (red arrows)—the first term of the mathematical equation in Figure 1—and the length of the arrows on the x -axis (yellow frame) in the y -direction (yellow arrows)—the second term of the mathematical equation. The colours refer to the visual cues as shown in Figure 2. The curl is zero if these two lengths do not change. The curl of the vector field plot shown in Figure 2 is therefore not zero.

3.3.2. Test Material. After the instruction, the students answered a cognitive load test. The questions were based on Leppink et al.'s [55] test for cognitive load. Students were asked to rate the instruction in terms of, e.g., its difficulty, language, graphical representation, and whether they understood the topic better as a result. An example statement would be (translated from German): “The content of the instruction was very complicated.”

The pre- and posttest developed by Dr. Küchemann consisted of six graphical representations of vector fields on the left side of the screen and the mathematical equation on the right (see Figure 1). The students' task was to determine whether a vector field's curl was zero or not. The vector fields were presented in a different order in the pre- and posttest. After each item, participants were asked to rate their confidence. To evaluate the test materials, we calculated the average item difficulty, the average item discriminatory index, and the average point biserial coefficient [57, 58]; pretest: $p = 0.58$, $D = 0.52$, $r_{\text{pbi}} = 0.72$, pretest: $p = 0.76$, $D = -0.09$, $r_{\text{pbi}} = 0.62$. The ideal item difficulty p is 0.5, the item discriminatory index D should be above 0.3, and the point biserial index r_{pbi} should be above 0.7 [57]. Although some values are not ideal due to ceiling effects, the tests taken together with Cronbach's α (see section 3.2) are sufficiently reliable for our purposes.

We scored pre- and posttest responses as either correct or incorrect. The responses were guess-corrected, i.e., the answers with the lowest confidence rating were always classified as incorrect. A more detailed description can be found below.

3.4. Eye-Tracking Apparatus and Measures

3.4.1. Apparatus. In line with Klein, Viiri, and Kuhn [4], the instructions were presented on a 22-inch computer screen with a resolution of 1280×960 pixels and a refresh rate of 75 Hz. The distance to the screen was about 60 cm. A Tobii X3-120 eye tracker with a sampling frequency of 120 Hz and an ideal accuracy of 0.40° of the visual angle (according to the manufacturer). As the system allowed a high degree of freedom, no chin rest was used. For more information, see Tobii [59]. Fixations and saccades were detected using an I-VT algorithm [60].

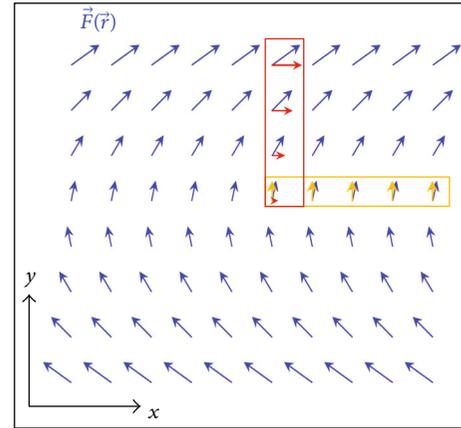


FIGURE 2: The vector field plot presented during instruction with visual cues indicating how to use the strategy explained in the text or the speech.

3.4.2. Measures. Areas of interest (AOIs), i.e., particularly interesting regions of the stimulus, form the basis of eye-tracking measures [44]. For example, a graphical representation could be an AOI; it could also be divided into several AOIs, such as coordinate systems, visual cues, and, for example, the vector field plot around it. We chose the most relevant AOIs based on their visibility in both spoken and written instruction, and whether they were considered by education experts and physicists to be most important for understanding graphical interpretation (along with comprehending the step-by-step process explained). Therefore, we selected five important AOIs: the equation, visual cues in x and y directions, the vector field plot not covered by visual cues (divided into parts above and below the cues), and the coordinate system. The equation referred to the cues in x and y directions. Since splitting the equation into the relevant parts related to each of the cues would result in very small AOIs, we analysed the equation as a whole. However, the cues could be analysed individually because of their location. We were also interested in the effectiveness of the visual cues and therefore looked at the part of the vector field plot that was not highlighted. This was divided into two parts to account for the spatial separation by the cues. We considered the coordinate system as potentially relevant, as the x and y directions given in the equation are indicated here. The full learning material, including the AOIs, can be seen in Figure 3 for instruction by written text (a) or speech (b).

In particular, the fixation duration on visual cues is an important indication of how much attention has been paid to them (see hypothesis 3). Fixations are points at which the eye is relatively static for a period of time and they indicate attention [8]. They are usually between 200-400 ms long and are identified based on their velocity (<100 deg/sec) using the I-VT filter (for details see [60]). Gaze switches between the equation, and the vector field indicate the extent to which participants have integrated the two types of representations. Gaze switches are switches between one AOI and the next (different) AOI and can be calculated manually based on AOI hits extracted from Tobii.

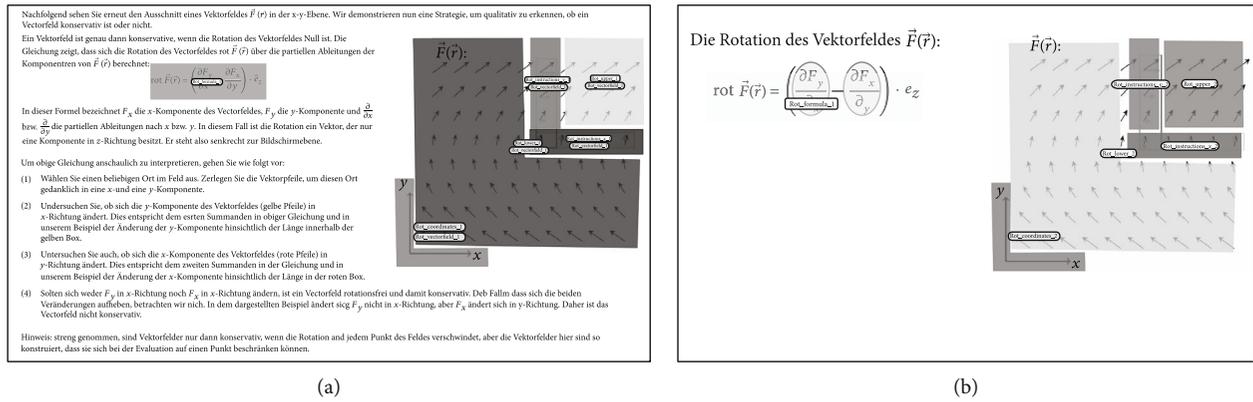


FIGURE 3: AOIs for written instruction (a) and spoken instruction (b).

4. Results

4.1. Vector Field Interpretation. Test results indicated performance in interpreting vector fields. We guess-corrected responses, i.e., the responses with the lowest confidence rating were always scored as incorrect. No outliers were identified in a box-plot of the total score; therefore, no data were excluded. Shapiro-Wilk tests showed that the data in both groups were not normally distributed in the pre- and posttest; written: pretest: $W(107) = 0.64, p < 0.001$, posttest: $W(107) = 0.62, p < 0.001$; spoken: pretest $W(119) = 0.62, p < 0.001$, posttest: $W(119) = 0.50, p < 0.001$. Levene's test indicated unequal variance in the posttest, $F(2,226) = 8.34, p = 0.004$; pretest: $F(2,226) = 2.24, p = 0.14$. As the group size was unequal and the data were nonparametric (binary), we compared pre- and posttest by Chi-squared tests using the number of correct and incorrect responses from each group. Cohen's h^1 is given for significant p values. The difference between the written and spoken instruction groups was significant in the posttest, $\chi^2(1) = 7.30, p = 0.007, h = -0.38$; pretest: $\chi^2(1) = 1.85, p = 0.17$. The proportion of correct responses between groups with written and spoken instruction can be seen in Figure 4, where the bars represent the error in sample proportions (p), indicating that learning gains were greater for the spoken instruction group; pretest: written text = 52%, spoken = 62%; posttest: written text = 62%, spoken = 79%.

We translated the learning times into z -scores. There were no outliers (z -score > 3). Learning time did not differ between the two groups; $t(36) = 0.0, p = 1.0$, written text: 294 sec (sd = 56), speech: 334 sec (sd = 46).

An overview of the mean values between groups for all variables can be seen in Table 1.

4.2. Cognitive Load. Participants reported cognitive load based on the items developed by Leppink et al. [55], which were slightly modified for this study. Items 11-14 are worded in opposite ways compared to the other items (see material [56]); therefore, the responses have been numerically reversed (reposed). No outliers were identified in the box plots of total cognitive load and the three types of cognitive load; therefore, no data were excluded. We compared total cognitive load between groups using a t -test; $t(31.34) = -0.92, p = 0.37$.

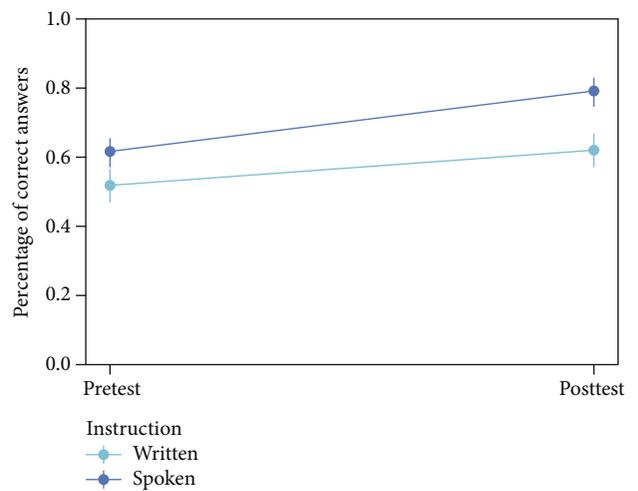


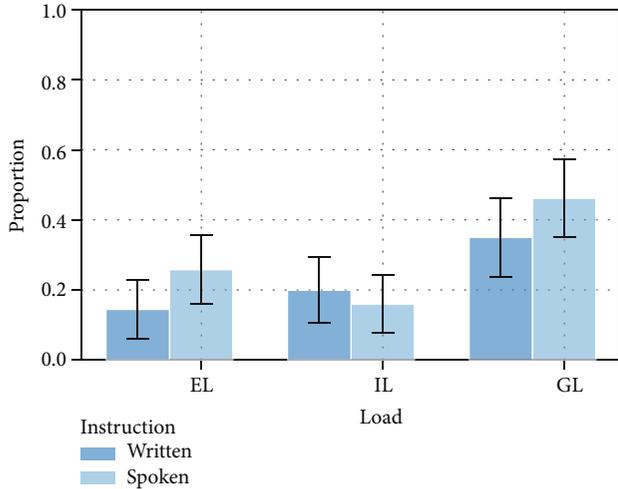
FIGURE 4: Proportion of correct answers in pre- and posttest for groups with written and spoken instruction; error calculated for proportion p .

The different types of cognitive load were distinguished using factors found by Leppink et al. [55] and Thees et al. [61]. Extraneous, intrinsic, and germane cognitive load were compared using t -tests; as Levene's test indicated unequal variances for extraneous cognitive load ($F = 4.89, p = 0.03$), a Welch-correction was used in this case; extraneous cognitive load: $t(36) = -1.37, p = 0.18$, written text: 14%, speech: 26%, intrinsic cognitive load: $t(36) = 0.66, p = 0.51$, written text: 20%, speech: 16%, germane cognitive load: $t(36) = -1.32, p = 0.19$, written text: 35%, speech: 46% (see Figure 5). The percentages indicate the average cognitive load of a possible 100% as reported by the participants.

4.3. Gaze Switches. We calculated the relative number of gaze switches for each participant by counting the number of gaze switches between particular AOIs and dividing this by the total number of gaze switches of the participant. Using a box-plot of the total number of gaze switches and a visual inspection of the data, we identified one outlier in the group with written instruction and two outliers in the group with spoken instruction and removed them for the analysis (written instruction: $N = 17$, spoken instruction:

TABLE 1: Mean values and standard deviation (in brackets) of each variable for groups with written and spoken instruction.

Variable	Pretest score	Posttest score	Learning time in sec	Overall cognitive load	Extraneous cognitive load	Intrinsic cognitive load	Germane cognitive load	Number of gaze switches	Fixation duration in %
Written	0.52 (0.50)	0.63 (0.49)	294 (56)	18.28 (9.47)	4.33 (3.53)	4.0 (3.79)	7.0 (4.90)	15.35 (8.23)	0.17 (0.07)
Spoken	0.62 (0.49)	0.83 (0.38)	334 (46)	22.15 (16.02)	7.57 (10.04)	3.2 (3.70)	9.25 (5.55)	58.78 (11.38)	0.17 (0.04)

FIGURE 5: Extraneous (EL), intrinsic (IL), and germane cognitive load (GL) for written and spoken instruction; error calculated for proportion p .

$N = 18$). Participants with spoken instruction switched between AOIs significantly more often than participants with written instruction, $t(33) = -12.86$, $p < 0.0001$, $d = 4.35$, written: 15.35 (sd = 8.24), spoken: 58.78 (sd = 11.38). We used t -tests to compare the gaze switches between the matched AOIs, the areas of the mathematical equation, and the vector field representation that were visible in both the written and the spoken instruction (consisting of cues in x and y directions, upper and lower part of the vector field, and coordinates). Statistically significant differences in gaze switches between groups with written and spoken instructions are shown in Table 2 including Cohen’s d . These were gaze switches between cues in x - and y -direction as well as between other parts of the vector field and the equation to the lower part of the vector field.

4.4. Proportional Fixation Duration. To calculate the proportion of fixation duration on an AOI, we first normalised the fixation duration for each participant by dividing the individual average fixation duration of each AOI by the sum of the average fixation duration. We calculated z -scores to remove outliers and removed participants with z -scores bigger than three. In this way, four participants with spoken instructions were excluded from further analysis ($N = 16$). Data from the remaining participants were standardised to get a mean of zero and a standard deviation of one for each group. There was no difference between groups, $t(202) = 0$, $p = 1$, written text: 17% (sd = 7), and spoken text: 17% (sd = 4). AOIs between groups were compared using t -tests and Cohen’s d (see Table 3). Participants with

TABLE 2: Statistically significant differences in gaze switches of groups with written and spoken instruction according to t -tests and Cohen’s d .

From	To	p	d
Cue x	Cue y	0.0002**	0.27
Coordinates	Lower VF	0.001**	-1.29
Cue y	Upper VF	0.03	-0.32
Equation	Lower VF	<0.0001***	-0.60

spoken instruction looked significantly longer at the coordinates and the lower part of the vector field. Participants with written instruction looked significantly longer at the cue in the x -direction (see Figure 6).

5. Discussion

We investigated the difference between written and spoken instruction regarding the assessment of the curl of vector fields. For this purpose, we conducted a study in which 38 students received either written or spoken instruction on the graphical interpretation of vector field plots. We asked participants to complete a test before and after instruction while recording their eye movements. In addition, participants rated their cognitive load immediately after instruction. The analysis of answer scores showed that students had answered correctly in slightly over half of the cases on the pretest (see Figure 4), which is in line with guessing probability. This indicated the need for further instruction in graphical representations of vector fields and their interpretation. We assumed that participants with spoken instruction examined the graphic more closely than those with written instruction (see [46]). Furthermore, we expected that participants who received spoken instruction would perform better and report lower extraneous cognitive load than those who received written instruction.

Hypothesis 1. Participants with spoken instruction perform better in the posttest than those with written instruction.

A comparison between pre- and posttest scores showed that the students with spoken instruction performed better than students with written instruction (see Figure 4). As there was no difference between groups in the pretest, these results supported our hypothesis and suggested a modality effect [5, 6, 15] as well as a beneficial effect of visual cues in combination with the modality effect [43]. In particular, participants with spoken instruction scored close to 80%

TABLE 3: T -, p -, and d -values for differences in fixation duration AOIs between groups with written and spoken instruction based on unpaired t -test and Cohen's d .

AOI	T	Dof	p	d
Coordinates	-2.70	34	0.01*	-0.9
Equation	0.01	34	1	—
Cue x	3.41	34	0.002**	1.14
Cue y	1.79	34	0.08	—
Lower VF	-2.80	34	0.01*	-0.93
Upper VF	0.30	34	0.77	—

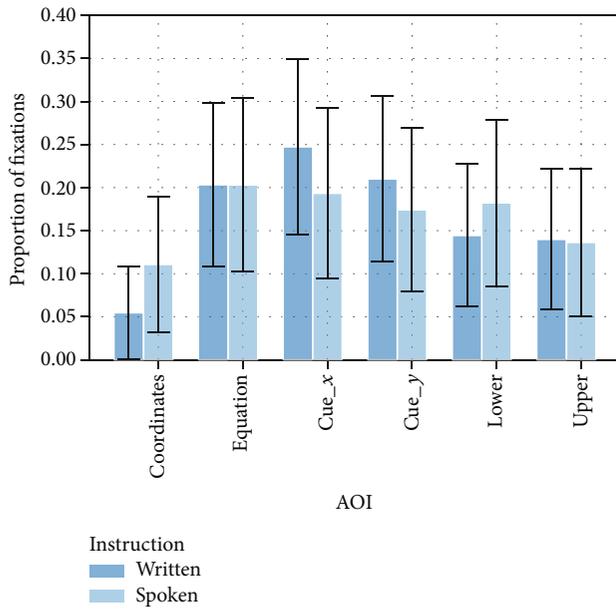


FIGURE 6: Proportion of fixations on different AOIs in groups with written and spoken instruction; error calculated for proportion p .

out of 100%, which could indicate that this type of instruction might be useful for students who have difficulty understanding curl (e.g., reported by [1, 2]). Klein, Viiri, and Kuhn [4] also noted a benefit of visual cues for written material, and our results suggest that signalling to promote coordination between equation and graphical representation (see, e.g., [3]) could be extended to multimodal types of instruction. The system-paced segmentation by the appearance of visual cues as soon as they were mentioned in the text did not seem to have any detrimental effects.

There were no differences between the groups in the pretest and the learning time of participants did not differ, suggesting that the learning gain was due to the instruction method (see [40]). As the pretest scores of the two groups were comparable, the prior knowledge of the participants seemed to be similar. Simonsmeier et al. [62] also did not find a high correlation between prior knowledge and knowledge gain in their review. However, our study design does not allow us to draw conclusions about whether modality (spoken vs. written explanation) or system-paced segmenta-

tion due to dynamic visual cues was the decisive factor. Nevertheless, the students who learned with system-paced segmentation had better results than those without segmentation. This outcome is in contrast to Soicher and Becker-Blease [36], who found no difference between segmented and nonsegmented instruction. Instead, our results are in line with Richter and Scheiter [42], who found a positive effect of cueing for learners with low prior knowledge. System-based segmentation combined with verbal instructions and abstract graphical representations could therefore be a good type of STEM learning material, especially useful in an online learning environment.

Hypothesis 2. Participants with spoken instruction report lower extraneous cognitive load than those with written instruction.

The types of cognitive load did not differ between written and spoken instructions, which contradicts the hypothesis that extraneous cognitive load was lower for instructions with visual-graphic and abstract-symbolic representations. This might be due to the need to remember previously heard information as a step-by-step process was explained. It is possible that the instruction was too long for working memory (see [34]). Moreover, the extraneous cognitive load was very low for both instruction types. This could indicate low variance in the measure as well as possible floor effects, which would make it impossible to reduce extraneous cognitive load. The reason for the low extraneous cognitive load could be that the positive effects of written text (multiple integration processes, no retrieval of content required from memory required) and verbal text (use of dual channels) compensate each other. However, Hughes et al. [27] found that extraneous load decreased with increasing media diversity. The authors suggested that students were able to choose the type of media that was associated with the least extraneous cognitive load for them. In the case of written and spoken instruction, this could mean that students who received written instruction were able to select the parts of the text and graphic representation that they found most relevant, thereby reducing their cognitive load. Since participants with spoken instruction performed better than those with written instruction, the visual cues with system-paced segmentation and the use of two modalities seemed to lead to a more effective allocation of mental capacity.

Hypothesis 3. Participants with spoken instruction examine the graphic more closely than those with written instruction.

Participants with spoken instruction made more gaze switches than participants with written instructions (see Table 1). The increase in gaze switches was particularly significant for gaze switches between AOIs that were part of the vector field. Students who were instructed aurally also made more gaze switches from the equation to the lower part of the vector field than those who were instructed via written text. The proportion of fixation duration differed significantly between the groups with written and spoken instruction. The group with written instruction spent more time

looking at the cue in the x -direction, whereas the group with spoken instruction looked more at the coordinate system and the lower part of the vector field plot. Therefore, we could not replicate the results of Schmidt-Weigand et al. [46].

There are three cognitive processes related to comprehending multimedia information: selection, organisation, and integration. First, the relevant information has to be selected, then it has to be organised into a coherent model, and finally, it can be integrated with each other and with prior knowledge [14]. Understanding these processes is important in order to design effective learning materials and understand possible learner problems. In this study, proportional fixation duration was chosen to enable the interpretation of proportional attention on specific AOIs and to account for individual differences in average fixation duration [44]. Participants with written instruction looked longer at the cue in the x -direction than participants with spoken instruction, whereas participants with spoken instruction looked longer at the coordinates and the lower part of the vector field, indicating differences in organisational processing [7]. This could also be due to the lower amount of visually presented information in spoken instruction but highlighted the importance of these AOIs for information selection. The difference is also interesting because the proportional fixation duration on the equation did not vary between instruction types. Participants seemed to pay equal attention to the equation. In the case of written instruction, the visual cue in the x -direction appeared to pull attention towards the area explained in the instruction as intended by signalling [3, 4]. However, students seemed to pay attention to both cues, as evidenced by longer proportional fixation durations [8]. This could indicate increased organisational processing necessary to structure information [7]. During auditory instructions, participants looked longer at the coordinate system and the adjoining lower part of the vector field, which could mean that they were trying to identify how the x and y “components” mentioned in the audio were related to each other. This is similar to Luzón and Letón [39] who found that the use of animated text led to more sense making and information selection processes. This could suggest that the use of animations is a good way of directing attention in instructional videos to help students deal with complex material (see e.g., [32]). In summary, what information is considered relevant seems to depend on the modality of instruction.

During written instruction, students looked less often from the coordinates to the equation and the lower part of the vector field. This suggests an increased amount of integration of the mathematical and graphical representations of a vector field and improved concept learning when instructions were presented aurally [7, 13]. Similarly, Smith et al. [48] interpreted gaze switches between mathematical and textual information as coordination between the two representations. The integration of equation and vector field corresponds with the result that participants with spoken instruction spent more time on the coordinates than those with written instruction. This could mean that students try to relate the directions in the graphical representation to the variables mentioned in the equation. In addition,

students with spoken instructions switched more often from the cue in the x -direction to the cue in the y -direction, indicating a slightly increased effort to integrate these directions. Relating between cues in a graphical representation could be a sign of local coherence formation within the graphical representation [18], in this case, a vector field plot. Together with the increased attention to the coordinate system and the better results in the posttest, heightened integration effort could indicate that the spoken instruction facilitates concept knowledge about the equation and its interpretation during the visual process. Possible problems of participants with written instruction to use the cue in x -direction when determining the curl support this interpretation.

However, the direction of the effect is not always clear: in some cases, participants with audio-based instruction looked between particular AOIs more often, whereas participants with written instruction switched between other AOIs more often. Some significant gaze switches also occurred between adjacent AOIs, such as the coordinate system and the lower vector field. This makes a detailed interpretation of the gaze switches difficult.

In conclusion, we found differences in visual behaviour between participants with spoken and those with written instructions, indicating different cognitive processing depending on the modality in which instructions were presented. Students with spoken instruction and dynamic visual cues acting as system-paced segmentation seemed to integrate the equation more fully with the graphical representation, helping students relate aspects, such as the x and y components. To our knowledge, few multimedia studies have examined convention-based graphical representations that are common in many STEM subjects. Research that finds multimedia principles not only for pictorial but also convention-based graphical representations could be very useful in other STEM contexts. It could help students learn problem-solving processes or acquire underlying conceptual knowledge, such as how to transfer a deep understanding of an equation to another graphical representation.

This study has several limitations. In terms of instructional pace, in the written condition, participants were able to jump between text and graphical representation as needed, possibly rereading some parts or skipping others. Participants who were instructed using audio-based material could not fast-forward, rewind, or pause the audio, and therefore had no choice in how they structured their own learning material. This might have had an impact on extraneous cognitive load, as experienced participants could, for example, focus on the graphic rather than the text and thus avoid for them redundant information. This could be avoided in self-paced learning material with sequenced audio-based instruction.

Second, there were several aspects of the study design that could be improved. Extraneous cognitive load was very low, with possible floor effects. This suggests that extraneous cognitive load might not have been measured with sufficient sensitivity. As Yang et al. [13] found that cognitive style can aid information integration, it would be beneficial to have information about participants' cognitive style, which might have reduced/increased instructional effectiveness. In addition,

the spoken instruction was segmented by dynamic visual cues, whereas there was no segmentation in the written instruction. This makes it impossible to determine whether the differences between the two types of instruction were due to modality or system-paced segmentation.

Third, prior knowledge was tested in a pretest identical to the posttest. A prior knowledge test that addresses specific features of the interpretation of vector fields might help explain the low response scores in the pretest. It could also help identify aspects that students feel unsure about and might be one aspect that explains the low confidence in spoken instruction in the posttest. For example, certain sections of text might have been reread or fixed for longer than others, which would not have been possible during spoken instruction. We also did not have the opportunity to validate the pre- and posttest with participants from different samples.

Fourth, because participants were recruited voluntarily and needed to have some prior mathematical knowledge to understand the equation, our sample size was limited and heterogeneous. However, Strohmaier et al. [63] reported an average sample size of $N = 29$ for eye-tracking studies in mathematics education research. Our sample size is therefore within the norm for our field.

There are several possibilities for future research. As mentioned earlier, replicating the study with a sequenced learner-paced audio instruction might account for differences in learning time and allow participants to study the material as they wish. This could provide additional insights into cognitive load in written and spoken instruction. Yang et al. [13] found that incorporating participants' cognitive styles, such as verbal presentation of information, into instructions could enhance information integration. This could also be used to further improve instructions, for example, by allowing participants to choose their preferred style of instruction. In addition, a more sensitive test for cognitive load could be useful due to the floor effects in extraneous cognitive load found in this study.

Another method for instructing a visual strategy is to use the eye movements of an expert as instructional material [51]. A comparison of this method with written and spoken instructions could prove beneficial in finding the best methods for teaching an efficient visual strategy. It would also be good to replicate our results with a more homogeneous participant sample, as well as in other contexts and with delayed posttests after some time to test the retention of the visual strategy. In addition, it would be useful to include participants from other disciplines to transfer this type of instruction to other fields where convention-based graphic representations are used.

6. Conclusion

This study compared instructional materials based on symbolic-mathematical and visual-graphic representations of vector fields, containing either written explanations with static visual cues or spoken explanations with dynamic visual cues. A difference in fixation durations and numbers of gaze switches between relevant parts of the instruction showed different visual strategies for text-based and audio-

based instructions, suggesting different cognitive strategies, in particular in organising and integrating information. Participants with audio-based instruction performed better in the posttest than those with text-based instruction, confirming our hypothesis. Contrary to expectations, there was no difference in extraneous cognitive load. This suggests that the instructions can be improved by spoken text and dynamic cues. However, we could not conclusively clarify whether the modality (e.g., spoken text) or the system-paced segmentation (due to dynamic cues) was the cause for this. This is a point for future research as well as for a more detailed investigation of learners' cognitive load.

7. Endnotes

Cohen's h is according to equations (1) and (2) with p being the sample proportion:

$$h = |\phi_1| - |\phi_2|, \quad (1)$$

$$\phi = 2 \cdot \arcsin \sqrt{p}. \quad (2)$$

Data Availability

The data that support the findings of this study are openly available in Open Science Framework (OSF) at https://osf.io/4uz8j/?view_only=bbb53db3d5fe47749aecadb41cc61f22.

Disclosure

The research was performed as part of the authors' employment at the TU Kaiserslautern.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

We are grateful to Sarah Malone for very helpful comments regarding the analysis and discussion of the results. Furthermore, we thank Susanne Schneider for the organisation of the study at the Georg-August University Göttingen and Johannes Lewing for his support during the performance of the measurements in Göttingen.

References

- [1] L. Bollen, P. van Kampen, C. Baily, M. Kelly, and M. De Cock, "Student difficulties regarding symbolic and graphical representations of vector fields," *Physical Review Physics Education Research*, vol. 13, no. 2, article 020109, 2017.
- [2] C. Singh and A. Maries, "Core graduate courses: a missed learning opportunity?," *AIP Conference Proceedings*, vol. 1513, no. 1, pp. 382–385, 2013.
- [3] T. Van Gog, "The signaling (or cueing) principle in multimedia learning," in *The Cambridge handbook of multimedia learning*, Richard E. Mayer, Ed., pp. 263–278, Cambridge University Press, 2014.
- [4] P. Klein, J. Viiri, and J. Kuhn, "Visual cues improve students' understanding of divergence and curl: evidence from eye

- movements during reading and problem solving,” *Physical Review Physics Education Research*, vol. 15, no. 1, article 010126, 2019.
- [5] R. Low and J. Sweller, “The modality principle in multimedia learning,” *The Cambridge handbook of multimedia learning*, vol. 147, p. 158, 2005.
 - [6] J. Sweller, “Implications of cognitive load theory for multimedia learning,” in *The Cambridge Handbook of Multimedia Learning*, pp. 19–30, Cambridge University Press, 2005.
 - [7] E. Alemdag and K. Cagiltay, “A systematic review of eye tracking research on multimedia learning,” *Computers & Education*, vol. 125, pp. 413–428, 2018.
 - [8] M. A. Just and P. A. Carpenter, “A theory of reading: from eye fixations to comprehension,” *Psychological Review*, vol. 87, no. 4, pp. 329–354, 1980.
 - [9] L. Bollen, P. van Kampen, C. Baily, and M. De Cock, “Qualitative investigation into students’ use of divergence and curl in electromagnetism,” *Physical Review Physics Education Research*, vol. 12, no. 2, article 020134, 2016.
 - [10] L. Bollen, P. van Kampen, and M. De Cock, “Students’ difficulties with vector calculus in electrodynamics,” *Physical Review Special Topics - Physics Education Research*, vol. 11, no. 2, article 020129, 2015.
 - [11] P. Klein, J. Viiri, S. Mozaffari, A. Dengel, and J. Kuhn, “Instruction-based clinical eye-tracking study on the visual interpretation of divergence: how do students look at vector field plots?,” *Physical Review Physics Education Research*, vol. 14, no. 1, article 010116, 2018.
 - [12] S. Mozaffari, M. Al-Naser, P. Klein et al., “Classification of visual strategies in physics vector field problem-solving,” *ICAART*, vol. 2, pp. 257–267, 2020.
 - [13] F.-Y. Yang, M.-J. Tsai, G.-L. Chiou, S. W.-Y. Lee, C.-C. Chang, and L.-L. Chen, “Instructional suggestions supporting science learning in digital environments based on a review of eye tracking studies,” *Journal of Educational Technology & Society*, vol. 21, no. 2, pp. 28–45, 2018.
 - [14] R. E. Mayer, “Cognitive theory of multimedia learning,” *The Cambridge handbook of Multimedia Learning*, vol. 41, pp. 31–48, 2005.
 - [15] R. E. Mayer, “Using multimedia for e-learning,” *Journal of Computer Assisted Learning*, vol. 33, no. 5, pp. 403–423, 2017.
 - [16] S. Y. Mousavi, R. Low, and J. Sweller, “Reducing cognitive load by mixing auditory and visual presentation modes,” *Journal of Educational Psychology*, vol. 87, no. 2, pp. 319–334, 1995.
 - [17] R. E. Mayer, “Multimedia learning: are we asking the right questions?,” *Educational Psychologist*, vol. 32, no. 1, pp. 1–19, 1997.
 - [18] T. Seufert and R. Brünken, “Cognitive load and the format of instructional aids for coherence formation,” *Applied Cognitive Psychology*, vol. 20, no. 3, pp. 321–331, 2006.
 - [19] A. J. Dontre, “The influence of technology on academic distraction: a review,” *Human Behavior and Emerging Technologies*, vol. 3, no. 3, pp. 379–390, 2021.
 - [20] D. Mutlu-Bayraktar, V. Cosgun, and T. Altan, “Cognitive load in multimedia learning environments: a systematic review,” *Computers & Education*, vol. 141, article 103618, 2019.
 - [21] J. Sweller, J. J. Van Merriënboer, and F. G. Paas, “Cognitive architecture and instructional design,” *Educational Psychology Review*, vol. 10, no. 3, pp. 251–296, 1998.
 - [22] J. Sweller, “Element interactivity and intrinsic, extraneous, and germane cognitive load,” *Educational Psychology Review*, vol. 22, no. 2, pp. 123–138, 2010.
 - [23] T. Zu, J. Hutson, L. C. Loschky, and N. S. Rebello, “Using eye movements to measure intrinsic, extraneous, and germane load in a multimedia learning environment,” *Journal of Educational Psychology*, vol. 112, no. 7, pp. 1338–1352, 2020.
 - [24] M. I. Siddiquei and S. Kathpal, “Challenges of Online Teaching during COVID-19: An Exploratory Factor Analysis,” *Human Behavior and Emerging Technologies*, vol. 3, no. 5, pp. 811–822, 2021.
 - [25] D. Vargo, L. Zhu, B. Benwell, and Z. Yan, “Digital technology use during COVID-19 pandemic: a rapid review,” *Human Behavior and Emerging Technologies*, vol. 3, no. 1, pp. 13–24, 2021.
 - [26] C. Müller and T. Mildenerger, “Facilitating flexible learning by replacing classroom time with an online learning environment: a systematic review of blended learning in higher education,” *Educational Research Review*, vol. 34, article 100394, 2021.
 - [27] C. Hughes, J. Costley, and C. Lange, “The effects of multimedia video lectures on extraneous load,” *Distance Education*, vol. 40, no. 1, pp. 54–75, 2019.
 - [28] J. Costley and C. H. Lange, “The effects of lecture diversity on germane load,” *The International Review of Research in Open and Distance Learning*, vol. 18, no. 2, pp. 27–46, 2017.
 - [29] H. Y. Lee and A. List, “Processing of texts and videos: a strategy-focused analysis,” *Journal of Computer Assisted Learning*, vol. 35, no. 2, pp. 268–282, 2019.
 - [30] R. Brünken, J. L. Plass, and D. Leutner, “Assessment of cognitive load in multimedia learning with dual-task methodology: auditory load and modality effects,” *Instructional Science*, vol. 32, no. 1/2, pp. 115–132, 2004.
 - [31] P. Ginns, “Meta-analysis of the modality effect,” *Learning and Instruction*, vol. 15, no. 4, pp. 313–331, 2005.
 - [32] R. E. Mayer, “Evidence-based principles for how to design effective instructional videos,” *Journal of Applied Research in Memory and Cognition*, vol. 10, no. 2, pp. 229–240, 2021.
 - [33] S. Herrlinger, T. N. Höffler, M. Opfermann, and D. Leutner, “When do pictures help learning from expository text? Multimedia and modality effects in primary schools,” *Research in Science Education*, vol. 47, no. 3, pp. 685–704, 2017.
 - [34] A. Wong, W. Leahy, N. Marcus, and J. Sweller, “Cognitive load theory, the transient information effect and e-learning,” *Learning and Instruction*, vol. 22, no. 6, pp. 449–457, 2012.
 - [35] W. Leahy and J. Sweller, “Cognitive load theory and the effects of transient information on the modality effect,” *Instructional Science*, vol. 44, no. 1, pp. 107–123, 2016.
 - [36] R. N. Soicher and K. A. Becker-Blease, “Testing the segmentation effect of multimedia learning in a biological system,” *Journal of Computer Assisted Learning*, vol. 36, no. 6, pp. 825–837, 2020.
 - [37] K. Stiller and P. Zinnbauer, “Does segmentation of complex instructional videos in big steps foster learning? The segmentation principle applied to a classroom video,” in *Proceedings of the Global Learn Asia Pacific 2011-Global Conference on Learning and Technology*, pp. 2044–2053, Melbourne, Australia, 2011.
 - [38] W. Bao, “COVID-19 and online teaching in higher education: a case study of Peking University,” *Human Behavior and Emerging Technologies*, vol. 2, no. 2, pp. 113–115, 2020.
 - [39] J. M. Luzón and E. Letón, “Use of animated text to improve the learning of basic mathematics,” *Computers & Education*, vol. 88, pp. 119–128, 2015.

- [40] S. M. Crooks, J. Cheon, F. Inan, F. Ari, and R. Flores, "Modality and cueing in multimedia learning: examining cognitive and perceptual explanations for the modality effect," *Computers in Human Behavior*, vol. 28, no. 3, pp. 1063–1071, 2012.
- [41] S. Schneider, F. Kriegelstein, M. Beege, and G. D. Rey, "How organization highlighting through signaling, spatial contiguity and segmenting can influence learning with concept maps," *Computers and Education Open*, vol. 2, article 100040, 2021.
- [42] J. Richter and K. Scheiter, "Studying the expertise reversal of the multimedia signaling effect at a process level: evidence from eye tracking," *Instructional Science*, vol. 47, no. 6, pp. 627–658, 2019.
- [43] H.-J. Jeung, P. Chandler, and J. Sweller, "The role of visual indicators in dual sensory mode instruction," *Educational Psychology*, vol. 17, no. 3, pp. 329–345, 1997.
- [44] K. Holmqvist and R. Andersson, *Eye Tracking: A Comprehensive Guide to Methods, Paradigms and Measures*, Lund Eye-Tracking Research Institute, 2017.
- [45] B. Park, L. Knörzer, J. L. Plass, and R. Brünken, "Emotional design and positive emotions in multimedia learning: an eye-tracking study on the use of anthropomorphisms," *Computers & Education*, vol. 86, pp. 30–42, 2015.
- [46] F. Schmidt-Weigand, A. Kohnert, and U. Glowalla, "Explaining the modality and contiguity effects: new insights from investigating students' viewing behaviour," *Applied Cognitive Psychology*, vol. 24, no. 2, pp. 226–237, 2010.
- [47] A. Schüler and M. Merkt, "Investigating text-picture integration in videos with the multimedia contradiction paradigm," *Journal of Computer Assisted Learning*, vol. 37, no. 3, pp. 718–734, 2021.
- [48] A. D. Smith, J. P. Mestre, and B. H. Ross, "Eye-gaze patterns as students study worked-out examples in mechanics," *Physical Review Special Topics - Physics Education Research*, vol. 6, no. 2, article 020118, 2010.
- [49] C.-Y. Wang, M.-J. Tsai, and C.-C. Tsai, "Multimedia recipe reading: predicting learning outcomes and diagnosing cooking interest using eye-tracking measures," *Computers in Human Behavior*, vol. 62, pp. 9–18, 2016.
- [50] L. Mason, P. Pluchino, and M. C. Tornatora, "Effects of picture labeling on science text processing and learning: evidence from eye movements," *Reading Research Quarterly*, vol. 48, no. 2, pp. 199–214, 2013.
- [51] L. Mason, P. Pluchino, and M. C. Tornatora, "Using eye-tracking technology as an indirect instruction tool to improve text and picture processing and learning," *British Journal of Educational Technology*, vol. 47, no. 6, pp. 1083–1095, 2016.
- [52] S. Malone, K. Altmeyer, M. Vogel, and R. Brünken, "Homogeneous and heterogeneous multiple representations in equation-solving problems: an eye-tracking study," *Journal of Computer Assisted Learning*, vol. 36, no. 6, pp. 781–798, 2020.
- [53] J. A. Harsh, M. Campillo, C. Murray, C. Myers, J. Nguyen, and A. V. Maltese, "“Seeing” data like an expert: an eye-tracking study using graphical data representations," *Education*, vol. 18, no. 3, p. ar32, 2019.
- [54] P. Klein, S. Küchemann, S. Brückner, O. Zlatkin-Troitschanskaia, and J. Kuhn, "Student understanding of graph slope and area under a curve: a replication study comparing first-year physics and economics students," *Physical Review Physics Education Research*, vol. 15, no. 2, article 020116, 2019.
- [55] J. Leppink, F. Paas, C. P. Van der Vleuten, T. Van Gog, and J. J. Van Merriënboer, "Development of an instrument for measuring different types of cognitive load," *Behavior Research Methods*, vol. 45, no. 4, pp. 1058–1072, 2013.
- [56] V. Ruf, S. Küchemann, J. Kuhn, and P. Klein, *Comparison of written and spoken instruction to foster coordination between diagram and equation in undergraduate physics education - data*, Open Science Framework (OSF), 2021, <https://osf.io/4uz8j/>.
- [57] L. Ding and R. Beichner, "Approaches to data analysis of multiple-choice questions," *Physical Review Special Topics - Physics Education Research*, vol. 5, no. 2, pp. 1–17, 2009.
- [58] S. Küchemann, S. Malone, P. Edelsbrunner et al., "Inventory for the assessment of representational competence of vector fields," *Physical Review Physics Education Research*, vol. 17, no. 2, article 20126, 2021.
- [59] A. B. Tobii, "Eye tracking for research," 2020, <https://www.tobiiipro.com/>.
- [60] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the 2000 symposium on Eye tracking research & applications*, pp. 71–78, Palm Beach Gardens Florida USA, 2000.
- [61] 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," *Computers in Human Behavior*, vol. 108, article 106316, 2020.
- [62] B. A. Simonsmeier, M. Flaig, A. Deiglmayr, L. Schalk, and M. Schneider, "Domain-specific prior knowledge and learning: a meta-analysis," *Educational Psychologist*, vol. 57, no. 1, pp. 31–54, 2022.
- [63] A. R. Strohmaier, K. J. MacKay, A. Obersteiner, and K. M. Reiss, "Eye-tracking methodology in mathematics education research: a systematic literature review," *Educational Studies in Mathematics*, vol. 104, no. 2, pp. 147–200, 2020.